Compendium of Mathematics & Physics

Nicolas Dewolf

May 31, 2025

Contents

Contents **1**

I Set Theory, Algebra & Category Theory **34**

1 Set Theory **38**
1.1 Axiomatization . . . 39
1.1.1 ZFC . . . 39
1.1.2 Material set theory . . . 40
1.1.3 Structural set theory . . . 40
1.1.4 ETCS ♣ . . . 40
1.1.5 Universes . . . 41
1.2 Finite sets . . . 41
1.3 Real numbers . . . 41
1.4 Set operations . . . 42
1.5 Functions . . . 43
1.5.1 (Co)domain . . . 43
1.5.2 Functions . . . 43
1.6 Collections . . . 44
1.6.1 Families and filters . . . 44
1.6.2 Algebra of sets . . . 45
1.7 Ordered sets . . . 46
1.7.1 Posets . . . 46
1.7.2 Ordinals ♣ . . . 48
1.7.3 Cardinals ♣ . . . 49
1.7.4 Lattices . . . 51
1.8 Limits . . . 53
1.9 Partitions . . . 54
1.9.1 Partition . . . 54
1.9.2 Superpartition ♣ . . . 55

2 Algebra **57**
2.1 Algebraic structures . . . 58

2.2 Group theory . . . 59
2.2.1 Abelian groups . . . 60
2.2.2 Cosets . . . 60
2.2.3 Sylow theorems . . . 61
2.2.4 Symmetric groups . . . 62
2.2.5 Group presentations . . . 63
2.2.6 Direct products . . . 64
2.2.7 Free groups . . . 66
2.3 Group actions . . . 67
2.4 Rings . . . 70
2.4.1 General . . . 70
2.4.2 Ideals . . . 73
2.4.3 Algebraic geometry . . . 75
2.4.4 Modules . . . 77
2.4.5 Semisimplicity . . . 79
2.5 Galois theory ♣ . . . 79
2.5.1 Polynomials . . . 79
2.5.2 Field extensions . . . 81
2.5.3 Algebraicity . . . 81
2.5.4 Galois extensions . . . 83
2.5.5 Radicals . . . 84
2.5.6 Nullstellensatz . . . 84

3 Category theory **85**
3.1 Categories . . . 86
3.2 Functors . . . 87
3.2.1 Natural transformations . . . 88
3.2.2 Equivalences . . . 90
3.2.3 Stuff, structure and property . . . 91
3.2.4 Adjunctions . . . 91
3.3 General constructions . . . 93
3.3.1 Fibred categories ♣ . . . 94
3.3.2 Monads . . . 96
3.4 Morphisms and diagrams . . . 99
3.4.1 Morphisms . . . 99
3.4.2 Initial and terminal objects . . . 100
3.4.3 Lifts . . . 101
3.4.4 Limits and colimits . . . 103
3.5 Internal structures . . . 111
3.5.1 Groupoids . . . 113

3.6 Lawvere theories ♣ . . . 114
3.7 Operad theory ♣ . . . 115
3.7.1 Operads . . . 115
3.7.2 Algebraic topology . . . 116

4 Higher-dimensional Algebra ♣ 117
4.1 Monoidal categories . . . 118
4.1.1 Braided categories . . . 120
4.1.2 Monoidal functors . . . 120
4.1.3 Closed categories . . . 122
4.2 Enriched category theory . . . 123
4.2.1 Enriched constructions . . . 125
4.2.2 Weighted (co)limits . . . 128
4.3 Abelian categories . . . 130
4.3.1 Additive categories . . . 130
4.3.2 Exact functors . . . 132
4.3.3 Finiteness . . . 133
4.4 Duality . . . 135
4.5 Tensor and fusion categories . . . 138
4.6 Ribbon and modular categories . . . 140
4.7 Module categories . . . 141
4.8 Higher vector spaces . . . 142
4.8.1 Kapranov–Voevodksy 2-vector spaces . . . 142
4.8.2 Baez–Crans 2-vector spaces . . . 142
4.9 Higher Lie theory . . . 143
4.9.1 Lie superalgebras . . . 143
4.9.2 Lie n-algebras . . . 145
4.10 Monoidal n-categories . . . 148
4.10.1 Relation with group cohomology . . . 149

5 Homological algebra 151
5.1 Chain complexes . . . 151
5.2 Exact sequences . . . 154
5.3 Resolutions . . . 155
5.4 Derived functors . . . 156
5.4.1 Modules . . . 158
5.5 Spectral sequences . . . 160
5.5.1 Exact couples . . . 160
5.5.2 Filtered complexes . . . 162
5.5.3 Convergence . . . 163
5.6 Examples . . . 164

5.6.1 Group cohomology 164
5.6.2 Hochschild and cyclic homology ♣ 166

II Topology **168**

6 General Topology **172**
6.1 Topological spaces 172
6.1.1 Neighbourhoods 174
6.1.2 Separation axioms 175
6.1.3 Convergence 176
6.2 Morphisms 176
6.2.1 Continuity 176
6.2.2 Homeomorphisms 178
6.3 Associated constructions 179
6.4 Connected spaces 181
6.5 Compact spaces 182
6.5.1 Compactness 182
6.5.2 Compactifications 184
6.6 Uniform spaces 184
6.7 Bornological spaces 185
6.8 Locales ♣ 185
6.8.1 Frames 185
6.8.2 Stone spaces 186

7 Algebraic Topology **188**
7.1 Homotopy theory 189
7.1.1 Homotopy 189
7.1.2 Homotopy groups 190
7.1.3 CW complexes 195
7.1.4 Fibrations 196
7.1.5 Rational homotopy theory ♣ 198
7.1.6 Equivariant homotopy theory ♣ 200
7.2 Simplicial homology 201
7.2.1 Simplices 201
7.2.2 Simplicial homology 202
7.2.3 Relative homology 205
7.2.4 Examples 206
7.3 Singular homology 207
7.4 Local coefficients 210
7.5 Axiomatic approach ♣ 210

7.6 Equivariant cohomology ♣ . . . 212
7.6.1 Borel construction . . . 213
7.6.2 Classifying topoi and stacks . . . 213

8 Sheaf Theory **215**
8.1 Presheaves . . . 215
8.2 Sheaves . . . 216
8.3 Equivariance . . . 219
8.4 Abelian sheaf cohomology . . . 220
8.4.1 Derived cohomology . . . 220
8.4.2 Čech cohomology . . . 221
8.5 Non-Abelian sheaf cohomology . . . 222
8.5.1 Čech cohomology . . . 222
8.6 Ringed spaces . . . 223

9 Metric Spaces **225**
9.1 Definition . . . 225
9.2 Topology . . . 226
9.3 Examples . . . 228
9.4 Completeness . . . 229
9.4.1 Convex metric spaces . . . 229
9.5 Compactness . . . 229

10 Algebraic Geometry **231**
10.1 Varieties . . . 231
10.1.1 Topology . . . 234
10.1.2 Varieties . . . 236
10.1.3 Projective varieties . . . 236
10.2 Schemes . . . 239
10.2.1 Morphisms . . . 242
10.2.2 Tangent space . . . 244
10.3 Sheaf theory ♣ . . . 245
10.4 Algebraic groups & Galois theory . . . 248
10.5 Tropical geometry . . . 248

11 Number Theory **250**
11.1 Adic numbers . . . 250
11.1.1 Finite fields . . . 250
11.1.2 Rational numbers . . . 251
11.1.3 Number and global fields . . . 252
11.2 Algebraic geometry ♣ . . . 254

11.2.1 Modular forms . . . 254
11.2.2 Algebraic functions . . . 256
11.3 L-functions . . . 256

III Higher Set Theory 259

12 Topos theory 262
12.1 Elementary topoi . . . 262
12.2 Morphisms . . . 264
12.3 Internal logic . . . 266
12.4 Grothendieck topoi . . . 268
12.4.1 Grothendieck topologies . . . 268
12.4.2 Topological sheaves . . . 273
12.4.3 Lawvere–Tierney topology . . . 275
12.4.4 Diaconescu's theorem . . . 277

13 Model theory ♣ 281
13.1 Simplicial sets . . . 282
13.1.1 Constructions . . . 283
13.1.2 Homological algebra . . . 285
13.2 Localization . . . 286
13.3 Model categories . . . 288
13.3.1 Model structures . . . 288
13.3.2 Monoidal structures . . . 290
13.3.3 Homotopy category . . . 291
13.3.4 Reedy model structure ♣ . . . 294
13.4 Simplicial spaces . . . 297
13.4.1 Kan complexes . . . 297
13.4.2 Simplicial localization . . . 298
13.4.3 Segal spaces . . . 299
13.4.4 Coherent nerve . . . 301
13.5 Cofibrantly generated categories . . . 303
13.5.1 Transfinite constructions . . . 303
13.5.2 Cofibrant generation . . . 305
13.6 Homotopy (co)limits . . . 307
13.6.1 Simplicially enriched diagrams . . . 308

14 Higher topos theory ♣ 310
14.1 Higher category theory . . . 310
14.1.1 n-categories . . . 310

14.1.2 n-functors . . . 313
14.1.3 Higher (co)limits . . . 314
14.2 ∞-Categories . . . 315
14.2.1 Simplicial approach . . . 315
14.2.2 Quasicategories . . . 315
14.2.3 Categorical notions . . . 316
14.3 Stacks . . . 316
14.3.1 2-sheaves . . . 316
14.3.2 Stacks on a 1-site . . . 318
14.4 Higher topos theory . . . 320
14.5 Cohomology . . . 321

15 Logic and Type Theory **324**
15.1 Logic . . . 325
15.1.1 Languages . . . 325
15.1.2 Propositional logic . . . 326
15.1.3 Sequent calculus . . . 327
15.1.4 Predicate logic . . . 328
15.2 Introduction to type theory . . . 328
15.3 Basic constructions . . . 329
15.3.1 Functions . . . 329
15.3.2 λ-calculus . . . 330
15.3.3 Products . . . 330
15.3.4 Propositions as types . . . 334
15.3.5 Identity types . . . 334
15.4 Categorical semantics . . . 337
15.4.1 Inductive types . . . 337
15.4.2 Polynomial functors . . . 337
15.5 Homotopy type theory . . . 338
15.5.1 Introduction . . . 338
15.5.2 Transport . . . 339
15.5.3 Equivalences . . . 341
15.5.4 Equality types: revisited . . . 342
15.6 Modal logic . . . 343
15.6.1 Modalities . . . 343
15.6.2 Type theory . . . 344

IV Calculus **347**

16 Calculus **351**

16.1 General definitions . . . 351
16.2 Continuity . . . 352
16.3 Convergence . . . 353
16.4 Series . . . 353
16.4.1 Convergence tests . . . 353
16.5 Differentiation . . . 355
16.6 Integration theory . . . 357
16.6.1 Riemann integral . . . 357
16.6.2 Euler integrals . . . 359
16.6.3 Gaussian integrals . . . 360
16.6.4 Generalizations . . . 361
16.7 Convexity . . . 362
16.8 Trigonometry . . . 364

17 Complex Analysis 365
17.1 Complex algebra . . . 365
17.2 Holomorphic functions . . . 366
17.3 Contour integrals . . . 367
17.4 Laurent series . . . 368
17.5 Singularities . . . 369
17.5.1 Poles . . . 369
17.5.2 Branch cuts . . . 370
17.5.3 Residue theorem . . . 371
17.6 Limit theorems . . . 372

18 Measure & Integration Theory 373
18.1 Measure theory . . . 374
18.1.1 General definitions . . . 374
18.1.2 Lebesgue measure . . . 376
18.1.3 Measurable functions . . . 377
18.1.4 Limit operations . . . 379
18.1.5 Polish spaces . . . 381
18.2 Lebesgue integral . . . 381
18.2.1 Simple functions . . . 381
18.2.2 Measurable functions . . . 382
18.2.3 Integrable functions . . . 382
18.2.4 Convergence theorems . . . 383
18.2.5 Relation to the Riemann integral . . . 385
18.3 Space of integrable functions . . . 385
18.3.1 Distance . . . 385
18.3.2 Hilbert space L^2 . . . 386

18.3.3 L^p-spaces . . . 386
18.4 Product measures . . . 388
18.4.1 Construction . . . 388
18.4.2 Fubini's theorem . . . 389
18.5 Radon–Nikodym theorem . . . 389
18.6 Lebesgue–Stieltjes integral . . . 391
18.7 Signed measures . . . 393
18.8 Bochner integral . . . 395

19 Distributions **396**
19.1 Distributions . . . 397
19.1.1 Derivatives . . . 399
19.1.2 Support . . . 399
19.1.3 Examples . . . 400
19.1.4 Growth rates . . . 401
19.2 Convolutions and kernels . . . 402
19.3 Transformations . . . 403
19.3.1 Fourier series . . . 403
19.3.2 Fourier transform . . . 404
19.3.3 Wave front sets . . . 406
19.3.4 Laplace transform . . . 407
19.3.5 Integral representations . . . 407
19.4 Harmonic analysis ♣ . . . 408

20 Ordinary differential equations **410**
20.1 Boundary conditions . . . 410
20.2 Existence and uniqueness . . . 411
20.3 First-order ODEs . . . 411
20.4 Second-order ODEs . . . 412
20.4.1 General solution . . . 412
20.4.2 Constant coefficients . . . 412
20.4.3 Frobenius' method . . . 413
20.5 Fourier transform . . . 414
20.6 Sturm–Liouville theory . . . 415
20.7 Applications . . . 416
20.7.1 Bessel functions . . . 416

21 Partial differential equations **418**
21.1 Important examples . . . 418
21.2 General linear equations . . . 419
21.3 First-order PDE . . . 420

21.4 Second-order PDE . . . 421
21.5 General methods . . . 423
21.5.1 Separation of variables . . . 423
21.6 Boundary conditions . . . 423
21.6.1 Dirichlet problem . . . 426
21.7 General . . . 426
21.7.1 Symbols . . . 426
21.8 Sobolev spaces . . . 427

V Linear Algebra 429

22 Linear Algebra 433
22.1 Vector spaces . . . 434
22.1.1 Linear independence . . . 434
22.1.2 Bases . . . 435
22.1.3 Sum and direct sum . . . 436
22.2 Linear maps . . . 437
22.2.1 Homomorphisms . . . 437
22.2.2 Dual maps . . . 438
22.3 Inner product . . . 439
22.3.1 Inner product space . . . 440
22.3.2 Orthogonality . . . 441
22.4 Matrices . . . 443
22.4.1 System of equations . . . 444
22.4.2 Coordinates and matrix representations . . . 444
22.4.3 Coordinate transformations . . . 446
22.4.4 Determinant . . . 448
22.4.5 Characteristic polynomial . . . 449
22.4.6 Matrix groups . . . 450
22.4.7 Matrix decompositions . . . 451
22.5 Eigenvectors . . . 452
22.5.1 Diagonalization . . . 453
22.5.2 Multiplicity . . . 453
22.6 Euclidean space . . . 454
22.7 Algebras . . . 455
22.8 Grassmanians . . . 458

23 Vector & Tensor Calculus 459
23.1 Nabla-operator . . . 460
23.2 Integration . . . 462

23.2.1 Line integrals . . . 462
23.2.2 Integral theorems . . . 463
23.3 Tensors . . . 463
23.3.1 Tensor product . . . 463
23.3.2 Transformation rules . . . 465
23.3.3 Tensor operations . . . 466
23.4 Exterior algebra . . . 467
23.4.1 Antisymmetric tensors . . . 467
23.4.2 Determinant . . . 468
23.4.3 Levi-Civita symbol . . . 468
23.4.4 Wedge product . . . 469
23.4.5 Exterior algebra . . . 470
23.4.6 Hodge star . . . 472
23.5 Graded vector spaces . . . 474
23.5.1 Supermatrices . . . 477
23.5.2 Berezin calculus . . . 478

24 Representation Theory 481
24.1 Group representations . . . 481
24.2 Irreducible representations . . . 482
24.3 Classification . . . 484
24.4 Generalized representations . . . 486
24.4.1 Projective representations . . . 486
24.4.2 Group cohomology . . . 487
24.5 Tensor operators . . . 487

25 Functional Analysis 489
25.1 Banach spaces . . . 490
25.1.1 Convex duality . . . 491
25.2 Hilbert spaces . . . 492
25.2.1 General . . . 492
25.2.2 Generalized Fourier series . . . 494
25.2.3 Orthogonality and projections . . . 495
25.2.4 Separable Hilbert spaces . . . 495
25.3 Seminorms . . . 496
25.3.1 Topology . . . 496
25.3.2 Locally convex spaces . . . 497
25.3.3 Tensor products . . . 500
25.3.4 Measure theory . . . 502
25.4 Riesz spaces . . . 502

26 Operator Theory **504**
26.1 Operators . . . 505
26.1.1 Operator topologies . . . 505
26.1.2 Bounded operators . . . 505
26.1.3 Self-adjoint operators . . . 507
26.1.4 Compact operators . . . 508
26.1.5 Spectrum . . . 510
26.1.6 Spectral theorem & functional calculus . . . 510
26.1.7 Dixmier trace . . . 513
26.2 C^*-algebras . . . 514
26.2.1 Involutive algebras . . . 514
26.2.2 Positive maps . . . 516
26.2.3 Traces . . . 517
26.2.4 Representations . . . 518
26.2.5 Gel'fand duality . . . 518
26.2.6 Hilbert modules ♣ . . . 521
26.3 von Neumann algebras . . . 522
26.3.1 Factors . . . 523

27 Clifford Algebra **524**
27.1 Clifford algebra . . . 524
27.2 Geometric algebra . . . 526
27.3 Bott periodicity ♣ . . . 527
27.4 Pin group . . . 528
27.4.1 Clifford group . . . 528
27.4.2 Pin and Spin . . . 530

28 Noncommutative Algebra **533**
28.1 Coalgebras . . . 533
28.2 Hopf algebras . . . 535
28.2.1 Drinfel'd double . . . 538
28.3 Quantum groups . . . 539
28.4 Differential structures . . . 541
28.4.1 Differential calculi . . . 541
28.4.2 Covariance . . . 542

VI (Differential) Geometry **543**

29 Elementary Geometry **551**
29.1 Straight lines . . . 551

29.2 Trigonometry . . . 551

30 Curves & Surfaces **554**
30.1 Curves . . . 554
30.1.1 Arc length . . . 555
30.1.2 Frenet–Serret frame . . . 555
30.2 Surfaces . . . 556
30.2.1 First fundamental form . . . 557
30.2.2 Isometries . . . 558
30.2.3 Second fundamental form . . . 558
30.2.4 Curvature of a surface . . . 559
30.2.5 Christoffel symbols and geodesics . . . 562
30.2.6 Theorema Egregium . . . 562

31 Manifolds **564**
31.1 Charts . . . 564
31.2 Tangent vectors . . . 567
31.3 Submanifolds . . . 569
31.3.1 Immersions and submersions . . . 569
31.3.2 Submanifolds . . . 570
31.4 Manifolds with boundary . . . 572
31.4.1 Cobordism theory ♣ . . . 572
31.5 Surgery theory ♣ . . . 573

32 Lie groups and Lie algebras **574**
32.1 Lie groups . . . 575
32.1.1 Left invariant vector fields . . . 575
32.1.2 One-parameter subgroups . . . 576
32.2 Lie algebras . . . 576
32.2.1 Definitions . . . 576
32.2.2 Examples . . . 579
32.2.3 Exponential map . . . 580
32.2.4 Solvable Lie algebras . . . 581
32.2.5 Simple Lie algebras . . . 582
32.2.6 Central extensions . . . 583
32.3 Representation theory . . . 583
32.3.1 Lie groups . . . 583
32.3.2 Lie algebras . . . 584
32.3.3 Coadjoint orbits . . . 584
32.4 Structure . . . 585
32.4.1 Killing form . . . 585

32.4.2 Weights, roots and Dynkin diagrams . . . 586
32.4.3 Highest weight theory . . . 592
32.4.4 Coxeter groups . . . 593
32.4.5 Kac–Moody algebras ♣ . . . 594
32.4.6 Universal enveloping algebra . . . 596
32.5 Group contractions . . . 598
32.6 Lie algebra cohomology ♣ . . . 598

33 Fibre Bundles **602**
33.1 Bundles . . . 602
33.2 Constructions . . . 605
33.3 Sections . . . 606

34 Vector Bundles **607**
34.1 Tangent bundle . . . 608
34.2 Vector bundles . . . 610
34.2.1 Sections . . . 611
34.2.2 Tubular neighbourhoods . . . 611
34.2.3 Sums and product . . . 611
34.2.4 Associated vector bundles . . . 612
34.3 Vector fields . . . 613
34.3.1 Integral curves . . . 615
34.3.2 Lie derivative . . . 616
34.3.3 Multivector fields . . . 618
34.4 Differential k-forms . . . 619
34.4.1 Definition . . . 619
34.4.2 Exterior derivative . . . 620
34.4.3 Lie derivative . . . 621
34.4.4 Interior product . . . 622
34.4.5 Lie derivative of tensor fields . . . 622
34.4.6 Vector-valued differential forms . . . 623
34.5 Grassmann bundle . . . 625
34.6 Frobenius' theorem . . . 627
34.7 Linear connections . . . 629
34.7.1 Koszul connections . . . 629
34.7.2 Induced connections . . . 630
34.7.3 Curvature & torsion . . . 632
34.8 Integration Theory . . . 633
34.8.1 Orientation and densities . . . 633
34.8.2 Orientation in homology . . . 635
34.8.3 Integration . . . 636

34.8.4 Stokes' theorem . . . 637
34.9 Cohomology . . . 638
34.9.1 de Rham complex . . . 638
34.9.2 Integration . . . 640
34.9.3 Cohomology with compact support . . . 641
34.9.4 Čech–de Rham complex . . . 645
34.9.5 Nonorientable manifolds . . . 646
34.9.6 Generalized cohomology ♣ . . . 648
34.10 Morse theory . . . 648
34.10.1 Morse functions . . . 648
34.10.2 Morse–Bott functions . . . 649
34.10.3 Morse homology . . . 649
34.11 Differential operators . . . 651
34.11.1 Elliptic complexes . . . 651

35 Principal Bundles 652
35.1 Principal bundles . . . 653
35.1.1 Associated bundles . . . 654
35.1.2 Sections . . . 656
35.2 Universal bundle . . . 657
35.3 Connections . . . 658
35.3.1 Vertical vectors . . . 658
35.3.2 Ehresmann connections . . . 660
35.3.3 Connection forms . . . 661
35.3.4 Local representations . . . 663
35.3.5 Parallel transport . . . 664
35.3.6 Holonomy group . . . 667
35.4 Covariant derivatives . . . 668
35.4.1 Koszul connections . . . 668
35.4.2 Exterior covariant derivative . . . 669
35.4.3 Curvature . . . 671
35.4.4 Torsion . . . 672
35.5 Reduction of the structure group . . . 673
35.6 Characteristic classes . . . 675
35.6.1 Chern–Weil theory . . . 676
35.6.2 Complex bundles . . . 677
35.6.3 Real bundles . . . 679
35.6.4 Lie group cohomology . . . 680
35.6.5 Chern–Simons forms . . . 680
35.7 Differential cohomology ♣ . . . 681

35.7.1 Differential characters . . . 681
35.7.2 Combining singular and de Rham cohomology . . . 682
35.7.3 Deligne cohomology . . . 683
35.8 Cartan geometry . . . 685
35.8.1 Klein geometry . . . 685
35.8.2 Cartan geometry . . . 686
35.9 $\mathcal{D}$-geometry ♣ . . . 688

36 Riemannian Geometry 689
36.1 Riemannian manifolds . . . 689
36.1.1 Metric . . . 689
36.1.2 Riemannian manifolds . . . 692
36.1.3 Levi-Civita connection . . . 692
36.1.4 Killing vectors . . . 694
36.2 Curvature . . . 695
36.3 Spinor bundles . . . 696
36.3.1 Spin structures . . . 696
36.3.2 Stiefel–Whitney classes . . . 698
36.3.3 Dirac operators ♣ . . . 699
36.4 Conformal structures . . . 702
36.5 Distributions ♣ . . . 704

37 Symplectic Geometry 705
37.1 Symplectic manifolds . . . 706
37.2 Poisson manifolds . . . 709
37.3 Lagrangians . . . 711
37.3.1 Lagrangian submanifolds . . . 711
37.3.2 Maslov classes . . . 713
37.3.3 Polarizations . . . 716
37.4 Hamiltonian dynamics . . . 717
37.4.1 Dynamical systems . . . 717
37.4.2 Hamilton–Jacobi equation . . . 719
37.4.3 Integrability . . . 720
37.5 Symplectic reduction . . . 723
37.5.1 Hamiltonian actions . . . 723
37.5.2 Symplectic reduction . . . 725
37.5.3 Poisson reduction . . . 725
37.5.4 Lie–Poisson reduction . . . 726
37.5.5 Infinite-dimensional manifolds . . . 727
37.6 Metaplectic structures . . . 727

38 Contact Geometry **729**
38.1 Contact structures . . . 729
38.1.1 Contact form . . . 729
38.1.2 Reeb vector fields . . . 730

39 Complex Geometry **731**
39.1 Complex structures . . . 731
39.2 Complex differential forms . . . 735
39.3 Holomorphic vector bundles . . . 736
39.4 Kähler manifolds . . . 737
39.4.1 Killing vectors . . . 740
39.4.2 Dirac operators . . . 741
39.5 Complex curves . . . 741
39.5.1 Riemann surfaces and orbifolds . . . 741
39.5.2 Pseudoholomorphic maps . . . 746
39.6 Cohomology . . . 748
39.6.1 Dolbeault cohomology . . . 748
39.6.2 Lagrangian Floer homology ♣ . . . 748

40 Calculus of Variations **753**
40.1 Introduction . . . 754
40.2 Jet bundles . . . 756
40.2.1 Topology . . . 756
40.2.2 Contact structure . . . 758
40.2.3 Differential operators . . . 758
40.3 Variational bicomplex ♣ . . . 759
40.3.1 Differential structure . . . 759
40.3.2 Euler operators . . . 763
40.3.3 Functional complex . . . 765
40.3.4 Structure of the bicomplex . . . 766
40.3.5 Variational problems . . . 770
40.3.6 Inverse problem . . . 772
40.3.7 Finite jet bundles . . . 773
40.4 Partial differential equations . . . 774
40.4.1 Algebraic formulation . . . 774
40.4.2 Symmetries . . . 776
40.4.3 Pseudogroups ♣ . . . 777

41 K-theory ♣ **778**
41.1 Preliminaries . . . 779
41.2 Topological K-theory . . . 779

41.2.1 Introduction . . . 779
41.2.2 Classification . . . 780
41.2.3 Negative degree . . . 782
41.2.4 Bott periodicity . . . 783
41.2.5 Clifford modules . . . 785
41.2.6 Cohomology theory . . . 786
41.2.7 Applications . . . 786
41.3 Twisted K-theory . . . 788
41.4 Differential K-theory . . . 788
41.5 Algebraic K-theory . . . 789
41.5.1 Determinant . . . 789
41.6 Operator K-theory . . . 790

42 Noncommutative Geometry 792
42.1 Noncommutative measure theory . . . 792
42.2 Quantum geometry . . . 795
42.3 Spectral geometry . . . 796
42.4 Dunkl operators . . . 800

43 Synthetic Differential Geometry ♣ 801
43.1 Neighbourhoods . . . 801
43.2 Affine connections . . . 802
43.3 Euclidean geometry . . . 802
43.3.1 Infinitesimal elements . . . 802
43.3.2 Calculus . . . 803

44 Higher-dimensional Geometry ♣ 805
44.1 Graded manifolds . . . 806
44.2 Infinite-dimensional geometry . . . 808
44.3 Cohesion . . . 809
44.4 Smooth spaces . . . 813
44.4.1 Frölicher spaces . . . 814
44.4.2 Smooth sets . . . 815
44.4.3 Smooth algebras . . . 819
44.4.4 Supergeometry . . . 819
44.5 Higher Lie theory . . . 824
44.6 Gauge theory ♣ . . . 824
44.6.1 Lie groupoids . . . 825
44.7 Weak groups . . . 826
44.7.1 Spaces . . . 828
44.8 2-Bundles . . . 829

44.9 Space and quantity . . . 831

VII Probability Theory & Statistics 833

45 Probability Theory 838
45.1 Probability . . . 839
45.2 Conditional probability . . . 840
45.3 Probability distributions . . . 841
45.3.1 Distribution functions . . . 841
45.3.2 Moments . . . 844
45.3.3 Correlation . . . 846
45.3.4 Conditional expectation . . . 847
45.4 Joint distributions . . . 848
45.4.1 Common distributions . . . 850
45.5 Information theory . . . 853
45.6 Extreme value theory . . . 854
45.7 Copulas . . . 855
45.8 Randomness ♣ . . . 856
45.9 Optimal transport . . . 857
45.9.1 Kantorovich duality . . . 857
45.9.2 Convex costs . . . 859
45.9.3 Concave costs . . . 860
45.9.4 Densities . . . 860
45.10 Categorical probability ♣ . . . 862
45.10.1 Probability monads . . . 862
45.10.2 Markov categories . . . 863

46 Stochastic Calculus 866
46.1 Stochastic processes . . . 867
46.2 Brownian motion . . . 869
46.2.1 Classical Wiener space . . . 869
46.2.2 Wiener process . . . 870
46.2.3 Abstract Wiener space . . . 871
46.3 Martingales . . . 875
46.3.1 Proper martingales . . . 875
46.3.2 Local martingales . . . 879
46.4 Itô calculus . . . 880
46.4.1 Wiener–Paley integral . . . 880
46.4.2 Stochastic integral . . . 882
46.4.3 Properties . . . 885

46.4.4 Lebesgue–Stieltjes integration . . . 889
46.4.5 Representation theorems . . . 890
46.5 Stochastic differential equations . . . 892
46.6 Markov processes . . . 893

47 Information Geometry ♣ 894
47.1 Statistical manifolds . . . 894
47.1.1 Divergences . . . 896
47.1.2 Exponential families . . . 899
47.1.3 Mixtures . . . 900
47.1.4 Compatible divergences . . . 901
47.1.5 Flat structures . . . 903
47.2 Projections . . . 903
47.2.1 Functional descent . . . 904

48 Statistics 906
48.1 Data samples . . . 907
48.1.1 Moment estimators . . . 907
48.1.2 Dispersion . . . 909
48.2 Empirical distribution . . . 910
48.3 Errors . . . 913
48.4 Parameter estimation . . . 914
48.4.1 General properties . . . 914
48.4.2 Common estimators . . . 915
48.4.3 Estimation error . . . 915
48.4.4 Likelihood function . . . 916
48.4.5 Maximum likelihood estimation . . . 916
48.4.6 Least squares estimation . . . 917
48.4.7 Geometric approach . . . 919
48.5 Bayesian modelling . . . 919
48.6 Confidence intervals . . . 920
48.6.1 Interval types . . . 920
48.6.2 General construction . . . 921
48.6.3 Interval for a sample mean . . . 921
48.7 Hypothesis testing . . . 922
48.7.1 Testing . . . 922
48.7.2 Comparison tests . . . 923
48.7.3 Post-hoc tests . . . 925
48.8 Goodness of fit . . . 926
48.8.1 χ^2-test . . . 926
48.8.2 Runs test . . . 927

49 Causal Inference **928**
49.1 Causal DAGs . . . 928
49.2 Do-calculus . . . 929

50 Data Analysis **930**
50.1 Data sampling . . . 931
50.1.1 Inverse CDF sampling . . . 931
50.1.2 Uniform rejection sampling . . . 932
50.1.3 Monte Carlo sampling . . . 932
50.2 Optimization Theory . . . 933
50.2.1 Linear problems . . . 933
50.2.2 Gradient descent . . . 935
50.2.3 Conjugate gradient . . . 936
50.2.4 Nonlinear conjugate gradients . . . 938
50.2.5 Krylov methods . . . 940
50.3 Constrained optimization . . . 940
50.3.1 Lagrange multipliers . . . 940
50.3.2 Riemannian gradient descent . . . 942
50.4 Approximation theory . . . 943
50.4.1 Bayes optimality . . . 943
50.4.2 PAC theory and empirical risk minimization . . . 943
50.5 Classification problems . . . 945
50.5.1 Clustering . . . 945
50.5.2 Nearest-neighbour search . . . 946
50.6 Support-vector machines . . . 946
50.6.1 Kernel methods . . . 946
50.6.2 Decision boundaries . . . 948
50.7 Complexity theory . . . 950
50.7.1 VC dimension . . . 950
50.7.2 Rademacher complexity . . . 952
50.7.3 Relation to Glivenko–Cantelli classes . . . 953
50.8 Time series analysis . . . 953
50.8.1 Stationarity . . . 953
50.8.2 Correlation . . . 954
50.8.3 Autoregressive models . . . 955
50.8.4 Granger causality . . . 957
50.9 Uncertainty quantification . . . 958
50.9.1 Prediction regions . . . 958
50.9.2 Classifier calibration . . . 959
50.9.3 Conditionality . . . 960

50.9.4 Distribution shift 961
50.10 Generative models 963
50.10.1 Normalizing flows 963

51 Fuzzy Sets & Imprecise Probabilities ♣ 965
51.1 Fuzzy sets 965
51.2 Imprecise probabilities 966
51.2.1 Gambles 966
51.2.2 Credal sets 967
51.2.3 Fuzzy measure theory 967

VIII Classical Physics 970

52 Classical Mechanics 974
52.1 Newtonian mechanics 975
52.1.1 Linear motion 975
52.1.2 Rotational motion 976
52.2 Lagrangian mechanics 979
52.2.1 Action 979
52.2.2 Euler–Lagrange equations 980
52.2.3 Kepler problem 985
52.3 Hamiltonian mechanics 986
52.3.1 Poisson bracket 987
52.3.2 Hamilton–Jacobi equation 988
52.4 Continuum mechanics 989
52.4.1 Liouville's theorem 989
52.4.2 Fluid mechanics 991
52.5 Koopman–von Neumann mechanics 992
52.6 Dynamical systems 994
52.7 Geometric mechanics 995
52.7.1 Newton's equations 995
52.7.2 Lagrangian mechanics 997
52.7.3 Hamiltonian mechanics 997
52.7.4 Contact structure 998

53 Classical Field Theory 1000
53.1 Lagrangian field theory 1000
53.1.1 Noether's theorem 1001
53.2 De Donder–Weyl theory 1003

54 BRST Theory 1004

54.1 Constrained systems 1005
54.1.1 Constraints 1005
54.1.2 Zero Hamiltonian 1010
54.1.3 Field theories 1011
54.1.4 Constraint surface 1012
54.1.5 Fermionic systems 1013
54.2 Hamiltonian BRST theory 1014
54.2.1 Introduction 1014
54.2.2 Irreducible constraints 1015
54.2.3 Reducible constraints 1019
54.2.4 Observables 1021
54.2.5 Actions and gauge fixing 1022
54.2.6 Nonminimal sectors 1022
54.2.7 Faddeev–Popov action 1023
54.3 Lagrangian BRST theory 1025
54.3.1 Field theory 1025
54.3.2 Gauge algebra 1026
54.3.3 Antifields 1028

55 Electromagnetism 1033
55.1 Electricity 1033
55.1.1 Conductors and resistors 1034
55.1.2 Ohm's law 1036
55.1.3 Electric dipoles 1036
55.2 Magnetism 1038
55.2.1 Lorentz force 1039
55.2.2 Energy and momentum 1039
55.2.3 Electric charges in a magnetic field 1040
55.3 Maxwell equations 1040
55.4 Potentials 1041
55.4.1 Decomposition in potentials 1041
55.4.2 Gauge theory 1041
55.4.3 Differential-geometric perspective 1042

56 Optics 1044
56.1 General 1044
56.1.1 Conservation of energy 1044
56.2 Wave mechanics 1045
56.3 Electromagnetic radiation 1046
56.4 Propagation in media 1048
56.4.1 Reflection 1048

56.4.2 Refraction 1048
56.4.3 Absorption 1049
56.4.4 Interference & Diffraction 1050

IX Quantum Theory 1052

57 Quantum Mechanics 1055
57.1 Introduction 1056
57.1.1 Dirac–von Neumann postulates 1056
57.2 Schrödinger picture 1058
57.3 Heisenberg–Born–Jordan picture 1060
57.3.1 Hydrogen atom 1061
57.3.2 Molecular dynamics 1062
57.4 Mathematical formalism 1062
57.4.1 Weyl systems 1062
57.4.2 Dirac–von Neumann postulates: revisited 1064
57.4.3 Symmetries 1064
57.4.4 Symmetric states 1065
57.5 Foundations ♣ 1066
57.5.1 Measurement problem 1066
57.5.2 Copenhagen interpretation 1066
57.5.3 Many-worlds interpretation 1066
57.5.4 Relational quantum mechanics 1067
57.6 Angular Momentum 1070
57.6.1 Angular momentum operator 1070
57.6.2 Rotations 1071
57.6.3 Spinor representation 1071
57.6.4 Coupling of angular momenta 1072
57.7 Electromagnetism 1073
57.7.1 Pauli theory 1073
57.7.2 Aharonov–Bohm effect 1073
57.8 Approximation methods 1074
57.8.1 WKB approximation 1074
57.9 Curved backgrounds ♣ 1076
57.9.1 Extending quantum mechanics 1076
57.9.2 WKB approximation 1077

58 Quantum Information Theory 1079
58.1 Entanglement 1079
58.1.1 Introduction 1079

58.1.2 Bell states 1080
58.1.3 SRE states 1081
58.2 Density operators 1081
58.3 Channels 1082
58.4 Quantum logic 1083
58.4.1 Birkhoff–von Neumann logic 1083
58.4.2 Linear logic 1084
58.5 Topos theory ♣ 1090
58.5.1 Bohr topos 1090
58.5.2 Internal logic 1091

59 Quantization ♣ 1093
59.1 Introduction 1093
59.2 Deformation quantization 1094
59.3 Geometric Quantization 1095
59.3.1 Prequantization 1095
59.3.2 Coadjoint orbits 1100
59.4 Toeplitz quantization 1100
59.5 Constrained systems 1102
59.5.1 Dirac procedure 1102
59.5.2 BRST quantization 1103
59.5.3 BFV quantization 1103

60 Perturbation Theory 1105
60.1 Interaction picture 1105
60.2 Rayleigh–Schrödinger theory 1106
60.3 Adiabatic approximation 1107
60.3.1 Berry phase 1107

61 Scattering Theory 1109
61.1 Cross sections 1109
61.2 Lippman–Schwinger equations 1109

X Statistical Mechanics & Condensed Matter Physics 1111

62 Thermodynamics 1114
62.1 General definitions 1114
62.2 Postulates 1115
62.3 Equations of state 1116
62.4 Gases 1116

63 Statistical Mechanics **1117**
63.1 Axioms . . . 1117
63.2 Temperature . . . 1118
63.3 Canonical ensemble . . . 1118
63.4 Grand canonical ensemble . . . 1119
63.5 Energy . . . 1120
63.6 Lattice systems . . . 1120
63.6.1 Ising model . . . 1120
63.7 Black-body radiation . . . 1122

64 Material Physics **1123**
64.1 Crystals . . . 1124
64.1.1 Analytic representation . . . 1124
64.2 Symmetries . . . 1124
64.3 Crystal lattice . . . 1124
64.3.1 Bravais lattice . . . 1125
64.3.2 Reciprocal lattice . . . 1125
64.4 Diffraction . . . 1125
64.4.1 Constructive interference . . . 1125
64.4.2 Intensity of diffracted beams . . . 1127
64.5 Alloys . . . 1127
64.6 Lattice defects . . . 1128
64.7 Electrical properties . . . 1129
64.7.1 Charge carriers . . . 1129
64.7.2 Band structure . . . 1129
64.7.3 Ferroelectricity . . . 1130
64.8 Magnetic properties . . . 1131
64.8.1 Paramagnetism . . . 1131
64.8.2 Ferromagnetism . . . 1132
64.8.3 Antiferromagnetism . . . 1132
64.8.4 Ferrimagnetism . . . 1133

65 Tensor Networks **1134**
65.1 Matrix Product States . . . 1134
65.1.1 Finite-dimensional lattices . . . 1134
65.1.2 Injectivity . . . 1136
65.1.3 Gauge freedom and canonical forms . . . 1136
65.1.4 Translation-invariant states . . . 1137
65.2 Matrix product operators . . . 1137
65.2.1 MPO-injectivity . . . 1138
65.2.2 MPO-symmetries for SPT phases . . . 1139

XI Relativity & Quantum Field Theory 1140

66 Special Relativity 1145
66.1 Lorentz transformations . . . 1145
66.2 Energy and momentum . . . 1146
66.3 Action principle . . . 1146

67 General Relativity 1149
67.1 Causal structure . . . 1149
67.2 Einstein field equations . . . 1151
67.3 Black holes . . . 1152
67.4 Conserved quantities . . . 1152
67.5 Tetradic formulation . . . 1153

68 Dirac Equation 1155
68.1 Dirac matrices . . . 1155
68.2 Spinors . . . 1156
68.2.1 Dirac equation . . . 1156
68.2.2 Chiral spinors . . . 1157
68.2.3 Dirac algebra in $d = 4$ . . . 1160
68.2.4 Fierz identities . . . 1160

69 Gauge Theory 1162
69.1 Gauge invariance . . . 1163
69.1.1 Gauge principle . . . 1163
69.1.2 Gauge fixing . . . 1164
69.1.3 Anomalies . . . 1166
69.2 Yang–Mills theory . . . 1166
69.2.1 Yang–Mills equations . . . 1166
69.2.2 Currents and matter coupling . . . 1167
69.2.3 2 dimensions . . . 1167
69.3 Spontaneous symmetry breaking . . . 1168
69.3.1 Higgs mechanism . . . 1168
69.4 Instantons . . . 1170
69.4.1 Instanton sectors . . . 1170
69.4.2 Large gauge transformations . . . 1171
69.4.3 Selfduality . . . 1172
69.4.4 ’t Hooft–Polyakov instantons . . . 1173
69.4.5 Dyons . . . 1173
69.4.6 Electric-magnetic duality . . . 1174
69.5 Chern–Simons theory . . . 1174

69.6 Quantization 1174
69.6.1 Auxiliary fields 1174

70 Quantum Field Theory 1175
70.1 Klein–Gordon field 1176
70.1.1 Lagrangian and Hamiltonian 1176
70.1.2 Raising and lowering operators 1176
70.1.3 Scalar propagator 1178
70.1.4 Normalization constant 1178
70.1.5 Invariant integration measure 1178
70.2 Contractions and Wick's theorem 1179
70.2.1 Bosonic fields 1179
70.2.2 Fermionic fields 1180
70.3 Gauge theory 1181
70.3.1 Schwinger–Dyson equation 1181
70.3.2 Faddeev–Popov integration 1183
70.3.3 BV-BRST theory 1184
70.4 Quantum chromodynamics 1186
70.5 Feynman rules 1186
70.5.1 Scalar theory 1186
70.5.2 Graph theory 1187
70.5.3 Degree of divergence 1188
70.5.4 Amplituhedron ♣ 1190
70.6 Renormalization 1190
70.6.1 Introduction: Statistical physics 1190
70.6.2 Dimensional regularization 1190
70.6.3 Wilsonian renormalization 1192
70.6.4 Connes–Kreimer renormalization 1192
70.7 Entanglement in QFT 1196
70.7.1 Lattice theories 1196

71 Particle Physics 1198
71.1 Overview 1198
71.2 Standard model 1200
71.2.1 Group theory 1200
71.2.2 Symmetry breaking 1201
71.3 Shortcomings 1203
71.3.1 Open problems 1203
71.3.2 Might-be problems 1203

72 Axiomatic QFT ♣ 1205

72.1 Topological QFT 1206
72.1.1 Atiyah–Segal axioms 1206
72.1.2 Open-closed TQFT 1207
72.2 Algebraic QFT 1207
72.2.1 Haag–Kastler axioms 1207
72.3 Perturbative AQFT 1209
72.3.1 Haag–Kastler axioms 1210
72.3.2 Functional analysis 1210
72.3.3 Equations of motion 1212
72.3.4 Symmetries 1215
72.3.5 Quantization 1216
72.4 Prequantum field theory 1218

73 Conformal Field Theory 1220
73.1 Introduction 1221
73.2 In dimension $d = 2$ 1221
73.2.1 Minkowski space 1222
73.3 Quantization in $d = 2$ 1223
73.3.1 Radial quantization 1224
73.3.2 Virasoro algebra 1225
73.3.3 Representation theory 1225
73.4 Vertex operator algebras 1226
73.4.1 Definition 1226
73.4.2 Rational CFTs 1227
73.5 Extensions ♣ 1227
73.5.1 TCFT 1227
73.5.2 Strange correlator 1227

74 Extensions and Unification 1229
74.1 Supersymmetry 1230
74.1.1 Supersymmetric quantum mechanics 1230
74.1.2 Loop space mechanics 1231
74.1.3 Extensions of the Standard Model 1231
74.1.4 Spinning particle 1232
74.2 Chern–Simons theory 1233
74.2.1 Holomorphic Chern–Simons theory 1233
74.3 Four-manifolds 1233
74.3.1 Donaldson–Witten theory 1234
74.3.2 Instanton Floer homology 1236
74.3.3 Seiberg–Witten theory 1237
74.4 String theory 1240

74.4.1 Worldsheets 1240
74.4.2 Branes 1241
74.4.3 Duality 1241
74.4.4 Superstrings 1241
74.5 M-theory 1242

75 Astronomy & Cosmology 1243
75.1 Stars 1243
75.2 Galaxies 1243
75.3 Black holes 1244
75.4 Inflation 1244
75.5 Tegmark multiverses 1244

XII Beyond Math & Physics 1246

76 Computer Science 1248
76.1 Computability 1248
76.1.1 Turing machines 1248
76.1.2 Functions 1249
76.2 Quantum computing 1249
76.3 Complexity theory 1249

77 Finance 1251
77.1 Securities 1251
77.2 Instruments and derivatives 1252
77.3 Markets 1254
77.3.1 Introduction 1254
77.3.2 Arbitrage 1257
77.3.3 Completeness 1258
77.4 Models 1261
77.4.1 Vašíček model 1261
77.4.2 Black–Scholes model 1261

78 Philosophy of science 1263
78.1 Modal logic 1263
78.1.1 Temporal logic 1263
78.1.2 Possibility & necessity 1263
78.2 Theories 1263
78.2.1 Emergence 1263

79 Epilogue 1264

79.1 Reading list 1264
79.2 Current topics 1265
79.3 Personal notes 1265

XIII Appendices 1266

Appendix A G-Structures 1268

Appendix B Notes 1271
B.1 Noether's Theorem and Gauge-Gravity Duality by S. De Haro 1271
B.1.1 Pseudotensors 1271
B.1.2 AdS-CFT duality 1273
B.2 A generalization of Noether's theorem and the information-theoretic approach to the study of symmetric dynamics by R. Spekkens 1273

List of Symbols 1275

References 1282

Index 1290

Introduction

Goals

This compendium originated out of the necessity for a compact summary of important theorems and formulas during physics and mathematics classes at university. When the interest in more (and more exotic) subjects grew, this collection lost its compactness and became the chaos it now is. Although there should exist some kind of overall structure, it was not always possible to keep every section self-contained or respect the order of the chapters.

It should definitely not be used as a formal introduction to any subject. It is neither a complete work nor a fact-checked one, so the usefulness and correctness is not guaranteed. However, it can be used as a look-up table for theorems and formulas, and as a guide to the literature. To this end, each chapter begins with a list of useful references. At the same time, only a small number of statements are proven in the text (or appendices). This was done to keep the text as concise as possible (a failed endeavour). However, in some cases the major ideas underlying the proofs are provided.

Structure and conventions

Sections and statements that require more advanced concepts, in particular concepts from later chapters or (higher) category theory, will be labelled by the *clubs* symbol ♣. Some definitions, properties or formulas are given with a proof or an extended explanation whenever I felt like it. These are always contained in a blue frame to make it clear that they are not part of the general compendium. When a section uses notions or results from a different chapter at its core, this will be recalled in a green box at the beginning of the section.

Definitions in the body of the text will be indicated by the use of **bold font**. Notions that have not been defined in this summary but that are relevant or that will be defined further on in the compendium (in which case a reference will be provided) are indicated by *italic text*. Names of authors are also written in *italic*.

Objects from a general category will be denoted by a lower-case letter (depending on the context, upper-case might be used for clarity), functors will be denoted by upper-case letters and the categories themselves will be denoted by symbols in **bold font**. In the later chapters on physics, specific conventions for the different types of vectors will often be adopted. Vectors in Euclidean space will be denoted by a bold font letter with an arrow above, e.g. $\vec{\mathbf{a}}$, whereas vectors in Minkowski space (4-vectors) and differential forms will be written without the arrow, e.g. $\mathbf{a}$. Matrices and tensors will always be represented by capital letters and, dependent on the context, a specific font will be adopted.

Part I

Set Theory, Algebra & Category Theory

Table of Contents

1 Set Theory **38**

1.1 Axiomatization . . . 39

1.1.1 ZFC . . . 39

1.1.2 Material set theory . . . 40

1.1.3 Structural set theory . . . 40

1.1.4 ETCS ♣ . . . 40

1.1.5 Universes . . . 41

1.2 Finite sets . . . 41

1.3 Real numbers . . . 41

1.4 Set operations . . . 42

1.5 Functions . . . 43

1.5.1 (Co)domain . . . 43

1.5.2 Functions . . . 43

1.6 Collections . . . 44

1.6.1 Families and filters . . . 44

1.6.2 Algebra of sets . . . 45

1.7 Ordered sets . . . 46

1.7.1 Posets . . . 46

1.7.2 Ordinals ♣ . . . 48

1.7.3 Cardinals ♣ . . . 49

1.7.4 Lattices . . . 51

1.8 Limits . . . 53

1.9 Partitions . . . 54

1.9.1 Partition . . . 54

1.9.2 Superpartition ♣ . . . 55

2 Algebra **57**

2.1 Algebraic structures . . . 58

2.2 Group theory . . . 59

2.2.1 Abelian groups . . . 60

2.2.2 Cosets . . . 60

2.2.3 Sylow theorems . . . 61
2.2.4 Symmetric groups . . . 62
2.2.5 Group presentations . . . 63
2.2.6 Direct products . . . 64
2.2.7 Free groups . . . 66
2.3 Group actions . . . 67
2.4 Rings . . . 70
2.4.1 General . . . 70
2.4.2 Ideals . . . 73
2.4.3 Algebraic geometry . . . 75
2.4.4 Modules . . . 77
2.4.5 Semisimplicity . . . 79
2.5 Galois theory ♣ . . . 79
2.5.1 Polynomials . . . 79
2.5.2 Field extensions . . . 81
2.5.3 Algebraicity . . . 81
2.5.4 Galois extensions . . . 83
2.5.5 Radicals . . . 84
2.5.6 Nullstellensatz . . . 84

3 Category theory **85**
3.1 Categories . . . 86
3.2 Functors . . . 87
3.2.1 Natural transformations . . . 88
3.2.2 Equivalences . . . 90
3.2.3 Stuff, structure and property . . . 91
3.2.4 Adjunctions . . . 91
3.3 General constructions . . . 93
3.3.1 Fibred categories ♣ . . . 94
3.3.2 Monads . . . 96
3.4 Morphisms and diagrams . . . 99
3.4.1 Morphisms . . . 99
3.4.2 Initial and terminal objects . . . 100
3.4.3 Lifts . . . 101
3.4.4 Limits and colimits . . . 103
3.5 Internal structures . . . 111
3.5.1 Groupoids . . . 113
3.6 Lawvere theories ♣ . . . 114
3.7 Operad theory ♣ . . . 115
3.7.1 Operads . . . 115

3.7.2 Algebraic topology . . . 116

4 Higher-dimensional Algebra ♣ **117**
4.1 Monoidal categories . . . 118
4.1.1 Braided categories . . . 120
4.1.2 Monoidal functors . . . 120
4.1.3 Closed categories . . . 122
4.2 Enriched category theory . . . 123
4.2.1 Enriched constructions . . . 125
4.2.2 Weighted (co)limits . . . 128
4.3 Abelian categories . . . 130
4.3.1 Additive categories . . . 130
4.3.2 Exact functors . . . 132
4.3.3 Finiteness . . . 133
4.4 Duality . . . 135
4.5 Tensor and fusion categories . . . 138
4.6 Ribbon and modular categories . . . 140
4.7 Module categories . . . 141
4.8 Higher vector spaces . . . 142
4.8.1 Kapranov–Voevodksy 2-vector spaces . . . 142
4.8.2 Baez–Crans 2-vector spaces . . . 142
4.9 Higher Lie theory . . . 143
4.9.1 Lie superalgebras . . . 143
4.9.2 Lie n-algebras . . . 145
4.10 Monoidal n-categories . . . 148
4.10.1 Relation with group cohomology . . . 149

5 Homological algebra **151**
5.1 Chain complexes . . . 151
5.2 Exact sequences . . . 154
5.3 Resolutions . . . 155
5.4 Derived functors . . . 156
5.4.1 Modules . . . 158
5.5 Spectral sequences . . . 160
5.5.1 Exact couples . . . 160
5.5.2 Filtered complexes . . . 162
5.5.3 Convergence . . . 163
5.6 Examples . . . 164
5.6.1 Group cohomology . . . 164
5.6.2 Hochschild and cyclic homology ♣ . . . 166

Chapter 1

Set Theory

Chapter Contents

1.1 Axiomatization **39**
1.1.1 ZFC 39
1.1.2 Material set theory 40
1.1.3 Structural set theory 40
1.1.4 ETCS ♣ 40
1.1.5 Universes 41
1.2 Finite sets **41**
1.3 Real numbers **41**
1.4 Set operations **42**
1.5 Functions **43**
1.5.1 (Co)domain 43
1.5.2 Functions 43
1.6 Collections **44**
1.6.1 Families and filters 44
1.6.2 Algebra of sets 45
1.7 Ordered sets **46**
1.7.1 Posets 46
1.7.2 Ordinals ♣ 48
1.7.3 Cardinals ♣ 49
1.7.4 Lattices 51
1.8 Limits **53**
1.9 Partitions **54**

1.9.1 Partition 54
1.9.2 Superpartition ♣ 55

1.1 Axiomatization

1.1.1 ZFC

The following set of axioms and axiom schemata gives a foundation for axiomatic set theory whilst fixing a number of issues in naive set theory, where the notion of a set is taken for granted. This theory is called **Zermelo–Frenkel** set theory (ZF). When extended with the axiom of choice (see further on), it is called ZFC, where the C stand for 'choice'.

Axiom 1.1 (Power set).

$$\forall x : \exists y : \forall z\big(z \in y \iff \forall w(w \in z \implies w \in x)\big) \tag{1.1}$$

The set y is called the power set $P(x)$ or 2^x of x.

Axiom 1.2 (Extensionality).

$$\forall x, y : \forall z\big(z \in x \iff z \in y\big) \implies x = y \tag{1.2}$$

This axiom allows to compare two sets based on their elements, i.e. two sets are (strictly) equal if and only if they have the same elements.

Remark 1.1.1. As will be made clear throughout this document, strict equality as in the extensionality axiom is often not what one cares about in practice. *Equivalence* and *isomorphism* will be the practically relevant notions.

Axiom 1.3 (Regularity[1]).

$$\forall x : (\exists z \in x) \implies (\exists a \in x) \wedge \neg(\exists b \in a : b \in x) \tag{1.3}$$

This axiom says that for every nonempty set x, one can find an element $a \in x$ such that x and a are disjoint. Among other things, this axiom implies that no set can contain itself.

The following axiom is technically not an axiom but an axiom schema, i.e. for every predicate φ, one obtains an axiom.

Axiom 1.4 (Specification).

$$\forall w_1, \dots, w_n, A : \exists B : \forall x\big[x \in B \iff \big(x \in A \wedge \varphi(x, w_1, \dots, w_n, A)\big)\big] \tag{1.4}$$

[1] Also called the **axiom of foundation**.

This axiom (schema) says that, for every set x, one can build another set of elements in x that satisfy a given predicate. By the axiom of extensionality, this subset $B \subseteq A$ is unique.

Notation 1.1.2 (Set builder notation). The sets constructed through the axiom of specification are often denoted as follows:

$$B = \{x \in A \mid \phi(x)\}\,, \tag{1.5}$$

where the parameters $w_1, \dots, w_n$ were suppressed for conciseness.

1.1.2 Material set theory

ZF(C) is an instance of *material set theory*. Every element of a set is itself a set and, hence, possesses some kind of internal structure.

Definition 1.1.3 (Pure set). A set U such that, for every sequence

$$x_n \in x_{n-1} \in \cdots \in x_1 \in U\,, \tag{1.6}$$

all elements x_i are also sets.

Definition 1.1.4 (Urelement[2]). An object that is not a set.

@@ COMPLETE @@

1.1.3 Structural set theory

In contrast to material set theory, the fundamental notions in this theory are the relations between sets. An element of a set does not have any internal structure and only becomes relevant if one specifies extra structure (given by relations) on the sets. This implies that elements of sets are not sets themselves. In fact, this would be a meaningless statement since, by default, they lack any internal structure. Even stronger, it is meaningless to compare two elements if one does not provide relations or extra structure on the sets.

@@ COMPLETE @@

1.1.4 ETCS ♣

Remark. ETCS is the abbreviation of 'Elementary Theory of the Category of Sets'.

Axiom 1.5. The category of sets is a *well-pointed* (*elementary*) *topos* (see Chapter 12).

@@ COMPLETE @@

[2]Sometimes called an **atom**.

1.1.5 Universes

To be able to talk about sets without running into problems such as *Russel's paradox*, where one needs (or wants) to talk about the collection of all things satisfying a certain condition, one can introduce the concept of a 'universal set' or 'universe (of discourse)'. This set takes the place of the 'collection of things' and all operations act within this universe.

Definition 1.1.5 (Grothendieck universe). A pure set U satisfying the following conditions:

1. **Transitivity**: If $x \in U$ and $y \in x$, then $y \in U$.
2. **Power set**: If $x \in U$, then $P(x) \in U$.
3. **Pairing**: If $x, y \in U$, then $\{x, y\} \in U$.
4. **Unions**: If $I \in U$ and $\{x_i\}_{i \in I} \subset U$, then $\bigcup_{i \in I} x_i \in U$.

1.2 Finite sets

Definition 1.2.1 (Bishop finiteness). A set S is said to be (Bishop) finite if there exists a bijection $S \cong [n]$ for some $n \in \mathbb{N}$.

In ZFC, any other reasonable definition of finite sets will be equivalent to the one above. However, in constructive mathematics, this is not true anymore. Because the *internal logic* of *elementary topoi* (see Chapter 12) is constructive, some more information is given here.

Definition 1.2.2 (Kuratowski finiteness). A set S is said to be Kuratowski finite or **finitely indexed set** if there exists a surjection $[n] \twoheadrightarrow S$ for some $n \in \mathbb{N}$.

Definition 1.2.3 (Dedekind finiteness). A set S is said to be Dedekind finite if any injection $S \hookrightarrow S$ is a bijection.

1.3 Real numbers

Axiom 1.6 (Ordering). The set of real numbers is an *ordered field* (see Definition 2.4.2).

Axiom 1.7 (Dedekind completeness). Every nonempty subset of $\mathbb{R}$ that is bounded from above has a *supremum* (see Definition 1.7.9 and Definition 1.7.38).

Axiom 1.8. The rational numbers form a subset of the real numbers: $\mathbb{Q} \subset \mathbb{R}$.

Remark 1.3.1. There is only one way (up to isomorphisms) to extend the field of rational numbers to the field of reals such that it satisfies the previous axioms.

Definition 1.3.2 (Extended real line).

$$\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\} \equiv [-\infty, \infty] \tag{1.7}$$

1.4 Set operations

Definition 1.4.1 (Cartesian product). Let X, Y be two sets. Their (Cartesian) product is defined as follows:

$$X \times Y := \{(x, y) \mid x \in X \wedge y \in Y\}. \tag{1.8}$$

Definition 1.4.2 (Symmetric difference).

$$A \Delta B := (A \backslash B) \cup (B \backslash A) \tag{1.9}$$

Definition 1.4.3 (Diagonal). The diagonal of a set X is defined as follows:

$$\Delta_X := \left\{(x, x) \in X \times X \mid x \in X\right\}. \tag{1.10}$$

Definition 1.4.4 (Complement). Let Ω be the universe of discours (Section 1.1.5) and consider $X \subseteq \Omega$. The complement of X is defined as follows:

$$X^c := \Omega \backslash X. \tag{1.11}$$

Formula 1.4.5 (de Morgan's laws).

$$\left(\bigcup_i X_i\right)^c = \bigcap_i X_i^c \tag{1.12}$$

$$\left(\bigcap_i X_i\right)^c = \bigcup_i X_i^c \tag{1.13}$$

Definition 1.4.6 (Relation). A relation between two sets X and Y is a subset of the Cartesian product $X \times Y$. A relation on X is then simply a subset of $X \times X$. This definition can easily be extended to n-ary relations by working with subsets of n-fold products.

Definition 1.4.7 (Converse relation). Consider a relation $R \subseteq X \times Y$. The converse relation R^t is defined as follows:

$$R^t := \left\{(y, x) \in Y \times X \mid (x, y) \in R\right\}. \tag{1.14}$$

Definition 1.4.8 (Composition of relations). Consider two relations $R \subseteq X \times Y$ and $S \subseteq Y \times Z$ between three sets X, Y and Z. The composition $S \circ R$ is defined as follows:

$$S \circ R := \left\{(x, z) \in X \times Z \mid \exists y \in Y : (x, y) \in R \wedge (y, z) \in S\right\}. \tag{1.15}$$

1.5 Functions

1.5.1 (Co)domain

Definition 1.5.1 (Domain). Let $f : X \to Y$ be a function. The set X is called the domain of f.

Notation 1.5.2. The domain of f is denoted by $\mathsf{dom}(f)$.

Definition 1.5.3 (Support). Let $f : X \to \mathbb{R}$ be a function. The support of f is defined as the set of points where f is nonzero.

Notation 1.5.4. The support of f is denoted by $\mathsf{supp}(f)$.

Notation 1.5.5. Let X, Y be two sets. The set of functions $f : X \to Y$ is denoted by Y^X or $\mathsf{Map}(X, Y)$. (See also Definition 4.1.22 for a generalization.)

Definition 1.5.6 (Codomain). Let $f : X \to Y$ be a function. The set Y is called the codomain of f.

Definition 1.5.7 (Image). Let $f : X \to Y$ be a function. The following subset of Y is called the image of f:

$$\mathsf{im}(f) := \{y \in Y \mid \exists x \in X : f(x) = y\}\,. \tag{1.16}$$

Remark. Some authors use the notions of codomain and image interchangeably.

Definition 1.5.8 (Level set). Consider a function $f : X \to \mathbb{R}$. The following set is called the level set of f at $c \in \mathbb{R}$:

$$f^{-1}(c) := \{x \in X \mid f(x) = c\}\,. \tag{1.17}$$

For $X = \mathbb{R}^2$ the level sets are called **level curves** and for $X = \mathbb{R}^3$ they are called **level surfaces**. More generally, for functions $f : X \to Y$, one calls $f^{-1}(y)$ the **preimage** or **fibre** of f over y.

1.5.2 Functions

Definition 1.5.9 (Injective). A function $f : A \to B$ is said to be injective or **one-to-one** if the following condition is satisfied:

$$\forall a, a' \in A : f(a) = f(a') \implies a = a'\,. \tag{1.18}$$

Notation 1.5.10 (Injective function).

$$f : A \hookrightarrow B$$

Definition 1.5.11 (Surjective). A function $f : A \to B$ is said to be surjective or **onto** if the following condition is satisfied:

$$\forall b \in B : \exists a \in A : f(a) = b\,. \tag{1.19}$$

Notation 1.5.12 (Surjective function).

$$f : A \twoheadrightarrow B$$

Definition 1.5.13 (Bijection). A function that has an inverse. Equivalently, a function that gives a one-to-one correspondence between the elements of the domain and those of the codomain.

Notation 1.5.14 (Isomorphic sets).

$$X \cong Y$$

Theorem 1.5.15 (Cantor–Bernstein–Schröder). *Consider two sets A, B. If there exist injections $A \hookrightarrow B$ and $B \hookrightarrow A$, there exists a bijection $A \cong B$.*

Definition 1.5.16 (Involution). A function $f : A \to A$ such that $f^2 = \mathbb{1}_A$, i.e. f is its own inverse. Every involution is, in particular, a bijection.

1.6 Collections

1.6.1 Families and filters

Definition 1.6.1 (Power set). Let X be a set. The power set is defined as the set of all subsets of X and is (often) denoted by $P(X)$ or 2^X. The existence of this set is enforced by the power set axiom (Axiom 1.1).

Corollary 1.6.2. All sets are elements of their power set: $X \in P(X)$.

Definition 1.6.3 (Collection). A collection of elements of a set X is a subset of X. Often, this collection is indexed by some **index set** I, in which case it is often denoted as $\{x_i\}_{i \in I} \subseteq X$.

Definition 1.6.4 (Family). Let X, I be two sets. A family of elements of X with **index set** I is a function $f : I \to X$. A family with index set I is often denoted by $(x_i)_{i \in I}$. In contrast to collections, a family can contain multiple copies of the same element.

Definition 1.6.5 (Cover). A cover of a set X is a collection of sets $\{V_i\}_{i \in I} \subseteq P(X)$ such that:

$$\bigcup_{i \in I} V_i = X\,. \tag{1.20}$$

Definition 1.6.6 (Partition). A partition of X is a collection of disjoint subsets $\{A_i\}_{i \in I} \subset P(X)$ such that $\bigcup_{i \in I} A_i = X$, i.e. it is a disjoint cover.

Definition 1.6.7 (Refinement). Let P be a partition of X. A refinement P' of P is a collection of subsets such that every $A \in P$ can be written as a disjoint union of elements in P'. It follows that every refinement is also a partition.

Definition 1.6.8 (Filter). Let X be a set. A family $\mathcal{F} \subseteq P(X)$ is a filter on X if it satisfies the following conditions:

1. **Nonemptiness**: $\emptyset \notin \mathcal{F}$.
2. **Downward closure**: $\forall A, B \in \mathcal{F} : A \cap B \in \mathcal{F}$.
3. **Isotony**: if $A \in \mathcal{F}$ and $A \subseteq B$, then $B \in \mathcal{F}$.

Definition 1.6.9 (Filtration). Consider a set X together with a collection of subsets $\{F_i\}_{i \in I}$ indexed by a *totally ordered set* I (see Definition 1.7.5). The collection is said to be a filtration of X if

$$i \leq j \implies F_i \subseteq F_j . \tag{1.21}$$

A filtration is said to be **exhaustive** if $\bigcup_{i \in I} F_i = X$ and **separated** if $\bigcap_{i \in I} F_i = \emptyset$.

Definition 1.6.10 (Associated grading). When one can define *quotient objects*, every filtration $\{F_i\}_{i \in I}$ of X defines an associated graded object $\{G_i := F_i / F_{i-1}\}_{i \in I}$.

1.6.2 Algebra of sets

Definition 1.6.11 (Algebra of sets). A collection $\mathcal{F} \subset P(X)$ is a called an algebra over X if it satisfies the following conditions:

1. **Totality**: $X \in \mathcal{F}$,
2. **Closed under complements**: $\forall E \in \mathcal{F} : E^c \in \mathcal{F}$, and
3. **Closed under finite unions**: $\forall E, E' \in \mathcal{F} : E \cup E' \in \mathcal{F}$.

The pair $(X, \mathcal{F})$ is also called a **field of sets** or **concrete Boolean algebra**.

Extending this definition to closure under countable unions gives another common notion.

Definition 1.6.12 (σ-algebra). A collection $\Sigma \subset P(X)$ is called a σ-algebra over a set X if it satisfies the following conditions:

1. **Totality**: $X \in \Sigma$,
2. **Closed under complements**: $\forall E \in \Sigma : E^c \in \Sigma$, and
3. **Closed under countable unions**: $\forall \{E_n\}_{n \in \mathbb{N}} \subset \Sigma : \bigcup_{n=1}^{+\infty} E_n \in \Sigma$.

The pair (X, Σ) is called a **measurable space**. The elements $E \in \Sigma$ are called **measurable sets**.

Remark 1.6.13. Conditions 1 and 2 imply that both σ-algebra and fields of sets always contain the empty set. Conditions 2 and 3, together with de Morgan's laws (1.12)

and (1.13), imply that a σ-algebra is also closed under countable intersections and, analogously, that a field of sets is closed under finite intersections.

Property 1.6.14 (Intersections). The intersection of a collection of σ-algebras is again a σ-algebra.

Definition 1.6.15 (Generated σ-algebras). A σ-algebra $\mathcal{G}$ is said to be generated by a collection of sets $\mathcal{A}$ if

$$\mathcal{G} = \bigcap \{\mathcal{F} \mid \mathcal{F} \text{ is a } \sigma\text{-algebra that contains } \mathcal{A}\}. \tag{1.22}$$

Equivalently, it is the smallest σ-algebra containing $\mathcal{A}$.

Notation 1.6.16. The σ-algebra generated by a collection of sets $\mathcal{A}$ is often denoted by $\sigma(\mathcal{A})$.

Definition 1.6.17 (Product σ-algebras). The product σ-algebra $\mathcal{F}_1 \otimes \mathcal{F}_2$ on $X_1 \times X_2$ can be defined in the following equivalent ways:

- $\mathcal{F}$ is generated by $\{A_1 \times \Omega_2 \mid A_1 \in \mathcal{F}_1\} \cup \{\Omega_1 \times A_2 \mid A_2 \in \mathcal{F}_2\}$.
- $\mathcal{F}$ is generated by $\{A_1 \times A_2 \mid A_1 \in \mathcal{F}_1 \wedge A_2 \in \mathcal{F}_2\}$.

Definition 1.6.18 (Monotone class). A collection of sets $\mathcal{A}$ that has the following two properties:

1. For every increasing sequence $A_1 \subset A_2 \subset \cdots : \bigcup_{n=1}^{+\infty} A_n \in \mathcal{A}$.
2. For every decreasing sequence $A_1 \supset A_2 \supset \cdots : \bigcap_{n=1}^{+\infty} A_n \in \mathcal{A}$.

Theorem 1.6.19 (Monotone class theorem). *Let $\mathcal{A}$ be an algebra of sets. If $\mathcal{G}_{\mathcal{A}}$ is the smallest monotone class containing $\mathcal{A}$, it coincides with the σ-algebra generated by $\mathcal{A}$.*

1.7 Ordered sets

1.7.1 Posets

Definition 1.7.1 (Partially ordered set). A set P equipped with a binary relation $\leq$ is called a partially ordered set (or **poset**) if the following 3 conditions are satisfied for all elements $x, y, z \in P$:

1. **Reflexivity**: $x \leq x$,
2. **Antisymmetry**: $x \leq y \wedge y \leq x \implies x = y$, and
3. **Transitivity**: $x \leq y \wedge y \leq z \implies x \leq z$.

If the antisymmetry condition is dropped, a **preorder** is obtained.

Definition 1.7.2 (Order-preserving map). A function $f : (P, \leq) \to (P', \preceq)$ of posets such that $x \leq y \implies f(x) \preceq f(y)$ for all $x, y \in P$. These functions are also said to be **monotone**.

Definition 1.7.3 (Hasse diagram). Consider a finite poset $(P, \leq)$. A Hasse diagram for P is a *graph* where the vertices are given by the elements of P and two vertices x, y are connected upwardly if and only if $x \leq y$ in P.

Example 1.7.4. Consider the power set of $\{0, 1\}$ with its canonical poset structure. A Hasse diagram of this poset is

$$\begin{tikzcd}[column sep=small] & \{0,1\} \arrow[dl, dash] \arrow[dr, dash] & \\ \{0\} \arrow[dr, dash] & & \{1\}. \arrow[dl, dash] \\ & \emptyset & \end{tikzcd}$$

Definition 1.7.5 (Totally ordered set). A poset P with the property that for all $x, y \in P$, either $x \leq y$ or $y \leq x$. This property is called **totality**.

Definition 1.7.6 (Strict total order). A nonstrict order $\leq$ has an associated strict order $<$ that satisfies $x < y \iff x \leq y \wedge x \neq y$.

Definition 1.7.7 (Linear order). A binary relation $<$ on a set P satisfying the following conditions for all $x, y, z \in P$:

1. **Irreflexivity**: $x \not< x$,
2. **Asymmetry**: $x < y \implies y \not< x$,
3. **Transitivity**: $x < y \wedge y < z \implies x < z$,
4. **Comparison**: $x < z \implies x < y \vee y < z$, and
5. **Connectedness**: $x \not< y \wedge y \not< x \implies x = y$.

Remark 1.7.8. By negation, one can freely pass between linear orders and total orders. However, without the law of the excluded middle, there exists no bijection between them.

Definition 1.7.9 (Supremum). The supremum $\sup(P)$ of a poset P is the least upper bound of P.

Definition 1.7.10 (Infimum). The infimum $\inf(P)$ of a poset P is the greatest lower bound of P.

Definition 1.7.11 (Maximum). If $\sup(P) \in P$, the supremum is called the maximum of P. This is denoted by $\max(P)$.

Definition 1.7.12 (Minimum). If $\inf(P) \in P$, the supremum is called the minimum of P. This is denoted by $\min(P)$.

Definition 1.7.13 (Chain). A totally ordered subset of a poset.

Theorem 1.7.14 (Zorn's lemma[3]). *Let $(P, \leq)$ be a poset. If every chain in P has an upper bound in P, then P has a maximal element.*

Definition 1.7.15 (Directed[4] set). A set P equipped with a preorder $\leq$ with the additional property that every 2-element subset has an upper bound, i.e. for every two elements $x, y \in P$, there exists an element $z \in P$ such that $x \leq z \wedge y \leq z$.

Definition 1.7.16 (Net). A net on a set X is a subset of X indexed by a directed set I.

Definition 1.7.17 (Cut). Let $(P, \leq)$ be a totally ordered set. A cut or **decomposition** of P is a pair (A, B) of disjoint subsets such that $P = A \cup B$ in the ordered sense, i.e. every element of A is smaller than every element of B. Cuts can be classified as follows:

- **Jumps**: A has a greatest element and B has a least element.
- **Dedekind cut**: Either A has a greatest element and B has no least element or A has no greatest element, but B has a least element.
- **Gap**: A has no greatest element and B has no least element.

The notion of a filter on a set X (Definition 1.6.8) is a specific example of a more general notion, where the underlying poset is the power set of X.

Definition 1.7.18 (Filter). Let X be a poset. A filter on X is a subset $\mathcal{F} \subseteq X$ such that:

1. **Nonemptiness**: $\mathcal{F} \neq \emptyset$,
2. **Downward closure**: $\forall x, y \in \mathcal{F} : \exists z \in \mathcal{F} : z \leq x \wedge z \leq y$, and
3. **Isotony**: If $x \in \mathcal{F}$ and $x \leq y$, then $y \in \mathcal{F}$.

If there exists no strictly greater filter $\mathcal{F} \subset \mathcal{F}'$, then $\mathcal{F}$ is called an **ultrafilter**.

The following theorem is independent of the ZF axioms, but strictly weaker than the *axiom of choice*.

Theorem 1.7.19 (Ultrafilter lemma). *Every proper filter on a set is contained in an ultrafilter.*

1.7.2 Ordinals ♣

Definition 1.7.20 (Well-ordering). A **well-founded** linear order, i.e. a linear order such that every nonempty subset has a minimal element.

[3]This theorem is equivalent to the *axiom of choice*.

[4]Sometimes called a **filtered** set or **upward** directed set. **Downward** directed sets are analogously defined with a lower bound for every two elements.

Definition 1.7.21 (Ordinal number). Consider the class of all well-ordered sets. An ordinal (type or rank) is an isomorphism class of well-ordered sets. The class of ordinals is itself well ordered by inclusion of 'initial segments'.

However, this definition gives problems within the ZF(C) framework of set theory, since these equivalence classes are proper classes and not sets. To overcome this problem, one can use a different approach. By using a well-defined construction, one can, for every class, select a particular representative and call this representative the **ordinal number** of all well-ordered sets isomorphic to it.

The most common construction is the one by *von Neumann*. For every well-ordered set W, there exists a function

$$W \to P(W) : w \mapsto W_{\leq w} := \{w' \in W \mid w' \leq w\} \tag{1.23}$$

that restricts to an order isomorphism on its image. This leads to the following definition.

Definition 1.7.22 (von Neumann ordinal). A set that is strictly well ordered by membership and such that every element is also a subset.

The first finite von Neumann ordinals are given as an example:

- $0 := \emptyset$,
- $1 := \{0\} = \{\emptyset\}$,
- $2 := \{0, 1\} = \{\emptyset, \{\emptyset\}\}$, and so on.

Property 1.7.23. Every ordinal type is uniquely order-isomorphic to an ordinal number. Consequently, every order-preserving isomorphism between an order type and itself is the identity.

Definition 1.7.24 (Successor). Every ordinal number α has a **successor** α^+. Using the von Neumann definition, this is simply $\alpha^+ := \alpha \cup \{\alpha\}$. An ordinal that is not the successor of another ordinal number is called a **limit ordinal**.

Definition 1.7.25 (Regular ordinal). A limit ordinal α that is not the limit of a set of smaller ordinals with order type less than α.

Property 1.7.26 (Burali–Forti paradox). The class of all ordinals (and, by extension, the class of all well-ordered sets) is not a set.

1.7.3 Cardinals ♣

There also exist numbers representing the sizes of sets, the **cardinal numbers**. These 'numbers' should satisfy the following conditions:

1. Every set has a well-defined cardinality.
2. Every cardinal number is the cardinality of some set.
3. Bijective sets have the same cardinality.

Guided by these conditions, one could naively use the following definition.

Definition 1.7.27 (Cardinal number[5]). An isomorphism class of sets (under bijections).

However, similar to the problem encountered for ordinals above, these classes are not sets. To solve this, one can also use a similar trick and select a specific representative. For cardinals, the following choice is common.

Alternative Definition 1.7.28 (Cardinal number). The cardinal number of a set is the smallest ordinal rank of any well-order on it, i.e. any ordinal number bijective to it.[6] The cardinal numbers inherit a well-ordering from the ordinal numbers.

Remark 1.7.29 (Ordering). The Cantor–Bernstein–Schröder theorem 1.5.15 induces a partial ordering on cardinal numbers. However, without the axiom of choice this can never be a total ordering. This problem is also apparent in the above definition since the ordinal rank of sets is used together with the *well-ordering theorem*, which is equivalent to the axiom of choice.

Similar to ordinal numbers, one can also define successors of cardinal numbers.

Definition 1.7.30 (Successor). Given a cardinal κ, its successor κ^+ is defined as the smallest cardinal greater than κ.

Remark 1.7.31. It should be noted that the successor of a cardinal number is not necessarily the same as its successor as an ordinal number (in fact, this is only the case for finite cardinals).

Definition 1.7.32 (Regular cardinal). An infinite cardinal κ such that there exists no set of cardinality κ that is the union of less than κ subsets of cardinality less than κ:

$$\left(\kappa = \sum_{i \in I} \lambda_i\right) \wedge \left(\forall i \in I : \lambda_i < \kappa\right) \implies |I| \geq \kappa. \tag{1.24}$$

The following theorem can easily be proven by a diagonal argument.

Theorem 1.7.33 (Cantor). *Let S be a set of cardinality κ. The power set $P(S)$ has cardinality strictly greater than κ.*

[5] Also called the **cardinality** of a set.

[6] The *well-ordering theorem* (if assumed) assures that this definition coincides with the naive one above.

1.7.4 Lattices

Definition 1.7.34 (Semilattice). A poset $(P, \leq)$ for which every 2-element subset has a supremum, called the **join**, is called a **join-semilattice**. Similarly, a poset $(P, \leq)$ for which every 2-element subset has an infimum, called the **meet**, is called a **meet-semilattice**.

Notation 1.7.35. The join of $\{x, y\}$ is denoted by $x \wedge y$. The meet of $\{x, y\}$ is denoted by $x \vee y$.

Definition 1.7.36 (Lattice). A poset that is both a join- and a meet-semilattice.

The above definition also allows for a purely algebraic formulation (in this case, some authors might speak about **lattice-ordered sets**).

Alternative Definition 1.7.37 (Lattice). A lattice is an algebraic structure that admits operations $\wedge, \vee$ that satisfy the following conditions:

1. Both $\wedge$ and $\vee$ are idempotent, commutative and associative.
2. The operations satisfy the **absorption laws**:

$$x \vee (x \wedge y) = x \qquad\qquad x \wedge (x \vee y) = x. \tag{1.25}$$

To go from this definition to the order-theoretic one, define the partial order

$$x \leq y \iff x \wedge y = x\,. \tag{1.26}$$

There exists an equivalent relation for the join.

Definition 1.7.38 (Complete lattice). A lattice is said to be σ-complete (resp. complete) if it admits all[7] countable (resp. all) joins. It can be proven that complete lattices also admit all meets (as a special case of the *Adjoint Functor Theorem* 13.5.10). If only bounded (from above) subsets admit a join, the lattice is said to be **Dedekind-complete**.

Definition 1.7.39 (Bounded lattice). A lattice $(L, \leq, \wedge, \vee)$ that contains a greatest element (denoted by $\top$ or 1) and a smallest element (denoted by $\bot$ or 0) such that

$$\bot \leq x \leq \top \tag{1.27}$$

for all $x \in P$. These elements are the identities for the join and meet operations:

$$x \wedge \top = x \qquad\qquad x \vee \bot = x\,. \tag{1.28}$$

[7]When working with *categories* (see Chapter 3), this has to be restricted to 'all small joins/meets' or, equivalently, the index category should be a set.

Definition 1.7.40 (Frame). A complete lattice $(L, \leq, \wedge, \vee)$ for which the **infinite distributivity law** is satisfied:

$$y \wedge \left(\bigvee_{i \in I} x_i \right) = \bigvee_{i \in I} (y \wedge x_i) \,. \tag{1.29}$$

Definition 1.7.41 (Distributive lattice). A lattice $(L, \leq, \wedge, \vee)$ such that

$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c) \tag{1.30}$$

and[8]

$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c) \tag{1.31}$$

for all $a, b, c \in L$. If only

$$a \leq c \implies a \vee (b \wedge c) = (a \vee b) \wedge c \tag{1.32}$$

holds, the lattice is said to be **modular**.

Definition 1.7.42 (Complemented lattice). A bounded lattice $(L, \leq, \wedge, \vee, \top, \bot)$ such that for every $a \in L$, there exists at least one $b \in L$ such that

$$a \wedge b = \bot \qquad \text{and} \qquad a \vee b = \top \,. \tag{1.33}$$

If a consistent assignment of **(ortho)complements** exists, i.e. for all $a, b \in L$:

1. **Involutivity**: $a^{\bot\bot} = a$, and
2. **Order-reversing**: $a \leq b \iff b^{\bot} \leq a^{\bot}$,

the lattice is said to admit an **orthocomplementation**. When, furthermore,

$$a \leq b \implies a \vee (a^{\bot} \wedge b) = b \tag{1.34}$$

for all $a, b \in L$, then L is said to be **orthomodular**. (Comparing to the previous definition, this shows that orthomodularity is an even weaker form of distributivity than modularity, where distributivity only holds with respect to complements.)

Property 1.7.43 (de Morgan). All (ortho)complemented lattices satisfy de Morgan's laws (1.12) and (1.13).

Definition 1.7.44 (Heyting algebra). A bounded lattice $(L, \leq, \wedge, \vee, \top, \bot)$ such that, for every two elements $a, b \in L$, there exists a greatest element $x \in L$ such that

$$a \wedge x \leq b \,. \tag{1.35}$$

[8] This second condition is actually a consequence of the first.

This element is denoted by $a \to b$. The **pseudocomplement** $\neg a$ of an element $a \in L$ is then defined as $a \to \bot$. Note that pseudocomplements do not define an orthocomplementation since

$$a \vee a^{\perp} = \top \tag{1.36}$$

does not have to hold.

Definition 1.7.45 (Boolean algebra). A Heyting algebra L in which the **law of excluded middle** holds:

$$\forall a \in L : \neg\neg a = a\,. \tag{1.37}$$

This can be equivalently stated as

$$\forall a \in L : a \vee \neg a = \top\,. \tag{1.38}$$

This is equivalent to a complemented, distributive lattice.

1.8 Limits

Definition 1.8.1 (Direct system). Let $(I, \leq)$ be a directed set (Definition 1.7.15) and let $(A_i)_{i\in I}$ be a family of objects. Consider a family of morphisms $(f_{ij} : A_i \to A_j)_{i,j\in I}$ between these objects with the following properties:

1. for every $i \in I$: $f_{ii} = \mathbb{1}_{A_i}$, and
2. for every $i \leq j \leq k \in I$: $f_{ik} = f_{jk} \circ f_{ij}$.

The pair (A_i, f_{ij}) is called a direct system (over I).

Definition 1.8.2 (Direct limit). Consider a direct system (A_i, f_{ij}) over a directed set I. The direct (or **inductive**) limit A of this direct system is defined as follows:

$$\varinjlim A_i := \bigsqcup_{i\in I} A_i \Big/ \sim\,, \tag{1.39}$$

where the equivalence relation is given by

$$x \in A_i \sim y \in A_j \iff \exists k \in I : f_{ik}(x) = f_{jk}(y)\,. \tag{1.40}$$

Informally put: two elements are equivalent if they eventually become the same. The operations on A are defined such that the inclusion maps $\phi_i : A_i \to A$ are morphisms.

Definition 1.8.3 (Inverse system). Let $(I, \leq)$ be a directed set (Definition 1.7.15) and let $(A_i)_{i\in I}$ be a family of objects. Consider a family of morphisms $(f_{ij} : A_j \to A_i)_{i,j\in I}$ between these objects with the following properties:

1. for every $i \in I$: $f_{ii} = \mathbb{1}_{A_i}$, and

2. for every $i \leq j \leq k \in I$: $f_{ik} = f_{ij} \circ f_{jk}$.

The pair (A_i, f_{ij}) is called an inverse system (over I).

Definition 1.8.4 (Inverse limit). Consider an inverse system (A_i, f_{ij}) over a directed set I. The inverse (or **projective**) limit A of this inverse system is defined as follows:

$$\varprojlim A_i := \left\{ \vec{a} \in \prod_{i \in I} A_i \middle| a_i = f_{ij}(a_j), \forall i \leq j \right\}. \tag{1.41}$$

For all $k \in I$, there exists a natural projection $\pi_k : \varprojlim A_i \to A_k$.

Remark 1.8.5. The direct and inverse limit are each other's (*categorical*) dual. The former is a *colimit* while the latter is a *limit* in category theory. (See Section 3.4.4.)

1.9 Partitions

1.9.1 Partition

Definition 1.9.1 (Composition). Let $k, n \in \mathbb{N}$. A k-composition of n is a k-tuple $(t_1, \dots, t_k) \subset \mathbb{N}$ such that $\sum_{i=1}^{k} t_k = n$.

Definition 1.9.2 (Partition). Let $n \in \mathbb{N}$. A partition of n is an ordered composition of n (usually in decreasing order). Hence multiple different composition can determine the same partition.

Definition 1.9.3 (Young diagram[9]). A Young diagram is a visual representation of the partition of an integer n. It is a left justified system of boxes, where every row corresponds to an element of the partition:

Figure 1.1: A Young diagram representing the partition $(5, 4, 4, 1)$ of 14.

Definition 1.9.4 (Conjugate partition). Let λ be a partition of n with associated Young diagram $\mathcal{D}$. The conjugate partition λ' is obtained by reflecting $\mathcal{D}$ across the main diagonal. Since the number of boxes is left invariant, conjugate partitions are partitions for the same integer.

Example 1.9.5. Conjugating Diagram 1.1 gives Diagram 1.2 below. The associated partition is $(4, 3, 3, 3, 1)$.

[9] Sometimes called a **Ferrers diagram**.

Figure 1.2: A Young diagram representing the partition $(4,3,3,3,1)$ of 14.

Definition 1.9.6 (Young tableau). Consider a Young diagram of shape λ. A Young tableau of shape λ is a filling of the corresponding Young diagram by the elements of a totally ordered set (with n elements). This tableau is said to be **standard** if every row and every column is increasing.

Formula 1.9.7 (Hook length formula). The **hook** $H_{i,j}$ is defined as the part of a Young diagram given by the cell (i,j) together with all cells below and to the right of (i,j). Given a hook $H_{i,j}$, define the **hook length** $h_{i,j}$ as the cardinality of $H_{i,j}$.

The number $f^\lambda \in \mathbb{N}$ of possible standard Young tableaux of shape λ, where λ defines a partition of n, is given by the following formula:

$$f^\lambda = \frac{n!}{\prod_{(i,j)\in\lambda} h_{i,j}} \,. \tag{1.42}$$

Definition 1.9.8 (Young tabloid). A Young tabloid of shape λ is defined as an equivalence class of Young tableaux that are related by permuting the elements within a row. These are often drawn as in Fig. 1.3. Note that every Young tabloid is represented by exactly one Young tableau.

1	2	3	5	8
4	6	9	10	
7	11	12	14	
15				

Figure 1.3: A Young tabloid associated to the Young diagram in Figure 1.1.

1.9.2 Superpartition ♣

For the physical background of the notions introduced in this section, see Section 57.4.

Definition 1.9.9 (Superpartition). Let $m, n \in \mathbb{N}$. A superpartition in the m-**fermion sector** is a sequence of integers of the following form:

$$\Lambda = (\Lambda_1, \dots, \Lambda_m; \Lambda_{m+1}, \dots, \Lambda_n) \,, \tag{1.43}$$

where the first m numbers are strictly ordered, i.e. $\Lambda_i > \Lambda_{i+1}$ for all $i < m$, and the last $n - m$ numbers form a normal partition.

Both sequences, separated by a semicolon, form in fact distinct partitions themselves. The first one represents the **antisymmetric, fermionic sector** (this explains the strict order) and the second one represents the **symmetric, bosonic sector**. This amounts to the following notation:

$$\Lambda \equiv (\lambda^a; \lambda^s)\,.$$

The **degree** of the superpartition is given by $|\Lambda| := \sum_{i=1}^{n} \Lambda_i$.

Notation 1.9.10. A superpartition of degree $n \in \mathbb{N}$ in the m-fermion sector is said to be a superpartition of $(n|m)$. To every superpartition Λ, one can also associate a unique partition Λ^* by removing the semicolon and reordering the numbers such that they form a partition of n. The superpartition Λ can then be represented by the Young diagram belonging to Λ^*, where the rows belonging to the fermionic sector are terminated by a circle.

Chapter 2

Algebra

For the later sections on (co)homology theory, some terminology from Chapter 5 is used. Basic introductions to group theory can be found in Jeevanjee (2015); Wigner (2012).

Chapter Contents

2.1 Algebraic structures **58**

2.2 Group theory **59**

2.2.1 Abelian groups 60

2.2.2 Cosets 60

2.2.3 Sylow theorems 61

2.2.4 Symmetric groups 62

2.2.5 Group presentations 63

2.2.6 Direct products 64

2.2.7 Free groups 66

2.3 Group actions **67**

2.4 Rings **70**

2.4.1 General 70

2.4.2 Ideals 73

2.4.3 Algebraic geometry 75

2.4.4 Modules 77

2.4.5 Semisimplicity 79

2.5 Galois theory ♣ **79**

2.5.1 Polynomials 79

2.5.2 Field extensions 81

2.5.3 Algebraicity 81
2.5.4 Galois extensions 83
2.5.5 Radicals 84
2.5.6 Nullstellensatz 84

2.1 Algebraic structures

Definition 2.1.1 (Semigroup). A set G equipped with a binary operation $\star$ such that the following axioms are satisfied:

1. **Closure**: G is closed under $\star$.
2. **Associativity**: $\star$ is associative.

If the associativity axiom is dropped, a **magma** is obtained.

Definition 2.1.2 (Monoid). A set M equipped with a binary operation $\star$ such that the following axioms are satisfied:

1. **Closure**: M is closed under $\star$.
2. **Associativity**: $\star$ is associative.
3. **Unitality**: M has an identity element e with respect to $\star$.

Definition 2.1.3 (Nilpotent). An element x of a monoid $(M, \star, e)$ for which there exists an integer $k \in \mathbb{N}$ such that $x^k = e$, where $x^k := \underbrace{x \star \cdots \star x}_{k \text{ times}}$.

Property 2.1.4 (Eckmann–Hilton argument). Let $(M, \circ), (M, \otimes)$ be two monoid structures (or even unital magma structures) on the same set M such that

$$(a \circ b) \otimes (c \circ d) = (a \otimes c) \circ (b \otimes d) \tag{2.1}$$

for all $a, b, c, d \in M$. The two monoid structures coincide and are in fact **Abelian**, i.e. they are **commutative**:

$$a \circ b = b \circ a \tag{2.2}$$

for all $a, b \in M$.

Remark 2.1.5. This property admits a vast generalization. See Property 3.5.1 for more information.

Definition 2.1.6 (Group). A set G equipped with a binary operation $\star$ such that the following axioms are satisfied:

1. **Closure**: G is closed under $\star$.

2. **Associativity**: $\star$ is associative.

3. **Unitality**: G has an identity element e with respect to $\star$.

4. **Inverses**: Every element in G has an inverse with respect to $\star$.

Notation 2.1.7 (Identity element). Although the identity element of a group is generally denoted by e, in certain cases, whenever this makes sense, the identity element will be denoted by 0 or 1 (additive and multiplicative conventions).

Definition 2.1.8 (Abelian group). Let $(G, \star)$ be a group. If $\star$ is commutative, then G is said to be an Abelian or **commutative** group.

Definition 2.1.9 (Morphism). A group **(homo)morphism** $\Phi : (G, \star, e_G) \to (H, \cdot, e_H)$ is a function $\Phi : G \to H$ such that

$$\Phi(g \star g') = \Phi(g) \cdot \Phi(g') \tag{2.3}$$

for all $g, g' \in G$. Note that this also implies that $\Phi(e_G) = e_H$.

Definition 2.1.10 (Kernel). The kernel of a group morphism $\Phi : G \to H$ is defined as the set

$$\ker(\Phi) := \{g \in G \mid \Phi(g) = e_H\}. \tag{2.4}$$

This set carries a group structure induced by that on G.

Theorem 2.1.11 (First isomorphism theorem). *Consider a group morphism $\Phi : G \to H$. If Φ is surjective, then $G/\ker(\Phi) \cong H$.*

2.2 Group theory

Definition 2.2.1 (Order of a group). The number of elements in the group. It is denoted by $|G|$ or $\mathrm{ord}(G)$.

Definition 2.2.2 (Order of an element). The order of an element $a \in G$ is the smallest integer $n \in \mathbb{N}$ such that

$$a^n = e\,, \tag{2.5}$$

where e is the identity element of G.

Definition 2.2.3 (Torsion group). A group in which all elements have finite order. The torsion set $\mathrm{Tor}(G)$ of a group G is the set of all elements $a \in G$ that have finite order. If G is Abelian, $\mathrm{Tor}(G)$ is a subgroup.

Theorem 2.2.4 (Lagrange). *Let G be a finite group and let H be a subgroup. Then $|H|$ is a divisor of $|G|$.*

Corollary 2.2.5. The order of any element $g \in G$ is a divisor of $|G|$.

Construction 2.2.6 (Grothendieck completion). Let $(A, \oplus)$ be a commutative monoid. From this monoid, one can construct an Abelian group $G(A)$, called the Grothendieck completion of A, as the quotient of $A \times A$ by the equivalence relation

$$(a_1, a_1') \sim (a_2, a_2') \iff \exists c \in A : a_1 \oplus a_2' \oplus c = a_1' \oplus a_2 \oplus c\,. \tag{2.6}$$

The identity element, denoted by 0, is given by the equivalence class of $(0, 0)$. By the definition of $G(A)$, this class contains all elements in Δ_A. In particular, $[(a, b)]$ is the additive inverse of $[(b, a)]$:

$$[(a, b)] + [(b, a)] = 0\,. \tag{2.7}$$

Universal Property 2.2.7. Let $G(A)$ be the Grothendieck completion of the Abelian monoid A. Every monoid morphism $m : A \to B$ between A and an Abelian group factors uniquely through a group morphism $\varphi : G(A) \to B$.

Example 2.2.8 (Integers). The Grothendieck completion of the natural numbers $\mathbb{N}$ is the additive group of integers $\mathbb{Z}$.

2.2.1 Abelian groups

Definition 2.2.9 (Commutator subgroup). The commutator subgroup or **derived subgroup** $[G, G]$ of G is defined as the group generated by the **commutators**

$$[g, h] := g^{-1}h^{-1}gh\,, \tag{2.8}$$

where $g, h \in G$. The **derived series** of a group G is defined as the following sequence:

$$G \subseteq [G, G] \subseteq \big[[G, G], [G, G]\big] \subseteq \cdots \tag{2.9}$$

To ease the notation, these groups are often denoted by $G^{(i)}$: $G^{(0)} := G$ and $G^{(i)} := [G^{(i-1)}, G^{(i-1)}]$.

Definition 2.2.10 (Solvable group). A group whose derived series ends in the trivial group.

Property 2.2.11 (Abelianization). The Abelianization $G/[G, G]$ is an Abelian group. Moreover, a group G is Abelian if and only if $[G, G]$ is trivial.

Corollary 2.2.12 (Abelian quotients). A quotient group G/H is Abelian if and only if $[G, G] \leq H$.

2.2.2 Cosets

Definition 2.2.13 (Coset). Let G be a group and let H be a subgroup. The left coset of H with respect to $g \in G$ is defined as the set

$$gH := \{gh \mid h \in H\}\,. \tag{2.10}$$

The right coset is defined analogously. If for all $g \in G$, the left and right cosets coincide, the subgroup H is said to be a **normal subgroup**, **normal divisor** or **invariant subgroup**.

Notation 2.2.14. The set of left (resp. right) cosets is denoted by G/H (resp. $H\backslash G$).

Definition 2.2.15 (Index). Let G be a group and let H be a subgroup. The index $|G : H|$ of H in G is defined as the number of cosets H in G. This can also be expressed as

$$|G| = |G : H||H| \,. \tag{2.11}$$

Property 2.2.16 (Tower law). Consider the subgroup inclusions $G \subset H \subset I$.

$$[I : G] = [I : H][H : G] \tag{2.12}$$

Alternative Definition 2.2.17 (Normal subgroup). Let G be a group and let H be a subgroup. Consider the **conjugacy classes** gHg^{-1} for all $g \in G$. If all classes coincide with H itself, then H is said to be a normal subgroup.

Notation 2.2.18. If N is a normal subgroup of G, this is often denoted by $N \lhd G$.

Property 2.2.19 (Quotient group). Let G be a group and let N be a normal subgroup. The coset space G/N can be turned into a group by equipping it with the product induced by that on G.

Definition 2.2.20 (Center). The center of a group G is defined as follows:

$$Z(G) := \{z \in G \mid \forall g \in G : zg = gz\} \,. \tag{2.13}$$

This is a normal subgroup of G.

Definition 2.2.21 (Normalizer). The normalizer of a subset $H \subseteq G$ is defined as follows:

$$N_G(H) := \{g \in G \mid gHg^{-1} = H\} = \{g \in G \mid [g, H] = H\} \,. \tag{2.14}$$

Definition 2.2.22 (Centralizer). The centralizer of a subset $H \subseteq G$ is defined as follows:

$$C_G(H) := \{g \in G \mid \forall h \in H : ghg^{-1} = h\} = \{g \in G \mid [g, H] = 0\} \,. \tag{2.15}$$

This group clearly satisfies $C_G(H) \lhd N_G(H)$.

2.2.3 Sylow theorems

Definition 2.2.23 (Sylow p-subgroup). Consider a finite group G. For every prime p, a Sylow p-subgroup of G is defined as a maximal p**-group** in G, i.e. every element has order a power of p and the subgroup is maximal with respect to this property. The set of all Sylow p-subgroups of G is denoted by $\mathrm{Syl}_p(G)$.

Theorem 2.2.24 (Sylow I). *Consider a finite group G. For every prime factor p of $|G|$ with multiplicity n, there exists a Sylow p-subgroup of order p^n.*

Theorem 2.2.25 (Sylow II). *Consider a finite group G and a prime factor p of $|G|$. All Sylow p-subgroups are conjugate and, in particular, isomorphic.*

Theorem 2.2.26 (Sylow III). *Consider a finite group G and a prime factor p of $|G|$ of multiplicity n such that $|G| = p^n m$ for some $m \in \mathbb{N}$ and write $n_p := |\mathsf{Syl}_p(G)|$.*

- *$m = |G : H|$ for every Sylow p-subgroup H of G.*
- *$n_p \mid m$, i.e. n_p divides m.*
- *$n_p = 1 \bmod p$.*
- *$n_p = |G : N_G(H)|$ for any Sylow p-subgroup H of G.*

2.2.4 Symmetric groups

Definition 2.2.27 (Symmetric group). The symmetric group S_n (on n elements) is defined as the set of all permutations of $\{1, \dots, n\}$. The integer $n \in \mathbb{N}$ is called the **degree** of the symmetric group. The symmetric group $\mathsf{Sym}(X)$ on a finite set X is defined similarly (by enumerating the elements)[1].

For infinite sets X, the symmetric group $\mathsf{Sym}(X)$ is defined as the group of all bijections from X to itself (where the multiplication is given by function composition).

Definition 2.2.28 (Automorphism group). Consider a set X with some extra structure (see Chapter 3 for a more formal treatment), e.g. a monoid structure or group structure. The subset of $\mathsf{Sym}(X)$ consisting of bijections that preserve this structure, the **automorphisms** of X, inherits the group structure of $\mathsf{Sym}(X)$ and is denoted by $\mathsf{Aut}(X)$.

Theorem 2.2.29 (Cayley). *Every group G is isomorphic to a subgroup of the permutation group $\mathsf{Sym}(G)$.*

Definition 2.2.30 (Cycle). A k-cycle is a permutation of the form $(a_1\ a_2\ \cdots\ a_k)$ sending a_i to a_{i+1} (and a_k to a_1). A **cycle decomposition** of an arbitrary permutation is the decomposition into a product of disjoint cycles.

Property 2.2.31 (Cycles are cyclic). Let τ be a k-cycle. Then τ is k-cyclic (hence the name): $\tau^k = e$.

Example 2.2.32. Consider the set $\{1, 2, 3, 4, 5, 6\}$. The permutation $\sigma : x \mapsto (x+2) \bmod 6$ can be decomposed as $\sigma = (1\ 3\ 5)(2\ 4\ 6)$.

[1] Two such choices are related through conjugation by a unique permutation and, hence, the resulting groups are isomorphic.

Definition 2.2.33 (Transposition). A permutation that exchanges two elements but leaves the other ones unchanged.

Property 2.2.34 (Decomposition). Any permutation can be decomposed as a product of transpositions. A permutation is said to be **even** (resp. **odd**) if the number of transpositions in its decomposition is even (resp. odd). One can prove that the parity of a permutation is well-defined, i.e. it is independent of the choice of decomposition.

Definition 2.2.35 (Alternating group). The alternating group A_n is the subgroup of S_n containing all even permutations.

Definition 2.2.36 (Shuffle). A permutation $\sigma \in S_n$ is called a (p,q)-shuffle (where $p + q = n$) if there exist disjoint increasing sequences $I = \{i_1 < \cdots < i_p\}$ and $J = \{j_1 < \cdots < j_q\}$ such that

$$\sigma(x) = \begin{cases} k & \text{if } x = i_k\,, \\ k+p & \text{if } x = j_k\,. \end{cases} \tag{2.16}$$

The name stems from the idea of dividing a deck of cards into two piles and interleaving them. This way the order in each pile is strictly preserved.

An unshuffle $\tau \in S_n$ is defined as a permuation such that τ^{-1} is a shuffle, i.e. there exist disjoint increasing sequences $I = \{i_1 < \cdots < i_p\}$ and $J = \{j_1 < \cdots < j_q\}$ such

$$\tau(k) = \begin{cases} i_k & \text{if } k \leq p\,, \\ j_{k-p} & \text{if } k > p\,. \end{cases} \tag{2.17}$$

2.2.5 Group presentations

Definition 2.2.37 (Generator). A set of elements $\{g_i\}_{i\in I} \subset G$ (where I can be infinite) is said to generate G if every element in G can be written as a finite product of the elements g_i and their inverses. These elements are then called generators of G.

Remark 2.2.38. If G is finite, the inverses can themselves be expressed as powers, so inverses are only required for infinite groups.

Definition 2.2.39 (Relation). Let G be a group. If the product of a number of elements $g \in G$ is equal to the identity e, this product is said to be a relation on G.

Definition 2.2.40 (Complete set of relations). Let G be a group generated by a set of elements S and let R be a set of relations on S. If G is uniquely (up to an isomorphism) determined by S and R, the set of relations is said to be complete.

Definition 2.2.41 (Presentation). Let G be a group with generators S and let R be a complete set of relations. The pair (S, R) is called a presentation of G.

If R is finite, G is said to be **finitely related**, while if S is finite, G is said to be **finitely generated**. If both S and R are finite, G is said to be **finitely presented**.

Notation 2.2.42. The presentation of a group G is often denoted by $\langle S \mid R \rangle$, where S is the set of generators and R the set of relations.

Remark 2.2.43. It should be clear that a group can have many different presentations and that it is (very) difficult to tell if two groups are isomorphic by just looking at their presentations.

2.2.6 Direct products

Definition 2.2.44 (Direct product). Let G, H be two groups. The direct product $G \times H$ is defined as the set-theoretic Cartesian product $G \times H$ equipped with a group operation defined as

$$(g_1, h_1) \cdot (g_2, h_2) = (g_1 g_2, h_1 h_2), \tag{2.18}$$

where the operations on the right-hand side are the group operations in G and H. This definition can be generalized to any number of groups, even an infinite number of them if the n-tuples are replaced by elements of the infinite Cartesian product.

If $g \in G_1 \times \cdots \times G_n$ can be written as $(g_1, \dots, g_n)$ for $g_i \in G_i$, the g_i are called the **components** of g.

Definition 2.2.45 (Weak direct product). Consider the direct product of groups. The subgroup consisting of all elements for which all components, except finitely many of them, are the identity, is called the weak direct product. In the case of Abelian groups, this is often called the **direct sum**. For a finite number of groups, the direct product and direct sum coincide.

Notation 2.2.46. The direct sum is often denoted by $\oplus$ (in accordance with the notation for *vector spaces* and other algebraic structures that will be introduced further on).

Definition 2.2.47 (Inner semidirect product). Let G be a group, $H \leq G$ a subgroup and $N \lhd G$ a normal subgroup. G is said to be the inner semidirect product of H and N, denoted by $N \rtimes H$, if it satisfies the following equivalent statements:

- $G = NH := \{nh \mid n \in N, h \in H\}$, where $N \cap H = \{e\}$.
- For every $g \in G$, there exist a unique $n \in N, h \in H$ such that $g = nh$.
- For every $g \in G$, there exist a unique $n \in N, h \in H$ such that $g = hn$.
- There exists a group morphism $\rho : G \to H$ that satisfies $\rho|_H = e$ and $\ker(\rho) = N$.
- The composition of the natural embedding $i : H \to G$ and the projection $\pi : G \to G/N$ gives an isomorphism between H and G/N.

Whenever G is isomorphic to $N \rtimes H$, it is said to **split** over N.

Property 2.2.48 (Normal subgroups). If both H and N in the above definition are normal, the inner semidirect product coincides with the direct product. If the subgroups H and N have presentations $\langle S_H \mid R_H \rangle$ and $\langle S_N \mid R_N \rangle$, the direct product is given by

$$H \times N = \langle S_H \cup S_N \mid R_H \cup R_N \cup R_C \rangle \,, \tag{2.19}$$

where R_C is the set of relations that enforce the commutativity of H and N.

Definition 2.2.49 (Outer semidirect product). Let G, H be two groups and let $\varphi : H \to \mathrm{Aut}(G)$ be a group morphism. The outer semidirect product $G \rtimes_\varphi H$ is defined as the set-theoretic Cartesian product $G \times H$ equipped with a binary relation $\cdot$ such that

$$(g_1, h_1) \cdot (g_2, h_2) = \big(g_1 \varphi(h_1)(g_2), h_1 h_2\big) \,. \tag{2.20}$$

The structure $(G \rtimes_\varphi H, \cdot)$ forms a group.

By noting that the set $N = \{(g, e_H) \mid g \in G\}$ is a normal subgroup isomorphic to G and that the set $B = \{(e_G, h) \mid h \in H\}$ is a subgroup isomorphic to H, one can also construct the outer semidirect product $G \rtimes_\varphi H$ as the inner semidirect product $N \rtimes B$.

Remark 2.2.50 (Direct products). The direct product of groups is a special case of the outer semidirect product where the group morphism is given by the trivial map $\varphi : h \mapsto e_G$.

Semidirect products can even be generalized further.

Definition 2.2.51 (Bicrossed product of groups[2]). Consider a group G with two subgroups $K, H \leq G$, with $K \cap H = \{e\}$, such that every element $g \in G$ can be uniquely decomposed as the product of an element in H and an element in K. This implies that for $k \in$ and $h \in H$, there exists a unique decomposition of kh of the form

$$kh = (k \cdot h) k^h \,, \tag{2.21}$$

where $k \cdot h \in H$ and $k^h \in K$.

It can be checked that the associativity of the product implies that $- \cdot - : K \times H \to H$ defines a *left action* of K on H and that $-^- : K \times H \to K$ defines a *right action* of H on K (see Section 2.3 further on). Some other properties are obtained in the same way:

- $e^h = e$,
- $k \cdot e = e$,
- $(kk')^h = k^{k' \cdot h} k'^h$, and
- $k \cdot (hh') = (k \cdot h)(k^h \cdot h')$.

[2] Also called the **Zappa–Szép product**.

Any two groups having this structure are said to form a **matched pair** (of groups). Given a matched pair of groups, one can define the bicrossed product $H \bowtie K$ as follows:

$$(h,k)(h',k') := \left(h(k \cdot h), k^{h'}k'\right). \tag{2.22}$$

2.2.7 Free groups

Definition 2.2.52 (Free group). Consider a set S. The free group on S is the group generated by **words** in S, i.e. finite sequences of elements in S.

The definition of a group presentation (Definition 2.2.41) can now be restated.

Alternative Definition 2.2.53 (Presentation). A group G is said to have a presentation $\langle S \mid R \rangle$ if it is isomorphic to the quotient of the free group on S by the normal subgroup generated by R.

Definition 2.2.54 (Free product). Consider two groups G, H. The free product $G * H$ is defined as the set consisting of all words composed of elements in G and H together with the concatenation (and reduction[3]) as multiplication. Due to the reduction, every element in $G * H$ has a unique expression of the form $g_1 h_1 g_2 h_2 \cdots g_n h_n$.

Remark 2.2.55 (Cardinality). For nontrivial groups, the free product is always infinite.

Alternative Definition 2.2.56 (Free product). The free product of two groups G and H can equivalently be defined as the free group on the set $G \cup H$. It follows that if G, H have presentations $\langle S_G \mid R_G \rangle$ and $\langle S_H \mid R_H \rangle$, respectively, the free product is given by

$$G * H = \langle S_G \cup S_H \mid R_G \cup R_H \rangle. \tag{2.23}$$

By comparing to Eq. (2.19), it can be seen that the free product is a generalization of the direct product.

Definition 2.2.57 (Free product with amalgamation). Consider three groups F, G, H together with two group morphisms $\phi : F \to G$ and $\psi : F \to H$. The free product with amalgamation $G *_F H$ is defined by adding the following set of relations to the presentation of the free product $G * H$:

$$\left\{\phi(f)\psi(f)^{-1} = e \mid f \in F\right\}. \tag{2.24}$$

This is the same as saying that the free product with amalgamation can be constructed as

$$G *_F H = (G * H)/N_F, \tag{2.25}$$

where N_F is the normal subgroup generated by elements of the form $\phi(f)\psi(f)^{-1}$.

[3] Two elements of the same group, written next to each other, are replaced by their product.

Definition 2.2.58 (Free Abelian group). An Abelian group G is said to be freely generated on the generators $\{g_i\}_{i\in I}$ if every element $g \in G$ can be uniquely written as a formal linear combination of the generators:

$$G = \left\{ \sum_{i\in J} a_i g_i \,\middle|\, a_i \in \mathbb{Z}, J \subseteq I \text{ is finite} \right\}. \tag{2.26}$$

The set of generators is called a **basis**[4] of G. The number of elements in the basis is called the **rank** of G.

Property 2.2.59 (Nielsen–Schreier). Every subgroup of a free group is free.

By definition, every finitely generated Abelian group can be decomposed as a quotient of two free, finitely generated Abelian groups. In fact, more is true.

Theorem 2.2.60 (Fundamental theorem of finitely generated Abelian groups). *Every finitely generated Abelian group G can be decomposed as a direct sum of a free, finitely generated Abelian group and a torsion group (Definition 2.2.3):*

$$G = \mathbb{Z}^n \oplus T \qquad \textit{with} \qquad T \equiv \mathbb{Z}_{h_1} \oplus \cdots \oplus \mathbb{Z}_{h_m}, \tag{2.27}$$

*where $n \in \mathbb{N}$ is called the **rank** of G and all $h_i \in \mathbb{N}$ are prime powers. The group T is called the **torsion subgroup**.*

*In fact, a unique (up to isomorphism) decomposition can be obtained by requiring the prime powers to satisfy $h_{i+1} \mid h_i$ for all $i < m$. This decomposition is called the **invariant factor decomposition** and the prime factors h_i are then called the **invariant factors** of G.*

Property 2.2.61. A free Abelian group G is finitely generated if and only if its rank n is finite, in which case $G \cong \mathbb{Z}^n$. On the other hand, a finitely generated Abelian group is finite if and only if its rank is 0.

2.3 Group actions

Definition 2.3.1 (Group action). Let G be a group with identity element $e \in G$ and let X be a set. A map $\rho : G \times X \to X$ is called an action of G on X if it satisfies the following conditions for all $g, h \in G$ and $x \in X$:

1. **Identity**: $\rho(e, x) = x$.
2. **Compatibility**: $\rho(gh, x) = \rho(g, \rho(h, x))$.

The set X is called a (left) G**-space** or G**-set**. Right G-spaces are defined a similar way.

Remark 2.3.2. Note that this definition already makes sense for monoids (Definition 2.1.2).

[4] In analogy with the basis of a *vector space* (see Section 22.1).

Alternative Definition 2.3.3. A group action can equivalently be defined as a group morphism $\rho : G \to \mathsf{Sym}(X)$. It assigns a permutation of X to every element $g \in G$. If the set X is equipped with some extra algebraic structure, one should replace $\mathsf{Sym}(X)$ by $\mathsf{Aut}(X)$, i.e. the action of G should respect the structure.

Notation 2.3.4. The action $\rho(g, x)$ is often denoted by $g \cdot x$ or $g \triangleright x$, or even gx if no confusion can arise.

Example 2.3.5 (Adjoint action). Consider a group G. This group acts on itself by conjugation:

$$\mathsf{Ad} : G \times G \to G : (g, h) \mapsto g^{-1}hg\,. \tag{2.28}$$

Definition 2.3.6 (Equivariant map). Let X, Y be two G-spaces. An equivariant map between these sets is a function $f : X \to Y$ satisfying

$$g \cdot f(x) = f(g \cdot x)\,, \tag{2.29}$$

where, by abuse of notation, the symbol $\cdot$ represents the group actions on both X and Y. An equivariant map is sometimes called a G**-map**.[5]

Example 2.3.7 (G-module). An Abelian group M equipped with a left group action $\varphi : G \to \mathsf{Aut}(M)$, i.e. an action that acts by group morphisms. Equivariant maps of G-modules are also called **intertwining maps** or **intertwiners**.

Definition 2.3.8 (Orbit). The orbit of an element $x \in X$ with respect to the action a group G is defined as the set

$$G \cdot x := \{g \cdot x \mid g \in G\}\,. \tag{2.30}$$

The relation

$$p \sim q \iff \exists g \in G : p = g \cdot q \tag{2.31}$$

induces an equivalence relation on X for which the equivalence classes coincide with the orbits of the G-action. The set of equivalence classes $X/\sim$, often denoted by X/G, is called the **orbit space**.

Definition 2.3.9 (Stabilizer). The stabilizer group (also called **isotropy group** or **little group**) of an element $x \in X$ with respect to the action of a group G is defined as the set

$$G_x := \{g \in G \mid g \cdot x = x\}\,. \tag{2.32}$$

This is a subgroup of G for all $x \in X$.

[5] G-spaces together with the G-maps constitute a category.

Theorem 2.3.10 (Orbit-stabilizer theorem). *Let G be a group acting on a set X and let G_x be the stabilizer of some $x \in X$. The following relation holds:*

$$|G \cdot x||G_x| = |G| \,. \tag{2.33}$$

Definition 2.3.11 (Free action). A group action is said to be free if $g \cdot x = x$ implies $g = e$ for any $x \in X$. Equivalently, a group action is free if the stabilizer group of all elements is trivial.

Definition 2.3.12 (Faithful action). A group action is said to be faithful or **effective** if the morphism $G \to \mathrm{Aut}(X)$ is injective. Alternatively, a group action is faithful if for every two group elements $g, h \in G$, there exists an element $x \in X$ such that $g \cdot x \neq h \cdot x$.

Definition 2.3.13 (Transitive action). A group action is said to be transitive if for every two elements $x, y \in X$, there exists a group element $g \in G$ such that $g \cdot x = y$. Equivalently, a group action is transitive if there is only one orbit.

Property 2.3.14. Let X be a set equipped with a transitive action of a group G. There exists a bijection $X \cong G/G_x$, where G_x is the stabilizer of any element $x \in X$.

Proof. Choose an element $x \in X$. The stabilizer of x with respect to G is the set

$$S_x = \{g \in G \mid g \cdot x = x\} \,.$$

Due to the transitivity of the group action one obtains that

$$\forall x, y \in X : \exists h \in G : h \cdot x = y \,.$$

So, for every $z \in X$, one can choose a group element g_z such that $g_z \cdot x = z$. For all elements in the coset $g_z S_x = \{g_z s \in G \mid s \in S_x\}$, the following equality is satisfied:

$$(g_z s) \cdot x = g_z \cdot (s \cdot x) = g_z \cdot x = z \,.$$

This implies that the map $\Phi : G/S_x \to X$ is surjective.

Now, one needs to prove that Φ is also injective. A proof by contradiction is given. Choose two distinct cosets gS_x and hS_x. There exist two elements $G, H \in X$ such that $g \cdot x = G$ and $h \cdot x = H$. Assume that $G = H$. This means that

$$\begin{aligned} & g \cdot x = h \cdot x \\ \iff & (h^{-1}g) \cdot x = x \\ \iff & h^{-1}g \in S_x \\ \iff & hS_x \ni h(h^{-1}g) = g \,. \end{aligned}$$

This would imply that $gS_x = hS_x$, which is in contradiction with the assumptions. It follows that $G \neq H$ and, hence, that Φ is injective. □

Definition 2.3.15 (Homogeneous space). A set equipped with a transitive group action.

Definition 2.3.16 (Principal homogenous space). If the action of a group G on a homogeneous space X is also free, then X is said to be a principal homogeneous space or G**-torsor**.

Example 2.3.17 (Affine space). The n-dimensional affine space $\mathbb{A}^n$ is an $\mathbb{R}^n$-torsor.

Definition 2.3.18 (Crossed module). A crossed module is a quadruple (G, H, t, α) where:

- G, H are two groups,
- t is a group morphism $H \to G$, and
- α is a group morphism $G \to \mathsf{Aut}(H)$.

These data are required to satisfy two compatibility conditions:

1. G**-equivariance**:

$$t\big(\alpha(g)h\big) = gt(h)g^{-1}\,. \tag{2.34}$$

2. **Peiffer identity**:

$$\alpha\big(t(h)\big)h' = hh'h^{-1}\,. \tag{2.35}$$

2.4 Rings

2.4.1 General

Definition 2.4.1 (Ring). Let R be a set equipped with two binary operations $+, \cdot$ (called **addition** and **multiplication**). $(R, +, \cdot)$ is a ring if it satisfies the following axioms:

1. $(R, +)$ is an Abelian group.
2. $(R, \cdot)$ is a monoid.
3. Multiplication is distributive with respect to addition.

Definition 2.4.2 (Field). A ring $(R, +, \cdot)$ for which the monoid $(R\backslash\{1_+\}, \cdot)$ is an Abelian group and $1_+ \neq 1_\cdot$.

Definition 2.2.22 can also be formulated in the context of ring theory, where two operations play a role.

Definition 2.4.3 (Centralizer). Let $(R, +, \cdot)$ be a ring and consider a subset $S \subseteq R$. The centralizer (or **commutant**) of S is defined as follows:

$$C_R(S) \equiv S' := \{r \in R \mid \forall s \in S : r \cdot s = s \cdot r\} = \{r \in R \mid [r, S] = 0\}\,. \tag{2.36}$$

This subset is actually a subring of R.

Definition 2.4.4 (Unit). An invertible element of a ring $(R, +, \cdot)$. The set of units, denoted by R^*, forms a group under multiplication.

Definition 2.4.5 (Integral domain). A commutative ring R in which the product of two nonzero elements is again nonzero.

Definition 2.4.6 (Field of fractions). Let I be an integral domain. Its field of fractions $\mathsf{Frac}(I)$ is the smallest field containing I, i.e. every injective ring morphism $I \to F$ factors uniquely through the embedding $I \to \mathsf{Frac}(I)$.

Construction 2.4.7. The field of fractions $\mathsf{Frac}(I)$ can be constructed as follows. On the set $I \times I/\{0\}$ one can define the equivalence relation

$$(i, j) \sim (p, q) \iff nq = pj\,. \tag{2.37}$$

The equivalence class of an element (i, j) is often denoted by the fraction $\frac{i}{j}$. The quotient set $\big(I \times I/\{0\}\big)/\sim$ together with the operations

- **Addition**: $\dfrac{i_1}{j_1} + \dfrac{i_2}{j_2} = \dfrac{i_1 j_2 + i_2 j_1}{j_1 j_2}$,
- **Multiplication**: $\dfrac{i_1}{j_1} \cdot \dfrac{i_2}{j_2} = \dfrac{i_1 i_2}{j_1 j_2}$.

These operations strongly resemble those of ordinary fractions, hence the name.

Example 2.4.8 (Integers). The field of fractions of $\mathbb{Z}$ is $\mathbb{Q}$.

Definition 2.4.9 (Rational fractions). Consider an integral domain I. The commutative ring of rational fractions $I(x)$ over I is given by the field of fractions of the polynomial ring $I[x]$.

Definition 2.4.10 (Reduced ring). A ring that contains no nonzero nilpotents.

Construction 2.4.11 (Localization). Let R be a commutative ring and let S be a multiplicatively closed set in R. Define an equivalence relation $\sim$ on $R \times S$ in the following way:

$$(r_1, s_1) \sim (r_2, s_2) \iff \exists t \in S : t(r_1 s_2 - r_2 s_1) = 0\,. \tag{2.38}$$

The set $S^{-1}R := (R \times S)/\sim$, called the localization of R with respect to S, can now be turned into a ring by defining an addition and a multiplication. By writing $(r, s) \in S^{-1}R$ as the formal fraction $\frac{r}{s}$, these operations are defined in analogy with the those of ordinary fractions:

- **Addition**: $\frac{r_1}{s_1} + \frac{r_2}{s_2} = \frac{r_1 s_2 + r_2 s_1}{s_1 s_2}$,
- **Multiplication**: $\frac{r_1}{s_1} \cdot \frac{r_2}{s_2} = \frac{r_1 r_2}{s_1 s_2}$.

Remark 2.4.12. The localization of R with respect to the set S can be interpreted as the ring obtained by formally adjoining inverses of S. This formalized by Property 2.4.14 below.

Notation 2.4.13. For specific cases, different notations are sometimes used. For example, choose an element $f \in R$ and let R_f denote the localization of R with respect to the set of powers of f, i.e. $S = \{f^n \mid n \in \mathbb{N}\}$. This is called the **localization at (the element)** f. Another example occurs when working with prime ideals. Let P be a *prime ideal* (see the next section). It is not hard to show that the complement $R \backslash P$ is multiplicatively closed. The localization of R with respect to this set is denoted by R_P and is called the **localization at (the prime ideal)** P.

Property 2.4.14. Consider the localization R_f at an element $f \in R$. This localization is isomorphic to the polynomial ring $R[f^{-1}]$.

Property 2.4.15 (Field of fractions). If R is an integral domain, the localization at the zero ideal coincides with the field of fractions $\mathsf{Frac}(R)$.

Definition 2.4.16 (Irreducible element). Let I be an integral domain. An element $r \in I$ is said to be irreducible if:

1. **Invertibility**: r is not a unit, i.e. it is not invertible.
2. **Decomposition**: r cannot be decomposed as a product of other noninvertible elements.

Definition 2.4.17 (Unique factorization domain). An integral domain in which every element admits a unique factorization into irreducible elements. These are often abbreviated as **UFD**s.

Definition 2.4.18 (Absolute value). Let K be a ring (usually a field). An absolute value (sometimes called an **exponential** or **multiplicative valuation**) is a multiplicative *seminorm* (see Definition 25.3.1), i.e. a function $|\cdot| : K \to \mathbb{R}$ satisfying the following conditions for all $x, y \in K$:

1. **Positivity**: $|x| \geq 0$,
2. **Nondegeneracy**: $|x| = 0 \iff x = 0$,
3. **Multiplicativity**: $|xy| = |x|\,|y|$, and
4. **Triangle inequality**: $|x + y| \leq |x| + |y|$.

If the triangle inequality is strengthened to

$$|x+y| \leq \max\big(|x|, |y|\big)\,, \tag{2.39}$$

the absolute value is said to be **ultrametric** or **non-Archimedean** since this implies that, for all $n \in \mathbb{N}$,

$$|nx| \leq |x| \tag{2.40}$$

holds and, hence, also that $|n| \leq 1$ holds for all $n \in \mathbb{Z}$. On the other hand, if the ultrametric condition does not hold, the **Archimedean** case, $0 < |x| < |y|$ implies that there exists a $n \in \mathbb{N}$ such that $|y| < |nx|$.

Definition 2.4.19 (Place). Two absolute values $|\cdot|_1 : K \to \mathbb{R}$, $|\cdot|_2 : K \to \mathbb{R}$ are equivalent if

$$|x|_1 < 1 \iff |x|_2 < 1 \tag{2.41}$$

for all $x \in K$. An equivalence class of absolute values is called a place.

(Non-Archimedean) absolute values can be extended in the following way.

Definition 2.4.20 (Valuation). Let K be a field and let Γ be a totally ordered Abelian group (Definition 1.7.5). The group law and the order relation on Γ can be extended to the union $\Gamma \cup \{\infty\}$ in the following way (the notation ∞ is only a convention):

- $g + \infty := \infty + g := \infty$ for all $g \in \Gamma$, and
- $g \leq \infty$ for all $g \in \Gamma$.

A valuation on K (with values in Γ) is a map $\nu : K \to \Gamma \cup \{\infty\}$ such that:

1. $\nu(a) = \infty \iff a = 0$,
2. $\nu(ab) = \nu(a) + \nu(b)$ for all $a, b \in K^\times$, and
3. $\min\big(\nu(a), \nu(b)\big) \leq \nu(a+b)$, where the equality holds if $\nu(a) \neq \nu(b)$.

Example 2.4.21 (Absolute values). Let $|\cdot| : K \to \mathbb{R}$ be a non-Archimedean absolute value and consider a number $b \in]1, +\infty[$. The function

$$\nu_b(x) := \begin{cases} -\log_b |x| & \text{if } x \neq 0 \\ \infty & \text{if } x = 0 \end{cases} \tag{2.42}$$

gives a valuation on K.

2.4.2 Ideals

Definition 2.4.22 (Ideal). Let $(R, +, \cdot)$ be a ring with $(R, +)$ its additive group. A subset $I \subseteq R$ is called a (two-sided) ideal of R if it satisfies the following conditions:

1. $(I, +)$ is a subgroup of $(R, +)$.
2. $\forall n \in I, \forall r \in R : n \cdot r, r \cdot n \in I$.

Definition 2.4.23 (Artinian ring). A ring is said to be Artin(ian) if it satisfies the **descending chain condition** on ideals, i.e. if it contains no infinite descending chain (Definition 1.7.13) of ideals.

Definition 2.4.24 (Noetherian ring). A ring is said to be Noether(ian) if it satisfies the **ascending chain condition** on ideals, i.e. if it contains no infinite ascending chain of ideals.

Definition 2.4.25 (Simple ring). A ring that has no nontrivial two-sided ideals. (Some authors require the ring to be Artinian.)

Definition 2.4.26 (Unit ideal). A ring considered as an ideal of itself.

Definition 2.4.27 (Proper ideal). An ideal that is not equal to the ring itself.

Definition 2.4.28 (Prime ideal). Let R be a ring. A proper ideal $I \subset R$ is a prime ideal if for any $a, b \in R$ the following relation holds:

$$ab \in I \implies a \in I \vee b \in I\,. \tag{2.43}$$

If only

$$ab \in I \implies a \in I \vee \exists n \in \mathbb{N} : b^n \in I \tag{2.44}$$

holds, then I is said to be **primary**.

Definition 2.4.29 (Maximal ideal). A proper ideal that is not contained in another proper ideal.

Property 2.4.30. Every maximal ideal is prime.

Definition 2.4.31 (Jacobson radical). The Jacobson radical of a ring R, often denoted by $J(R)$, is the ideal obtained as the intersection of all maximal left (or right) ideals. Equivalently, it is the intersection of the *annihilators* of all simple, left (or right) R-modules.

Definition 2.4.32 (Radical). Let I be an ideal in a commutative ring R. The radical $\sqrt{I}$ of I is defined as follows:

$$\sqrt{I} := \{r \in R \mid \exists n \in \mathbb{N} : r^n \in I\}\,. \tag{2.45}$$

Property 2.4.33. If I is primary, $\sqrt{I}$ is prime and, in fact, it is the smallest prime ideal containing I.

Theorem 2.4.34 (Lasker–Noether). *Let I be an ideal in a Noetherian ring R. There exists a unique decomposition in primary ideals*

$$I = Q_1 \cap \ldots \cap Q_n \tag{2.46}$$

such that[6]

- *the decomposition is irredundant, i.e.*

$$\forall i \leq n : \bigcap_{i \neq j} Q_j \not\subseteq Q_i \,, \tag{2.47}$$

- *the prime ideals* $\sqrt{Q_i}$ *are all distinct,*
- *the prime ideals* $\sqrt{Q_i}$ *are uniquely determined by* I*, and*
- *if* $\sqrt{Q_i}$ *is minimal (with respect to inclusion), then* Q_i *itself is uniquely determined by* I*.*

Construction 2.4.35 (Generating ideals). Let R be a ring and let X be a subset of R. The two-sided ideal generated by X is defined as the intersection of all two-sided ideals containing X. An explicit construction is given by

$$I = \left\{ \sum_{i=1}^{n} l_i x_i r_i \,\middle|\, n \in \mathbb{N}, \forall i \leq n : l_i, r_i \in R \wedge x_i \in X \right\}. \tag{2.48}$$

Left and right ideals are generated in a similar fashion.

Notation 2.4.36. If the ideal I is generated by the elements $\{f_j\}_{j \in J}$ (for some index set J), it is often denoted by

$$I \equiv (f_1, f_2, \ldots) \,. \tag{2.49}$$

Construction 2.4.37 (Extension). Let I be an ideal of a ring R and let $\iota : R \to S$ be a ring morphism. The extension of I with respect to ι is the ideal generated by the set $\iota(I)$.

Definition 2.4.38 (Principal ideal). An ideal that is generated by a single element.

Definition 2.4.39 (Principal ideal domain). An integral domain (Definition 2.4.5) in which every ideal is principal. These are often abbreviated as **PID**s.

Property 2.4.40 (UFD). Every PID is also a UFD.

Definition 2.4.41 (Fractional ideal). Let R be an integral domain. A fractional ideal I of R is a submodule of $\mathsf{Frac}(R)$ such that $rI \subseteq R$ for some $r \in R$. In the case of $R = \mathbb{Z}$, r would correspond to the common denominator of the fractions in I.

2.4.3 Algebraic geometry

This section lays the foundations for Chapter 10.

Definition 2.4.42 (Local ring). A ring for which a unique maximal ideal exists.

[6] The last two conditions can be dropped, but then the decomposition is not unique anymore.

Property 2.4.43 (Characterization by invertible complements). A ring R is local if and only if there exists a maximal ideal M such that every element in the complement $R \backslash M$ is invertible.

Property 2.4.44 (Prime localization). The localization of a ring R with respect to a prime ideal P is a local ring, where the maximal ideal is given by the extension of P with respect to the ring morphism $\iota : R \to R_P$. Equivalently, this says that the maximal ideal is given by PR_P.

Definition 2.4.45 (Residue field). Consider a local ring R and let I be its maximal ideal. The quotient ring R/I forms a field, called the residue field.

The following definition is similar to the one for groups (cf. Definition 2.2.41).

Definition 2.4.46 (Presentation). Consider a ring morphism $f : R \to S$. This morphism is said to be of **finite type** (or S is called an R-algebra of finite type) if S is isomorphism to a quotient of $R[x_1, \dots, x_n]$ for some $n \in \mathbb{N}$. The morphism is said to be **finitely presented** if there exist integers $m, n \in \mathbb{N}$ such that $S \cong R[x_1, \dots, x_n]/(f_1, \dots, f_m)$ for some polynomials $\{f_1, \dots, f_m\}$ over R.

Definition 2.4.47 (Reduction). An **infinitesimal extension** or **infinitesimal thickening** is a ring morphism $R \to R/I$, where $I \subseteq R$ is a nilpotent ideal. If I is the nilradical of R, the quotient is called a **reduced ring**. The forgetful functor $\mathsf{Red} : \mathbf{CRing}^{\mathbf{ext}} \to \mathbf{CRing}$ is called the reduction functor.

Definition 2.4.48 (Formally étale morphism). A ring morphism $f : R \to S$ such that for all R-rings A and nilpotent ideals $I \subset A$ the map $\mathsf{Hom}_R(S, A) \to \mathsf{Hom}_R(S, A/I)$ is a bijection. If this map is merely injective (resp. surjective), the morphism f is said to be **formally smooth** (resp. **formally unramified**).

The map $\mathsf{Hom}_R(S, A) \to \mathsf{Hom}_R(S, A/I)$ corresponds to lifts of the following form

$$\begin{tikzcd} R \arrow[r] \arrow[d, "f"'] & A \arrow[d] \\ S \arrow[r] \arrow[ur, dashed] & A/I \end{tikzcd}$$

In other words, formally étale morphisms have the unique left lifting property with respect to infinitesimal thickenings. Saying that $f : R \to S$ is formally smooth (resp. formally unramified) means that there exists at least (resp. at most) one lift.

Remark 2.4.49 (Topological rings). The previous definition can be generalized to *topological rings* R and S. In this case A ranges over discrete algebras and all morphisms are taken to be continuous morphisms.

[7] This form of ring completion is sometimes called the **formal completion** or **adic completion**.

Definition 2.4.50 (Completion[7]). Consider a commutative ring R together with a proper ideal I. The completion $\widehat{R}$ of R with respect to I is the inverse limit $\varprojlim_{n\in\mathbb{N}}(R/I^n)$.

Property 2.4.51 (Adic topology). When the filtration ends in the trivial ideal, for example when R is a Noetherian local ring, the completion $\widehat{R}$ is a topological ring (see Chapter 6). The **Krull topology** or **I-adic topology** is the topology for which a basis of 0 is given by the powers of I.

Example 2.4.52. The completion of $R[x]$ with respect to (x) is given by the ring of formal power series $R[[x]]$. In general, one can think of these completions as containing (formal) *Taylor series expansions*.

2.4.4 Modules

Definition 2.4.53 (R-module). Let $(R, +, \cdot)$ be a ring. An Abelian group $(M, \oplus)$ is said to be a left R-module if there exists a left (monoid) action $\triangleright : (R, \cdot) \times M \to M$ that satisfies the following axioms:

1. **Left distributivity**: $r \triangleright (m \oplus n) = r \triangleright m \oplus r \triangleright n$ for all $r \in R$ and $m, n \in M$.
2. **Right distributivity**: $(r + s) \triangleright m = r \triangleright m \oplus s \triangleright m$ for all $r, s \in R$ and $m \in M$.

These conditions make sure that both the additive structure $(R, +)$ and the group structure $(M, \oplus)$ are compatible with the action of $(R, \cdot)$. Due to these compatibility conditions one can identify $\cdot \sim \triangleright$ and $+ \sim \oplus$ without confusion.

Remark 2.4.54 (Categorical perspective ♣). The definition of a ring can be defined more concisely in categorical terms. Recall the definition of an algebra over a monad (Definition 3.3.19). Modules over a monoid object A are defined as algebras over the monad $A \otimes -$. A ring R is a monoid object in the category **Ab** of Abelian groups. So a module M over R consists of a morphism $\alpha : R \otimes M \to M$ satisfying the algebra axioms. The distributivity laws come for free since α is a morphism in **Ab** and, hence, is bilinear (in both arguments).

Alternative Definition 2.4.55 (R-module). The above two formulations can be restated similar to that of group modules (Example 2.3.7). Consider the Abelian group $(M, \oplus)$. Its endomorphism set $\mathsf{End}(M, \oplus)$ can be given the structure of a ring where the addition is induced by that on M and the multiplication is given by composition. A left R-module structure is then simply a ring morphism $R \to \mathsf{End}(M, \oplus)$.

Definition 2.4.56 (Morita equivalence). Two rings R, S are said to be Morita equivalent if their categories of left modules (or, equivalently, right modules) are equivalent.

Property 2.4.57. Two rings R, S are Morita equivalent if and only if

$$S \cong eM_n(R)e \tag{2.50}$$

for some $n \in \mathbb{N}$ and *full idempotent* $e \in M_n(R)$.

Definition 2.4.58 (Bimodule). Let R, S be two rings. An Abelian group $(M, \oplus)$ is an $R - S$-bimodule if

1. it is a left R-module,
2. it is a right S-module, and
3. the actions are compatible:

$$(r \triangleright m) \triangleleft s = r \triangleright (m \triangleleft s) \tag{2.51}$$

for all $r \in R$, $s \in S$ and $m \in M$.

Definition 2.4.59 (Enveloping algebra). Let A be an *algebra over a commutative ring* R (see Remark 22.7.2). The enveloping algebra $A^e := A \otimes A^{\mathrm{op}}$ is defined as follows:

$$(a \otimes b) \cdot (a' \otimes b') := (a \cdot a') \otimes (b' \cdot b) \tag{2.52}$$

for all $a, a', b, b' \in R$.[8]

Alternative Definition 2.4.60 (Bimodule). Let A be an algebra. An A-bimodule is a left A^e-module.

Definition 2.4.61 (Free module). An R-module M that admits a **basis**, i.e. there exists a set $\{x_i\}_{i \in I}$, where I can be infinite, such that:

1. Every element $m \in M$ can be written as a linear combination $\sum_{j \in J_m} r_j x_j$, where $J_m \subseteq I$ is finite.
2. The set $\{x_i\}_{i \in I}$ is linearly independent in the sense that

$$\sum_{j \in J \subseteq I} r_j x_j = 0 \implies \forall j \in J : r_j = 0. \tag{2.53}$$

Example 2.4.62 (Dual numbers). Let R be a ring. The R-algebra of dual numbers, often denoted by $R[\varepsilon]$, is defined as the free R-module with basis $\{1, \varepsilon\}$ subject to the relation $\varepsilon^2 = 0$.

Property 2.4.63 (Division rings). For a general R-module, the existence of a basis is not guaranteed unless R is a *division ring*. (See Construction 22.1.10 for more information.)

Corollary 2.4.64. Since every field is in particular a division ring, the existence of a basis follows from the above property for R-modules.

[8] Note that this is simply a consequence of the definitions of the **opposite ring** ($a \cdot_{\mathrm{op}} b := b \cdot a$) and the *tensor product of rings*.

Definition 2.4.65 (Projective module). A module P that can be expressed as

$$P \oplus M = F\,, \tag{2.54}$$

where M is a module and F is a free module, i.e. P is projective if it is a direct sumand of a free module.

Theorem 2.4.66 (Nakayama's lemma[9]). *Consider a commutative ring R. If $I \subset R$ is an ideal and M a finitely generated R-module such that $IM = M$, there exists an element $r \in R$ with $r = 1$ in R/I such that $rM = 0$.*

The following corollary is also often called Nakayama's lemma:

Corollary 2.4.67. If M is a finitely generated R-module and $J(R)M = M$, then $M = 0$.

Corollary 2.4.68. If M is a finitely generated R-module and there exist elements $m_1, \dots, m_k \in M$ whose image in $M/J(R)M$ generate this quotient, the elements generate M.

2.4.5 Semisimplicity

Definition 2.4.69 (Simple module). A module over a ring that contains no nontrivial submodules. A module is said to be **semisimple** if it admits a decomposition as a direct sum of simple modules. A ring is said to be semisimple if it is semisimple as a module over itself.

Property 2.4.70 (Jacobson radical). A ring is semisimple if and only if it is Artinian and if its Jacobson radical (Definition 2.4.31) vanishes.

Theorem 2.4.71 (Artin–Wedderburn). *Every semisimple ring is isomorphic to a direct sum of matrix rings over division rings D_i with multiplicity n_i. Furthermore, the integers D_i and n_i are unique (up to a permutation of the indices).*

2.5 Galois theory ♣

2.5.1 Polynomials

Definition 2.5.1 (Polynomial ring). Let R be a (commutative unital) ring and consider a set X. The polynomial ring on the **indeterminates** X is defined as the free commutative R-algebra on X. Often, X will be a finite set: $R[X] \equiv R[x_1, \dots, x_n]$.

Theorem 2.5.2 (Hilbert's basis theorem). *If a ring is Noetherian (Definition 2.4.24), any polynomial ring with finitely many indeterminates over it is also Noetherian.*

In the context of polynomials, Definition 2.4.16 becomes the following definition.

[9] Also called the **Krull–Azumaya lemma**.

Definition 2.5.3 (Irreducible polynomial). A polynomial that cannot be factorized as the product of two nonconstant polynomials. Over a field, a polynomial is irreducible if and only if the ideal it generates is maximal.

Definition 2.5.4 (Degree). The degree of a polynomial f is equal to the largest integer $d \in \mathbb{N}$ for which f contains a monomial $x_1^{i_1} \cdots x_n^{i_n}$ such that $i_1 + \cdots + i_n = d$. It is denoted by $\deg(f)$.

Definition 2.5.5 (Monic polynomial). A polynomial for which the highest degree term has coefficient 1.

Definition 2.5.6 (Split). A polynomial $f \in R[x]$ is said to split (over R) if it can be factorized as a product of linear factors, i.e. there exist elements $\lambda, r_1, \dots r_n$ such that

$$f = \lambda(x - r_1) \cdots (x - r_n), \tag{2.55}$$

where $n = \deg(f)$. The numbers $r_i \in R$ are called the **roots** of the polynomial.

Theorem 2.5.7 (Fundamental theorem of algebra). *Consider a $\mathbb{C}$-valued polynomial of degree ≥ 1. It has at least one root in $\mathbb{C}$.*

Corollary 2.5.8. If $f \in \mathbb{C}[x]$ is a monic polynomial with $\deg(f) \geq 1$, it can be factorized as follows:

$$f(x) = \prod_{i=1}^{k} (x - a_i)^{n_i},$$

where $a_1, \dots, a_k \in \mathbb{C}$ and $n_1, \dots, n_k \in \mathbb{N}$.

Formula 2.5.9 (Vieta). Consider a polynomial of degree n. By the fundamental theorem of algebra, this polynomial has n complex roots. Vieta's formulas relate the coefficients of the polynomial to its roots r_i:

$$\sum_{1 \leq i_1 \leq \dots \leq i_k \leq n} \prod_{j=1}^{k} r_{i_j} = (-1)^k \frac{a_{n-k}}{a_n}, \tag{2.56}$$

where $k \leq n$. For $k = 1$ and $k = n$, this gives the well-known sum and product formulas:

$$r_1 + r_2 + \cdots + r_n = -\frac{a_{n-1}}{a_n}, \tag{2.57}$$

$$r_1 r_2 \cdots r_n = (-1)^n \frac{a_0}{a_n}. \tag{2.58}$$

Example 2.5.10. For quadratic polynomials

$$f(x) = ax^2 + bx + c, \tag{2.59}$$

one recovers the following well-known formulas:

$$r_1 + r_2 = -\frac{b}{a}, \tag{2.60}$$

$$r_1 r_2 = \frac{c}{a}. \tag{2.61}$$

2.5.2 Field extensions

Definition 2.5.11 (Field extension). Let K be a field. A field extension of K is a field L such that $K \subseteq L$ and such that the operations of K are the restrictions of those in L.

Notation 2.5.12. A field extension L of K is often denoted by L/K.

Notation 2.5.13 (Generation). Consider a field extension L/K together with a subset $X \subseteq L$. The smallest field extension of K that contains X is denoted by $K(X)$. L is said to be **simple** if $L = K(a)$ for some $a \in L$.

Definition 2.5.14 (Degree). If L/K is a field extension, L can be given the structure of a *K-vector space* (see Definition 22.1.1). The dimension of this vector space is called the degree of the extension K. It is often denoted by $[L : K]$. If the degree is finite, the extension is also said to be **finite**.

The following property is very similar to the tower rule for finite groups (Property 2.2.16).

Property 2.5.15 (Tower property). Consider two field extensions M/L and L/K.

$$[M : K] = [M : L][L : K] \tag{2.62}$$

Remark. Galois theory will show that this is no mere coincidence.

Definition 2.5.16 (Number field). A finite extension of the rational numbers $\mathbb{Q}$.

2.5.3 Algebraicity

Definition 2.5.17 (Algebraic element). Consider a field extension L/K. An element $x \in L$ for which there exists a nontrivial polynomial $p \in K[x]$ such that $p(x) = 0$, is said to be algebraic, otherwise it is said to be **transcendental**.

Definition 2.5.18 (Algebraic dependence). Consider a commutative ring R and a subring $S \subset R$. An element $r \in R$ is said to be algebraically dependent on S if it is the root of a polynomial in $S[x]$. If r is the root of a monic polynomial in $S[x]$, it is said to be **integrally dependent** on S.

Remark 2.5.19. Since every nonzero element in a field is invertible, one can always turn a general polynomial into a monic polynomial. Hence, over a field, the concepts of algebraic and integral dependence coincide.

Definition 2.5.20 (Algebraic integer). Let K be a number field. An algebraic integer of K is an element satisfying a polynomial with integer coefficients.

Definition 2.5.21 (Ideal class group). Consider the ring of algebraic integers $\mathcal{O}_K$ for a number field K (or *Dedekind domain*). The fractional ideals of (Definition 2.4.41) of $\mathcal{O}_K$ form a group $\mathcal{I}_K$ with the principal fractional ideals forming a subgroup $\mathcal{P}_K$. The

quotient group

$$\mathcal{C}_K := \mathcal{J}_K / \mathcal{P}_K \tag{2.63}$$

is called the ideal class group of K. The order of $\mathcal{C}_K$ is called the **class number** h_K.

Property 2.5.22 (UFD). The ring of integers $\mathcal{O}_K$ is a UFD if and only if $h_K = 1$.

Property 2.5.23 (Abelian groups). Every Abelian group can be obtained as the ideal class group over a *Dedekind domain*.

Definition 2.5.24 (Algebraic extension). An extension L/K such that every element of L is algebraic over K.

Property 2.5.25. Every finite extension is algebraic.

Definition 2.5.26 (Algebraic closure). A field is said to be **algebraically closed** if every polynomial over it has a root. This implies that every polynomial splits. An algebraic closure of a field K is an algebraic extension $\overline{K}/K$ such that $\overline{K}$ is algebraically closed.

Definition 2.5.27 (Splitting field). Consider a field K and a polynomial $f \in K[x]$. The smallest field extension of K over which f splits is called the splitting field of f.

Property 2.5.28. A splitting field of a polynomial of degree n has degree at most $n!$.

Example 2.5.29. Theorem 2.5.7 implies that $\mathbb{C}$ is an algebraic closure of $\mathbb{R}$. In fact, it is also the splitting field over $\mathbb{R}$.

Definition 2.5.30 (Minimal polynomial). Consider a field extension L/K with an element $a \in L$ that is algebraic over K. The monic polynomial $f^a \in K[x]$ of lowest order such that $f^a(a) = 0$ is called the minimal polynomial of a. This polynomial is obtained as the generator[10] of the kernel ideal of the **evaluation morphism**

$$K[x] \to L : f \mapsto f(a)\,. \tag{2.64}$$

Property 2.5.31. Let f^a be the minimal polynomial of an algebraic element $a \in L$. If $f(a) = 0$ for some polynomial $f \in K[x]$, the minimal polynomial f^a divides f.

Corollary 2.5.32. The minimal polynomial of an algebraic element is irreducible.

Property 2.5.33. Consider two fields K, L together with an element $a \in L$. The 'polynomial ring' $K[a]$, the image of the evaluation morphism, coincides with $K(a)$ if and only if a is algebraic over K.

Definition 2.5.34 (Separable element). An element $a \in \overline{K}$ of an algebraic closure $\overline{K}/K$ is said to be separable (over K) if the minimal polynomial $f^\alpha \in K[x]$ is separable, i.e. has no multiple zeros. A subextension $L \subset \overline{K}$ is said to be separable if every element of L is separable over K.

[10] Polynomial rings in one variable are PIDs, so the kernel ideal admits a generator.

Definition 2.5.35 (Separable closure). Let K be a field with algebraic closure $\overline{K}$. The separable closure K_S is defined as follows:

$$K_S := \{x \in \overline{K} \mid x \text{ is separable over } K\} . \tag{2.65}$$

Example 2.5.36 (Separable splitting field). The splitting field of a separable polynomial is separable.

Definition 2.5.37 (Perfect field). A field K satisfying any of the following equivalent statements:

1. Every irreducible polynomial over K is separable.
2. Every algebraic extension of K is separable.
3. Every finite extension of K is separable.
4. The separable closure K_S is algebraically closed.

Example 2.5.38. All finite fields and fields of characteristic zero are perfect.

Corollary 2.5.39. For a perfect field, the algebraic closure and separable closure coincide.

Definition 2.5.40 (Normal extension). A field extension L/K with $L \subseteq \overline{K}$ is said to be **normal** if every irreducible polynomial in $K[x]$ that has a root in L splits over L.

Property 2.5.41. A field extension is normal if and only if it is the splitting field for some set of polynomials. Moreover, it is finite normal if and only if it is the splitting field of a single polynomial.

Definition 2.5.42 (Normal closure). Consider a field extension L/K. The normal closure of L is the smallest normal extension of L over K.

2.5.4 Galois extensions

Definition 2.5.43 (Galois extension). An algebraic field extension L/K such that there exists a subgroup $G \subseteq \mathsf{Aut}(L)$ satisfying

$$K \cong L^G , \tag{2.66}$$

where L^G denotes the set of G-invariants of L. This group, often denoted by $\mathsf{Gal}(L/K)$, is called the **Galois group** of the extension. If L is the splitting field of a polynomial $f \in K[x]$, the Galois group is sometimes denoted by $\mathsf{Gal}(f)$.

When L/K is finite, the Galois group is a finite group and, hence, can simply be endowed with the discrete topology. However, when the extension is not finite, one should endow the group with a topology. The **Krull topology** is defined as the *profinite topology* (see Definition 6.8.7) with respect to all finite subgroups of $\mathsf{Gal}(L/K)$.[11]

[11] One can show that every *profinite group* arises as the Galois group for some field extension.

Property 2.5.44. Consider a field extension L/K. The following statements are equivalent:

- L is a Galois extension.
- L is algebraic, normal and separable over K.

Property 2.5.45. A finite extension L/K is Galois if and only if $|\mathsf{Gal}(L/K)| = [L : K]$.

Theorem 2.5.46 (Fundamental theorem of Galois theory). *Consider a Galois extension L/K. The intermediate field extensions K'/K are in bijection with the closed subgroups of $\mathsf{Gal}(L/K)$.*

Property 2.5.47. Consider a subgroup $H \subset \mathsf{Gal}(L/K)$. The correspondence in the fundamental theorem has the following properties:

- L/L^H is Galois.
- If H is open, L^H/K is finite.
- If H is normal, L^H/K is Galois with $\mathsf{Gal}\left(L^H/K\right) \cong \mathsf{Gal}(L/K)/H$.

2.5.5 Radicals

Definition 2.5.48 (Radical extension). A simple radical extension L/K is a simple extension $L = K(a)$ such that $a^n = b$ for some $b \in K$ and $n \in \mathbb{N}_0$. A radical extension or radical series is a finite sequence of simple radical extensions.

Property 2.5.49. If L/K is radical and K' is an intermediate field extension, $\mathsf{Aut}(K'/K)$ is solvable.

Theorem 2.5.50 (Galois). *A polynomial $f \in K[x]$, where $\mathsf{char}(K)$ does not divide $\mathsf{deg}(f)$, can be solved by radicals if and only if its Galois group $\mathsf{Gal}(f)$ is solvable (Definition 2.2.10).*

2.5.6 Nullstellensatz

For the geometric incarnation, see Chapter 10.

Theorem 2.5.51 (Weak Nullstellensatz). *Let K be an algebraically closed field and consider the polynomial ring $R = K[x_1, \dots, x_n]$. An ideal $I \subset R$ is maximal if and only if it is of the form*

$$(x_1 - a_1, \dots, x_n - a_n)$$

with $a_i \in K$ for all $i \leq n$.

Corollary 2.5.52. There exists a bijection between K^n and the set of maximal ideals of $K[x_1, \dots, x_n]$.

Corollary 2.5.53. Consider a collection of polynomials $\{f_i\}_{i \in I} \subset K[x_1, \dots, x_n]$. If these polynomials do not have a common root, the ideal they generate is the unit ideal.

Chapter 3

Category theory

For the general theory of categories, the classical reference is Mac Lane (2013) or the more modern account Riehl (2017). The main reference for (co)end calculus is Loregian (2021), while a thorough introduction to the theory of enrichment is given in Kelly (1982). For the theory of higher categories and its applications to topology and algebra, the reader is referred to the book by Baez and May (2009). A good starting point for bicategories (and more) is the paper by Leinster (1998).

Chapter Contents

3.1 Categories **86**

3.2 Functors **87**

3.2.1 Natural transformations 88

3.2.2 Equivalences 90

3.2.3 Stuff, structure and property 91

3.2.4 Adjunctions 91

3.3 General constructions **93**

3.3.1 Fibred categories ♣ 94

3.3.2 Monads 96

3.4 Morphisms and diagrams **99**

3.4.1 Morphisms 99

3.4.2 Initial and terminal objects 100

3.4.3 Lifts 101

3.4.4 Limits and colimits 103

3.5 Internal structures **111**

3.5.1 Groupoids 113

3.6 Lawvere theories ♣ **114**

3.7 Operad theory ♣ 115
3.7.1 Operads 115
3.7.2 Algebraic topology 116

3.1 Categories

Definition 3.1.1 (Category). A category $\mathbf{C}$ consists of two collections, the objects $\mathsf{ob}(\mathbf{C})$ and the morphisms $\mathsf{hom}(\mathbf{C})$ or $\mathsf{mor}(\mathbf{C})$, that satisfy the following conditions:

1. **Source and target**: For every morphism $f \in \mathsf{hom}(\mathbf{C})$, there exist two objects $s(f), t(f) \in \mathsf{ob}(\mathbf{C})$, the source and the target. The collection of all morphisms with source x and target y is denoted by $\mathsf{Hom}_{\mathsf{C}}(x, y)$ or $\mathbf{C}(x, y)$.

2. **Composition**: For every two morphisms $f \in \mathbf{C}(y, z)$ and $g \in \mathbf{C}(x, y)$, the composite $f \circ g$ is an element of $\mathbf{C}(x, z)$. Moreover, composition is required to be associative.

3. **Identity**: For every $x \in \mathsf{ob}(\mathbf{C})$, there exists an identity morphism $\mathbb{1}_x \in \mathbf{C}(x, x)$. Identity morphisms are required to satisfy $f \circ \mathbb{1}_x = f = \mathbb{1}_y \circ f$ for all morphisms $f \in \mathbf{C}(x, y)$.

Remark 3.1.2. One technically does not need to consider objects as a separate notion, since every object can be identified with its identity morphism (which exists by definition) and, hence, one can work solely with morphisms. It should be noted that for higher categories this remark can be omitted since the objects are always regarded as 0-morphisms in that context.

Definition 3.1.3 (Subcategory). Consider two categories $\mathbf{C}$ and $\mathbf{S}$. $\mathbf{S}$ is called a subcategory of $\mathbf{C}$ if $\mathsf{ob}(\mathbf{S})$ and $\mathsf{hom}(\mathbf{S})$ are subcollections of $\mathsf{ob}(\mathbf{C})$ and $\mathsf{hom}(\mathbf{C})$, respectively.

A subcategory is said to be **full** if for every two objects $x, y \in \mathsf{ob}(\mathbf{S})$:

$$\mathbf{S}(x, y) = \mathbf{C}(x, y)\,. \tag{3.1}$$

A subcategory is said to be **wide** or **lluf** if it contains all objects:

$$\mathsf{ob}(\mathbf{S}) = \mathsf{ob}(\mathbf{C})\,. \tag{3.2}$$

Definition 3.1.4 (Replete subcategory). A subcategory $\mathbf{S} \subseteq \mathbf{C}$ such that if $x \in \mathsf{ob}(\mathbf{S})$ and $f : x \cong y \in \mathsf{hom}(\mathbf{C})$, then also $y \in \mathsf{ob}(\mathbf{S})$ and $f \in \mathsf{hom}(\mathbf{S})$.

Definition 3.1.5 (Small category). A category $\mathbf{C}$ for which both $\mathsf{ob}(\mathbf{C})$ and $\mathsf{hom}(\mathbf{C})$ are sets. A category $\mathbf{C}$ is said to be **locally small** if for every two objects $x, y \in \mathsf{ob}(\mathbf{C})$ the collection of morphisms $\mathbf{C}(x, y)$ is a set. A category *equivalent* (see further down below) to a small category is said to be **essentially small**.

Definition 3.1.6 (Opposite category). Let $\mathbf{C}$ be a category. The opposite category $\mathbf{C}^{\mathrm{op}}$ is constructed by reversing all arrows in $\mathbf{C}$, i.e. a morphism in $\mathbf{C}^{\mathrm{op}}(x, y)$ is a morphism in $\mathbf{C}(y, x)$.

Property 3.1.7 (Involution). From the definition of the opposite category it readily follows that *op* is an involution:

$$(\mathbf{C}^{\mathrm{op}})^{\mathrm{op}} = \mathbf{C}\,. \tag{3.3}$$

3.2 Functors

Definition 3.2.1 (Covariant functor). Let $\mathbf{C}, \mathbf{D}$ be categories. A (covariant) functor is an assignment $F : \mathbf{C} \to \mathbf{D}$ satisfying the following conditions:

1. F maps every object $x \in \mathsf{ob}(\mathbf{C})$ to an object $Fx \in \mathsf{ob}(\mathbf{D})$.
2. F maps every morphism $\phi \in \mathbf{C}(x, y)$ to a morphism $F\phi \in \mathbf{D}(Fx, Fy)$.
3. F preserves identities, i.e. $F\mathbb{1}_x = \mathbb{1}_{Fx}$.
4. F preserves compositions, i.e. $F(\phi \circ \psi) = F\phi \circ F\psi$.

Remark 3.2.2 (Category of categories). Small categories, together with (covariant) functors between them, form a category $\mathbf{Cat}$. The restriction to small categories is important since otherwise one would obtain an inconsistency similar to *Russell's paradox*. In certain foundations one can also consider the 'category' $\mathbf{CAT}$ of all categories, but this would not be a large category anymore. It would be something like a 'very large' category.

Definition 3.2.3 (Contravariant functor). Let $\mathbf{C}, \mathbf{D}$ be categories. A contravariant functor is an assignment $F : \mathbf{C} \to \mathbf{D}$ satisfying the following conditions:

1. F maps every object $x \in \mathsf{ob}(\mathbf{C})$ to an object $Fx \in \mathsf{ob}(\mathbf{D})$.
2. F maps every morphism $\phi \in \mathbf{C}(x, y)$ to a morphism $F\phi \in \mathbf{D}(Fy, Fx)$.
3. F preserves identities, i.e. $F\mathbb{1}_x = \mathbb{1}_{Fx}$.
4. F reverses compositions, i.e. $F(\phi \circ \psi) = F\psi \circ F\phi$.

A contravariant functor can also be defined as a covariant functor from the opposite category and, accordingly, from now on the word 'covariant' will be dropped when talking about functors.

Definition 3.2.4 (Endofunctor). A functor of the form $F : \mathbf{C} \to \mathbf{C}$.

Definition 3.2.5 (Presheaf). A functor $G : \mathbf{C}^{\mathrm{op}} \to \mathbf{Set}$. The collection of all presheaves on a (small) category $\mathbf{C}$ forms a category $\mathbf{Psh}(\mathbf{C})$. This is sometimes also denoted by $\widehat{\mathbf{C}}$.

Example 3.2.6 (Hom-functor). Let $\mathbf{C}$ be a locally small category. Every object $x \in \mathsf{ob}(\mathbf{C})$ induces a functor $h^x : \mathbf{C} \to \mathbf{Set}$ defined as follows:

- h^x maps every object $y \in \mathsf{ob}(\mathbf{C})$ to the set $\mathbf{C}(x, y)$.
- For all $y, z \in \mathsf{ob}(\mathbf{C})$, h^x maps every morphism $f \in \mathbf{C}(y, z)$ to the fucntion

$$f \circ - : \mathbf{C}(x, y) \to \mathbf{C}(x, z) : g \mapsto f \circ g\,.$$

Remark 3.2.7. The contravariant hom-functor h_x is defined by replacing $\mathbf{C}(x, -)$ with $\mathbf{C}(-, x)$ and replacing postcomposition with precomposition.

Definition 3.2.8 (Faithful functor). A functor $F : \mathbf{C} \to \mathbf{D}$ for which the map

$$\mathbf{C}(x, y) \to \mathbf{D}(Fx, Fy)$$

is injective for all objects $x, y \in \mathsf{ob}(\mathbf{C})$.

Definition 3.2.9 (Full functor). A functor $F : \mathbf{C} \to \mathbf{D}$ for which the map

$$\mathbf{C}(x, y) \to \mathbf{D}(Fx, Fy)$$

is surjective for all objects $x, y \in \mathsf{ob}(\mathbf{C})$.

Definition 3.2.10 (Embedding). A fully faithful functor.

Definition 3.2.11 (Concrete category). A category equipped with an embedding into $\mathbf{Set}$. The objects of such categories can be interpreted as sets with additional structure.

Definition 3.2.12 (Essentially surjective functor). A functor $F : \mathbf{C} \to \mathbf{D}$ such that for every object $y \in \mathsf{ob}(\mathbf{D})$, there exists an object $x \in \mathsf{ob}(\mathbf{C})$ with $Fx \cong y$.

Definition 3.2.13 (Profunctor[1]). A functor of the form $F : \mathbf{D}^{\mathrm{op}} \times \mathbf{C} \to \mathbf{Set}$. Such a functor is often denoted by $F : \mathbf{C} \nrightarrow \mathbf{D}$.[2] Elements of the set $F(x, y)$ are called **heteromorphisms** (between x and y).

It should be noted that presheaves on $\mathbf{C}$ are profunctors of the form $1 \nrightarrow \mathbf{C}$.

Definition 3.2.14 (Reflection). A functor $F : \mathbf{C} \to \mathbf{D}$ is said to **reflect** a property if whenever the property holds for Fc, it also holds for $c \in \mathbf{C}$. (Here, c could also be a morphism.)

3.2.1 Natural transformations

Definition 3.2.15 (Natural transformation). Let $F, G : \mathbf{C} \to \mathbf{D}$ be functors. A natural transformation $\psi : F \Rightarrow G$[3] consists of a collection of morphisms satisfying the following two conditions:

[1] Sometimes called a **distributor**.

[2] This is the convention by *Borceux*. Some other authors, such as Johnstone (2014), use the opposite convention.

[3] This notation is in analogy with the general notation for *2-morphisms*. See Section 14.1 for more information.

1. For every object $x \in \mathsf{ob}(\mathbf{C})$, there exists a morphism $\psi_x : Fx \to Gx$ in $\mathsf{hom}(\mathbf{D})$. This morphism is called the **component** of ψ at x. (It is often said that ψ_x **is natural in** x.)

2. For every morphism $f \in \mathbf{C}(x, y)$, the diagram below commutes:

$$\begin{tikzcd} Fx \arrow[r, "Ff"] \arrow[d, "\psi_x"'] & Fy \arrow[d, "\psi_y"] \\ Gx \arrow[r, "Gf"'] & Gy \end{tikzcd}$$

Definition 3.2.16 (Functor category). Consider two categories $\mathbf{C}$ and $\mathbf{D}$, where $\mathbf{C}$ is small. The functors $F : \mathbf{C} \to \mathbf{D}$ form the objects of a category with the natural transformations as morphisms. This category is denoted by $[\mathbf{C}, \mathbf{D}]$ or $\mathbf{D}^{\mathbf{C}}$ (the latter is a generalization of Notation 1.5.5).

Definition 3.2.17 (Dinatural transformation). Consider two profunctors $F, G : \mathbf{C} \nrightarrow \mathbf{C}$ or, more generally, two functors $F, G : \mathbf{C}^{\mathrm{op}} \times \mathbf{C} \to \mathbf{D}$. A dinatural transformation is a family of morphisms

$$\eta_x : F(x, x) \to G(x, x)$$

that make Diagram 3.1 commute for every morphism $f : y \to x$.

$$\begin{tikzcd} & F(x,y) \arrow[dl, "F(\mathbb{1}_x, f)"'] \arrow[dr, "F(f, \mathbb{1}_y)"] & \\ F(x,x) \arrow[d, "\eta_x"'] & & F(y,y) \arrow[d, "\eta_y"] \\ G(x,x) \arrow[dr, "G(f, \mathbb{1}_x)"'] & & G(y,y) \arrow[dl, "G(\mathbb{1}_y, f)"] \\ & G(y,x) & \end{tikzcd}$$

Figure 3.1: Dinatural transformation.

Definition 3.2.18 (Representable functor). Let $\mathbf{C}$ be a locally small category. A functor $F : \mathbf{C} \to \mathbf{Set}$ is said to be representable if there exists an object $x \in \mathsf{ob}(\mathbf{C})$ such that F is naturally isomorphic to h^x. The pair $(x, \psi : F \Rightarrow h^x)$ is called a **representation** of F.

Theorem 3.2.19 (Yoneda lemma). *Let $\mathbf{C}$ be a locally small category and let $F : \mathbf{C} \to \mathbf{Set}$ be*

a functor. For every object $x \in \mathsf{ob}(\mathbf{C})$, there exists a natural isomorphism[4]

$$\eta_x : \mathsf{Nat}(h^x, F) \to Fx : \psi \mapsto \psi_x(\mathbb{1}_x)\,. \tag{3.4}$$

Corollary 3.2.20 (Yoneda embedding). When F is another hom-functor h^y, the following result is obtained:

$$\mathsf{Nat}(h^x, h^y) \cong \mathbf{C}(y, x)\,. \tag{3.5}$$

Note that y appears in the first argument on the right-hand side.

Let $\mathbf{C}(f, -)$ denote the natural transformation corresponding to the morphism $f \in \mathbf{C}(y, x)$. The functor h^-, mapping an object $x \in \mathsf{ob}(\mathbf{C})$ to its hom-functor $\mathbf{C}(x, -)$ and a morphism $f \in \mathbf{C}(y, x)$ to the natural transformation $\mathbf{C}(f, -)$, can also be interpreted as a covariant functor $G : \mathbf{C}^{\mathrm{op}} \to \mathbf{Set}^{\mathbf{C}}$. This way, the Yoneda lemma can be seen to give rise to an embedding h^- of $\mathbf{C}^{\mathrm{op}}$ in the functor category $\mathbf{Set}^{\mathbf{C}}$.

As usual, all of this can be done for contravariant functors. This gives an embedding

$$\mathcal{Y} := h_- : \mathbf{C} \hookrightarrow \mathbf{Psh}(\mathbf{C})\,, \tag{3.6}$$

called the Yoneda embedding.

Definition 3.2.21 (Local object). Consider a collection of morphisms $S \subseteq \mathsf{hom}(\mathbf{C})$. An object $c \in \mathsf{ob}(\mathbf{C})$ is said to be S-local if the Yoneda embedding $\mathcal{Y}c$ maps morphisms in S to isomorphisms in $\mathbf{Set}$. A morphism $f \in \mathsf{hom}(\mathbf{C})$ is said to be S-local if its image under the Yoneda embedding of every S-local object is an isomorphism in $\mathbf{Set}$.

3.2.2 Equivalences

Definition 3.2.22 (Equivalence of categories). Two categories $\mathbf{C}, \mathbf{D}$ are said to be equivalent if there exist functors $F : \mathbf{C} \to \mathbf{D}$ and $G : \mathbf{D} \to \mathbf{C}$ such that $F \circ G$ and $G \circ F$ are naturally isomorphic to the identity functors.

A weaker notion is that of a **weak equivalence**. Two categories $\mathbf{C}, \mathbf{D}$ are said to be weakly equivalent if there exist functors $F : \mathbf{C} \to \mathbf{D}$ and $G : \mathbf{D} \to \mathbf{C}$ that are fully faithful and essentially surjective. Assuming the axiom of choice, every weak equivalence is also a (strong) equivalence (in fact this statement is equivalent to the axiom of choice).

Definition 3.2.23 (Skeletal category). A category in which every isomorphism is necessarily an identity morphism. The **skeleton** of a category is an equivalent skeletal category (often taken to be a subcategory by choosing a representative from every isomorphism class).

If one does not assume the axiom of choice, the skeleton is merely a weakly equivalent skeletal category.

[4] Here, the fact that $\mathsf{Nat}(h^-, -)$ can be seen as a functor $\mathbf{Set}^{\mathbf{C}} \times \mathbf{C} \to \mathbf{Set}$ is used.

Definition 3.2.24 (Decategorification). Let $\mathbf{C}$ be an (essentially) small category. The set of isomorphism classes of $\mathbf{C}$ is called the decategorification of $\mathbf{C}$. This amounts to a functor $\mathsf{Decat} : \mathbf{Cat} \to \mathbf{Set}$.

3.2.3 Stuff, structure and property

To classify properties of objects and the *forgetfulness* of functors, it is interesting to make a distinction between stuff, structure and property. Consider for example a group. This is a set (*stuff*) equipped with a number of operations (*structure*) that obey some relations (*properties*).

Using these notions one can classify forgetful functors in the following way:

- A functor forgets nothing if it is an equivalence of categories.
- A functor forgets at most properties if it is fully faithful.
- A functor forgets at most structure if it is faithful.
- A functor forgets at most stuff if it is just a functor.

@@ COMPLETE (see e.g. nLab or the paper "Why surplus structure is not superfluous" by Nicholas Teh et al.) @@

3.2.4 Adjunctions

Definition 3.2.25 (Hom-set adjunction). Let $F : \mathbf{C} \to \mathbf{D}$ and $G : \mathbf{D} \to \mathbf{C}$ be two functors. These functors form a (hom-set) adjunction $F \dashv G$ if the following isomorphism is natural in both x and y:

$$\Phi_{x,y} : \mathbf{D}(Fx, y) \cong \mathbf{C}(x, Gy)\,. \tag{3.7}$$

The functor F (resp. G) is called the left (resp. right) adjoint and the image of a morphism under either of the natural isomorphisms is called the adjunct of the other morphism.[5].

Notation 3.2.26. An adjunction $F \dashv G$ between categories $\mathbf{C}, \mathbf{D}$ is often denoted by

$$\begin{tikzcd} \mathbf{D} \arrow[r, shift right, "G"'] \arrow[r, phantom, "\perp"] & \mathbf{C} \arrow[l, shift right, "F"'] \end{tikzcd}\,.$$

Definition 3.2.27 (Unit-counit adjunction). Let $F : \mathbf{C} \to \mathbf{D}$ and $G : \mathbf{D} \to \mathbf{C}$ be two functors. These functors form a unit-counit adjunction if there exist natural transfor-

[5]The terms 'adjunct' and 'adjoint' are sometimes used interchangeably (cf. French versus Latin).

mations

$$\varepsilon : F \circ G \Rightarrow \mathbb{1}_{\mathbf{D}} \tag{3.8}$$

$$\eta : \mathbb{1}_{\mathbf{C}} \Rightarrow G \circ F \tag{3.9}$$

such that the following compositions are identity morphisms:

$$F \xrightarrow{F\eta} FGF \xrightarrow{\varepsilon F} F\,, \tag{3.10}$$

$$G \xrightarrow{\eta G} GFG \xrightarrow{G\varepsilon} G\,. \tag{3.11}$$

These identities are sometimes called the **triangle** or **zig-zag identities** (the latter results from the shape of the associated *string diagram*). The transformations η and ε are called the **unit** and **counit**, respectively.

Property 3.2.28 (Equivalence). Every hom-set adjunctions induces a unit-counit adjunction where the counit ε_y is obtained as the adjunct $\Phi^{-1}_{Gy,y}(\mathbb{1}_{Gy})$ of the identity morphism on $Gy \in \mathsf{ob}(\mathbf{C})$ and the unit η_x is given by the adjunct $\Phi_{c,Fc}(\mathbb{1}_{Fx})$ of the identity morphism at $Fx \in \mathsf{ob}(\mathbf{D})$.

Conversely, every unit-counit adjunction induces a hom-set adjunction. The (right) adjunct of a morphism $f : Fx \to y$ is given by the composition

$$\tilde{f} := Gf \circ \eta_x : x \to (G \circ F)x \to Gy \tag{3.12}$$

and the (left) adjunct of a morphism $\tilde{g} : x \to Gy$ is given by:

$$g := \varepsilon_y \circ F\tilde{g} : Fx \to (F \circ G)y \to y\,. \tag{3.13}$$

Definition 3.2.29 (Reflective subcategory). A full subcategory is said to be reflective (resp. coreflective) if the inclusion functor admits a left (resp. right) adjoint.

Property 3.2.30 (Adjoint equivalence). Any equivalence of categories is part of an adjoint equivalence, i.e. an adjunction for which the unit and counit morphisms are invertible.

Property 3.2.31. Given an adjunction, one obtains an adjoint equivalence by restricting to the full subcategories on which the unit and counit become isomorphisms.

Adjunctions can also be defined through a third alternative. This links the definition to universal properties.

Definition 3.2.32 (Universal morphism). Consider a functor $F : \mathbf{C} \to \mathbf{D}$ and an object $d \in \mathsf{ob}(\mathbf{D})$. A universal morphism from d to F is a pair $(c, f : d \to Fc)$ such that all other morphisms $f' : d \to Fc'$ factor uniquely through f by the image of a morphism in $\mathbf{C}$.

Alternative Definition 3.2.33 (Adjoint functor). A functor $F : \mathbf{C} \to \mathbf{D}$ is a left adjoint if for each object $c \in \mathsf{ob}(\mathbf{C})$ there exists a universal morphism from F to c. A functor G :

$\mathbf{C} \to \mathbf{D}$ is a right adjoint if for each object $d \in \mathsf{ob}(\mathbf{D})$ there exists a universal morphism from d to G.

The functor to which these functors are adjoint can be recovered as that mapping objects to the object-part of their universal morphism.

3.3 General constructions

Definition 3.3.1 (Dagger category). A category equipped with a contravariant involutive endofunctor, this functor is often denoted by $\dagger : \mathbf{C} \to \mathbf{C}$, similar to the adjoint operator for Hermitian matrices.

Remark 3.3.2. The concept of a dagger structure allows the usual definition of **unitary** and **self-adjoint** morphisms, i.e. morphism satisfying

$$f^\dagger = f^{-1} \qquad \text{or} \qquad f^\dagger = f\,. \tag{3.14}$$

Definition 3.3.3 (Comma category). Let $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ be three categories and let $F : \mathbf{A} \to \mathbf{C}$ and $G : \mathbf{B} \to \mathbf{C}$ be two functors. The comma category $F \downarrow G$ is defined as follows:

- **Objects**: The triples (x, y, γ) where $x \in \mathsf{ob}(\mathbf{A}), y \in \mathsf{ob}(\mathbf{B})$ and $\gamma : Fx \to Gy$.
- **Morphisms**: The morphisms $(x, y, \gamma) \to (k, l, \sigma)$ are pairs (f, g) with $f : x \to k \in \mathsf{hom}(\mathbf{A})$ and $g : y \to l \in \mathsf{hom}(\mathbf{B})$ such that $\sigma \circ Ff = Gg \circ \gamma$.

Composition of morphisms is defined componentwise.

Definition 3.3.4 (Arrow category). The comma category of the pair of functors $(\mathbb{1}_{\mathbf{C}}, \mathbb{1}_{\mathbf{C}})$. This is equivalently the functor category $[\mathbf{2}, \mathbf{C}]$, where $\mathbf{2}$ is the **interval category/walking arrow** $\{0 \to 1\}$.

Definition 3.3.5 (Functorial factorization). A *section* (see Definition 3.4.1) of the composition functor

$$\circ : [\mathbf{3}, \mathbf{C}] \to [\mathbf{2}, \mathbf{C}]\,,$$

where $\mathbf{3}$ is the poset $\{0 \to 1 \to 2\}$.

If F is the identity functor and $G : \mathbf{1} \to \mathbf{C}$ picks out a single object, the notion of a slice category is obtained (by interchanging these choices, one can also define **coslice categories**).

Definition 3.3.6 (Slice category). Let $\mathbf{C}$ be a category and consider an object $x \in \mathsf{ob}(\mathbf{C})$. The slice category (or **overcategory**) $\mathbf{C}_{/x}$ of $\mathbf{C}$ over x is defined as follows:

- **Objects**: The morphisms in $\mathbf{C}$ with codomain x, and
- **Morphisms**: The morphisms $f \to g$ are morphisms h in $\mathbf{C}$ such that $g \circ h = f$.

By dualizing one obtains the **undercategory** of x.

3.3.1 Fibred categories ♣

Definition 3.3.7 (Fibre category). Let $\Pi : \mathbf{C} \to \mathbf{D}$ be a functor. The fibre category (of Π) over $y \in \mathsf{ob}(\mathbf{D})$ is the subcategory of $\mathbf{C}$ consisting of all objects $x \in \mathsf{ob}(\mathbf{C})$ such that $\Pi x = y$ and all morphisms $f \in \mathsf{hom}(\mathbf{C})$ such that $\Pi f = \mathbb{1}_y$. It will be denoted by $\mathbf{C}_y$.

Morphisms in $\mathbf{C}$ that are mapped to a morphism g in $\mathbf{D}$ are called **g-morphisms** and, in particular (using the identification of objects and their identity morphisms), morphisms in $\mathbf{C}_y$ are called **y-morphisms**. Similarly, **D-categories** are defined as the categories equipped with a (covariant) functor to $\mathbf{D}$. (It is not hard to see that these form a *2-category* under composition of functors that respects the $\mathbf{D}$-category structure.)

Definition 3.3.8 (Cartesian morphism). Consider a $\mathbf{D}$-category $\Pi : \mathbf{C} \to \mathbf{D}$. A morphism f in $\mathbf{C}$ is called Π-Cartesian if every Πf-morphism factors uniquely through a y-morphism, where y is the domain of Πf.

There also exists a stronger notion. A **strongly Cartesian morphism** is a morphism $f \in \mathsf{hom}(\mathbf{C})$ such that for every morphism $\varphi \in \mathsf{hom}(\mathbf{C})$ with the same target and every factorization of $\Pi\varphi$ through Πf, there exists a unique factorization of φ through f that maps to the given factorization of $\Pi\varphi$.

The following diagram, where the triangles commute, should clarify the above (technical) definitions:

$$\begin{tikzcd}
\forall x' \arrow[d, dashed, "\exists! g"'] \arrow[dr, "\forall\varphi"] & \\
x_1 \arrow[r, "f"'] & x_2
\end{tikzcd}
\quad \overset{\Pi}{\mapsto} \quad
\begin{tikzcd}
\Pi x' \arrow[d, "\forall\nu"'] \arrow[dr, "\Pi\varphi"] & \\
\Pi x_1 \arrow[r, "\Pi f"'] & \Pi x_2
\end{tikzcd}$$

The diagram for (weak) Cartesian morphisms is obtained by identifying the objects $\Pi x'$ and Πx_1, i.e. by restricting to the case $\nu = \mathbb{1}_{\Pi x_1}$.

The Cartesian morphisms are said to be **inverse images** of their projections under Π and the object x_1 is called an **inverse image** of x_2 by Πf. The Cartesian morphisms of a fibre category are exactly the isomorphisms of that category.

Definition 3.3.9 (Fibred category). A $\mathbf{D}$-category $\Pi : \mathbf{C} \to \mathbf{D}$ is called a fibred category or **Grothendieck fibration** if the following conditions are satisfied:

1. For each morphism in $\mathbf{D}$, whose codomain lies in the range of Π, and each lift of this codomain to $\mathbf{C}$, there exists at least one inverse image with the given codomain (in the weak sense).

2. The composition of two Cartesian morphisms is again Cartesian (in the weak sense).

If one instead works with strongly Cartesian morphisms, the second condition follows from the first one. However, it should be noted that, in a fibred category, a morphism is weakly Cartesian if and only if it is strongly Cartesian.

Definition 3.3.10 (Cleavage). Given a $\mathbf{D}$-category $\Pi : \mathbf{C} \to \mathbf{D}$, a cleavage is the choice of a Cartesian g-morphism $f : x \to y$ for every $y \in \mathrm{ob}(\mathbf{C})$ and morphism $g : d \to \Pi y$. A $\mathbf{D}$-category equipped with a cleavage is said to be **cloven**.

The existence of cleavage is sufficient for a category to be fibred and, conversely (assuming the axiom of choice), every fibred category admits a cleavage.

The following example can be obtained as a Grothendieck fibration with discrete fibres.

Example 3.3.11 (Discrete fibration). A functor $F : \mathbf{C} \to \mathbf{D}$ such that, for every object $x \in \mathrm{ob}(\mathbf{C})$ and every morphism $f : y \to Fx$ in $\mathbf{D}$, there exists a unique morphism $g : z \to x$ in $\mathbf{C}$ such that $Fg = f$.

Example 3.3.12 (Groupoidal fibration). If every morphism is required to be Cartesian, the notion of a groupoid(al) fibration or a **category fibred in groupoids** is obtained. The reason for this name is that every fibre is a groupoid. An equivalent definition is that the associated *pseudofunctor* (see the construction below) factors through the embedding $\mathbf{Grpd} \hookrightarrow \mathbf{Cat}$.

Property 3.3.13 (Grothendieck construction ♣). Every fibred category $\Pi : \mathbf{C} \to \mathbf{D}$ defines a *pseudofunctor*[6] $F : \mathbf{D}^{\mathrm{op}} \to \mathbf{Cat}$ that sends objects to fibre categories and arrows $f : c \to c'$ to the pullback functor $f^* : \mathbf{C}_{c'} \to \mathbf{C}_c$ constructed from a Cartesian lift of f. This pullback functor acts as follows:

- For every object $x \in \mathbf{C}_{c'}$, f^*x is the domain of the Cartesian lift of f through x.
- For every morphism $(\alpha : x \to y) \in \mathbf{C}_{c'}$ there exists a diagram of the form

$$\begin{tikzcd} f^*x \arrow[r] \arrow[d, dashed, "f^*\alpha"'] & x \arrow[d, "\alpha"] \\ f^*y \arrow[r] & y \end{tikzcd}$$

  Because the horizontal morphism are both projected to f and α is projected to the identity, there exists a unique factorization of the diagram through a morphism $f^*\alpha : f^*x \to f^*y$.

Conversely, every *pseudofunctor* gives rise to a fibred category through the Grothendieck construction $\int : [\mathbf{C}^{\mathrm{op}}, \mathbf{Cat}] \to \mathbf{Cat}/\mathbf{C}$ as follows. (These two constructions constitute a 2-equivalence of 2-categories.). Consider a *pseudofunctor* $F : \mathbf{C}^{\mathrm{op}} \to \mathbf{Cat}$. The 'bundle' $\int F$ consists of the following data:

[6] See Definition 14.1.8.

- The objects are pairs (x, y) with $x \in \mathsf{ob}(\mathbf{C})$ and $y \in \mathsf{ob}(Fx)$.
- The morphisms $(x, y) \to (x', y')$ are pairs $\big(f : x \to x', \alpha : y \to Ff(y')\big)$.

Given a cleavage, the morphisms of the Grothendieck construction are exactly the factorizations of f-morphisms through the canonical lifting of f in the cleavage.

Property 3.3.14 (Functors). A *pseudofunctor* is a functor if and only if the cleavage of the associated fibred category is **split(ting)**, i.e. it contains all identities and is closed under composition.

Definition 3.3.15 (Category of elements). Consider a presheaf $F : \mathbf{C}^{\mathrm{op}} \to \mathbf{Set}$. Its category of elements $\mathsf{El}(F)$ or $\int_{\mathbf{C}} F$ is defined as the comma category $(\mathcal{Y} \downarrow !_F)$, where $!_F : * \to [\mathbf{C}^{\mathrm{op}}, \mathbf{Set}]$ sends the unique object to F itself. Equivalently, it is the category with objects the pairs $(c, x) \in \mathsf{ob}(\mathbf{C}) \times Fc$ and morphisms $f \in \mathbf{C}(c, c')$ such that $c = Ff(c')$, i.e. it is the Grothendieck construction applied to F.

This category comes equipped with a canonical forgetful functor

$$\mathbf{C}_F : \mathsf{El}(F) \to \mathbf{C} : (c, x) \mapsto c\,. \tag{3.15}$$

Remark 3.3.16. The category of elements is often defined for covariant functors. To obtain that definition, one should take the opposite of the category of elements (and also take the opposite of the forgetful functor).

3.3.2 Monads

Definition 3.3.17 (Monad). A monad is a triple (T, μ, η) where $T : \mathbf{C} \to \mathbf{C}$ is an endofunctor and $\mu : T^2 \to T, \eta : \mathbb{1}_{\mathbf{C}} \to T$ are natural transformations satisfying the following (coherence) conditions:

1. As natural transformations from T^3 to T:

$$\mu \circ T\mu = \mu \circ \mu_T\,. \tag{3.16}$$

2. As natural transformations from T to itself:

$$\mu \circ T\eta = \mu \circ \eta_T = \mathbb{1}\,. \tag{3.17}$$

These conditions say that a monad is a monoid (Definition 2.1.2) in the category $\mathbf{End}_{\mathbf{C}}$ of endofunctors on $\mathbf{C}$. Accordingly, η and μ are often called the **unit** and **multiplication** maps.

Example 3.3.18 (Adjunction). Every adjunction $F \dashv G$, with unit ε and counit η, induces a monad of the form $(GF, G\varepsilon F, \eta)$.

Definition 3.3.19 (Algebra over a monad[7]). Consider a monad (T, μ, η) on a category $\mathbf{C}$. An algebra over T or T-algebra is a couple (x, κ), where $x \in \mathrm{ob}(\mathbf{C})$ and $\kappa : Tx \to x$, such that the following conditions are satisfied:

1. $\kappa \circ T\kappa = \kappa \circ \mu_x$, and
2. $\kappa \circ \eta_x = \mathbb{1}_x$.

Morphisms $(x, \kappa_x) \to (y, \kappa_y)$ of T-algebras are morphisms $f : x \to y$ in $\mathbf{C}$ such that $f \circ \kappa_x = \kappa_y \circ Tf$. An algebra of the form (Tx, μ_x) is said to be **free**. The object x is called the **carrier** of the algebra

Definition 3.3.20 (Eilenberg–Moore category). Given a monad T over a category $\mathbf{C}$, the Eilenberg–Moore category $\mathbf{C}^T$ is defined as the category of T-algebras.

Definition 3.3.21 (Kleisli category). Consider a monad T on a category $\mathbf{C}$. The Kleisli category $\mathbf{C}_T$ is defined as the full subcategory of $\mathbf{C}^T$ on the free T-algebras. This is equivalently the category with objects $\mathrm{ob}(\mathbf{C}_T) := \mathrm{ob}(\mathbf{C})$ and morphisms $\mathbf{C}_T(x, y) := \mathbf{C}(x, Ty)$.

Morphisms in the Kleisly category are composed in the 'obvious way':

$$f \circ_{\mathbf{C}_T} g := \mu_Z \circ Tf \circ g \tag{3.18}$$

for all $f \in \mathbf{C}_T(Y, TZ)$ and $g \in \mathbf{C}_T(X, TY)$.

Definition 3.3.22 (Monadic adjunction). Consider an adjunction $L \dashv R$ between categories $\mathbf{C}$ and $\mathbf{D}$ with the induced monad T. The natural morphism $R\varepsilon : T \circ R \Rightarrow R$ endows R with a T-algebra structure and, hence, induces a functor $\mathbf{C} \to \mathbf{D}^T$ between $\mathbf{C}$ and the Eilenberg-Moore category of T. $L \dashv R$ is said to be monadic if this functor is an equivalence.

Definition 3.3.23 (Monadic functor). A functor is said to be monadic if it admits a left adjoint such that the adjunction is monadic.

The converse of 3.3.18 is also true:

Property 3.3.24. Every monad $T : \mathbf{C} \to \mathbf{C}$ can be obtained from an adjunction. The canonical choice is the adjunction

$$\mathbf{C} \underset{U_T}{\overset{F_T}{\underset{\longrightarrow}{\overset{\longleftarrow}{\perp}}}} \mathbf{C}^T, \tag{3.19}$$

where F_T is the forgetful functor and U_T sends an object to the free T-algebra on it.

[7] A more suitable name would be "module over a monad", since these are modules over a monoid if monads are regarded as monoids in $\mathbf{End}_{\mathbf{C}}$.

The following theorem characterizes monadic functors (for more information on some of the concepts, see Section 3.4 further below).

Theorem 3.3.25 (Beck's monadicity theorem). *Consider a functor $F : \mathbf{C} \to \mathbf{D}$. This functor is monadic if and only if the following conditions are satisfied:*

- *F admits a left adjoint.*
- *F reflects isomorphisms.*
- *$\mathbf{C}$ has all coequalizers of F-split parallel pairs*[8] *and F preserves these coequalizers.*

Remark 3.3.26 (Crude monadicity theorem). A sufficient condition for monadicity is obtained by replacing the third condition above by the following weaker statement: "$\mathbf{C}$ has all coequalizers of reflexive pairs and F preserves these coequalizers."

Definition 3.3.27 (Closure operator). Consider a monad $(T : \mathbf{C} \to \mathbf{C}, \eta, \mu)$. This monad is called a closure operator or **modal operator** if the multiplication map is a natural isomorphism, i.e. if the monad is idempotent. Equivalently, it is idempotent if and only if $\eta \circ T$ is a natural isomorphism.

Given a closure operator $T : \mathbf{C} \to \mathbf{C}$, the object Tx is called the closure of $x \in \mathsf{ob}(\mathbf{C})$ and the associated morphism η_x is called the **closing map**. An object $x \in \mathsf{ob}(\mathbf{C})$ itself is said to be **T-closed** exactly if its closing map is an isomorphism.

An object $x \in \mathsf{ob}(\mathbf{C})$ is called a **modal type** if the unit $\eta_x : x \to Tx$ is an isomorphism.

Property 3.3.28. Every (co)reflective subcategory inclusion (Definition 3.2.29) induces a closure operator. Conversely, every closure operator is induced by a (co)reflective subcategory.

Remark 3.3.29 (Bicategories ♣). A monad can be defined in any bicategory as a 1-morphism $t : x \to x$ together with two 2-morphisms that satisfy conditions similar to the ones above. The above definition is then just a specific case of this more general definition in **Cat**.

In the general setting one can then also define a **module** over a monad. First of all, one can regard any object $x \in \mathsf{ob}(\mathbf{C})$ as a functor from the terminal category $\mathbf{1}$. By replacing $\mathbf{1}$ by any other category in the ordinary definition one obtains a general algebra (or module). It is this definition that readily generalizes to bicategories, i.e. a module is a 1-morphism $a : x \to y$ together with a 2-morphism that satisfies the same conditions as an algebra over a monad in **Cat**.

[8] These are parallel pairs f, g such that the images Ff, Fg under F admit a split coequalizer.

3.4 Morphisms and diagrams

3.4.1 Morphisms

Definition 3.4.1 (Section). A section of a morphism $f : x \to y$ is a right inverse, i.e. a morphism $g : y \to x$ such that $f \circ g = \mathbb{1}_y$. f itself is called a **retraction** of g and y is called a **retract** of x.

Definition 3.4.2 (Monomorphism). Let $\mathbf{C}$ be a category. A morphism $\mu \in \mathbf{C}(x, y)$ is called a monomorphism, **mono** or **monic morphism** if for every object $z \in \mathsf{ob}(\mathbf{C})$ and every two morphisms $\alpha_1, \alpha_2 \in \mathbf{C}(z, x)$ such that $\mu \circ \alpha_1 = \mu \circ \alpha_2$, one can conclude that $\alpha_1 = \alpha_2$.

Definition 3.4.3 (Epimorphism). Let $\mathbf{C}$ be a category. A morphism $\varepsilon \in \mathbf{C}(x, y)$ is called an epimorphism, **epi** or **epic morphism** if for every object $z \in \mathsf{ob}(\mathbf{C})$ and every two morphisms $\alpha_1, \alpha_2 \in \mathbf{C}(y, z)$ such that $\alpha_1 \circ \varepsilon = \alpha_2 \circ \varepsilon$, one can conclude that $\alpha_1 = \alpha_2$.

A family of morphisms $\{f_i : x_i \to y\}_{i \in I}$ is called **jointly epimorphic** if

$$\alpha_1 \circ f_i = \alpha_2 \circ f_i \tag{3.20}$$

for all $i \in I$ implies that $\alpha_1 = \alpha_2$.

Definition 3.4.4 (Split monomorphism). A morphism $f : x \to y$ that is a section of some other morphism $g : y \to x$. It can be shown that every split mono is in fact a mono and even an **absolute mono**, i.e. it is preserved by all functors.

The morphism g can be seen to satisfy the dual condition and, hence, is called a **split epimorphism**. It can be shown to be an absolute epi.

Definition 3.4.5 (Balanced category). A category in which every monic epi is an isomorphism.

Definition 3.4.6 (Reflexive pair). Two parallel morphisms $f, g : x \to y$ are said to form a reflexive pair if they have a common section, i.e. if there exists a morphism $\sigma : y \to x$ such that $f \circ \sigma = g \circ \sigma = \mathbb{1}_y$.

Definition 3.4.7 (Subobject). Let $\mathbf{C}$ be a category and let $x \in \mathsf{ob}(\mathbf{C})$ be any object. A subobject y of x is a mono $y \hookrightarrow x$.

In fact, one should work up to isomorphisms and, accordingly, the formal definition goes as follows: a subobject y of x is an isomorphism class of monos $i : y \hookrightarrow x$ in the slice category $\mathbf{C}_{/x}$.

Definition 3.4.8 (Well-powered category). A category $\mathbf{C}$ such that for every object $x \in \mathsf{ob}(\mathbf{C})$ the class of subobjects $\mathsf{Sub}(x)$ is small.

3.4.2 Initial and terminal objects

Definition 3.4.9 (Initial object). An object $\emptyset$ such that for every other object x there exists a unique morphism $\iota_x : \emptyset \to x$. If one drops the uniqueness, the notion of a **weakly initial object** is obtained.

Definition 3.4.10 (Terminal object). An object 1 such that for every other object x there exists a unique morphism $\tau_x : x \to 1$.

Property 3.4.11 (Uniqueness). If an initial (or terminal) object exists, it is unique (up to isomorphisms).

Definition 3.4.12 (Zero object). An object that is both initial and terminal. The zero object is often denoted by 0.

Property 3.4.13 (Zero morphism). From the definition of the zero object it follows that for any two objects x, y there exists a unique morphism $0_{xy} : x \to 0 \to y$.

Definition 3.4.14 (Pointed category). A category containing a zero object.

Definition 3.4.15 (Global element). Let $\mathbf{C}$ be a category with a terminal object 1. A global element of an object $x \in \mathsf{ob}(\mathbf{C})$ is a morphism $1 \to x$.

Property 3.4.16. Every global element is monic.

Definition 3.4.17 (Pointed object). An object x equipped with a global element $1 \to x$. This morphism is sometimes called the **basepoint**.

Remark 3.4.18. In the category **Set**, the elements of a set S are in one-to-one correspondence with the global elements of S. Furthermore, there is the the important property (*axiom of functional extensionality*) that two functions $f, g : S \to S'$ coincide if their values at every element $s \in S$ coincide or, equivalently, if their precompositions with global elements coincide.

However, this way of checking equality can fail in other categories. Consider for example **Grp**, the category of groups, with its zero object $0 = \{e\}$. The only morphism from this group to any other group G is the one mapping e to the unit in G. It is obvious that precomposition with this morphism says nothing about the equality of other morphisms. To recover the extensionality property from **Set**, the notion of an 'element' should be generalized:

Definition 3.4.19 (Generalized element). Let $\mathbf{C}$ be a category and consider an object $x \in \mathsf{ob}(\mathbf{C})$. For any object $y \in \mathsf{ob}(\mathbf{C})$, a morphism $y \to x$ is called a generalized element of x. These morphisms are also called **y-elements** in x or elements of **shape** y in x.

Definition 3.4.20 (Generator). Let $\mathbf{C}$ be a category. A collection of objects $\mathcal{O} \subset \mathsf{ob}(\mathbf{C})$ is called a collection of generators or **separators** for $\mathbf{C}$ if the generalized elements of

shape $\mathcal{O}$ are sufficient to distinguish between all morphisms in $\mathbf{C}$:

$$\forall x, y \in \mathsf{ob}(\mathbf{C}) : \forall f, g \in \mathbf{C}(x, y) : \big(f \neq g \implies \exists o \in \mathcal{O} : \exists h \in \mathbf{C}(o, x) : f \circ h \neq g \circ h\big). \tag{3.21}$$

Definition 3.4.21 (Well-pointed category). A category for which the terminal object is a generator.

Definition 3.4.22 (Free object). Consider a forgetful functor $U : \mathbf{C} \to \mathbf{D}$ (whatever this may mean). An object $c \in \mathsf{ob}(\mathbf{C})$ is said to be free over an object $x \in \mathsf{ob}(\mathbf{D})$ if there exists a universal morphism $\eta_x : x \to Uc$. These are the initial objects of the comma categories $x/\mathbf{U}$.

Property 3.4.23 (Free functor). Note that if the forgetful functor admits a left adjoint $F : \mathbf{D} \to \mathbf{C}$, every object in the image of F is free according to the previous definition. Moreover, if U admits a free object for every $x \in \mathsf{ob}(\mathbf{D})$, it has a left adjoint. For this reason, left adjoints to fortgetful functors are often called free functors.

Definition 3.3.19 can be generalized to endofunctors as follows:

Definition 3.4.24 (Algebra over an endofunctor). Consider an endofunctor $F : \mathbf{C} \to \mathbf{C}$. An algebra over F is a pair $(x, f : Fx \to x)$, where $x \in \mathsf{ob}(\mathbf{C})$ is often called the **carrier**.

Property 3.4.25. The category of algebras over an endofunctor is equivalent to the Eilenberg–Moore category of the **(algebraically-)free monad** it generates (if it exists).

Construction 3.4.26. This property could actually be interpreted as the definition of the free monad generated by an endofunctor. If it exists, it can be obtained as the monad induced by the free-forgetful adjunction induced by $F\mathbf{Alg} \to \mathbf{C}$.

When the free functor exists, it can be constructed as follows. Consider an endofunctor $F : \mathbf{C} \to \mathbf{C}$. The term introduction of an inductive type corresponds to a morphism $Fc \to c$, i.e. to an algebra over F. Now, algebras over F should correspond to the algebras over its free monad, the functor $F^* := U \circ \mathcal{F}$, where $U : F\mathbf{Alg} \to \mathbf{C}$ is the forgetful functor and $\mathcal{F} : \mathbf{C} \to F\mathbf{Alg}$ the free functor. The latter sends every object $d \in \mathsf{ob}(\mathbf{C})$ to the initial object of the comma category d/U, i.e. to an object $(c, \alpha : Fc \to c, \beta : d \to c)$. However, as long as $\mathbf{C}$ admits coproducts, such a triple is equivalent to a pair $(c, \gamma : d + Fc \to c)$. Since the latter is an algebra over $\mathbb{1}_{\mathbf{C}} + F$, one finds that algebras over F are equivalent to initial algebras over $\mathbb{1}_{\mathbf{C}} + F$.

Theorem 3.4.27 (Lambek). *If $F : \mathbf{C} \to \mathbf{C}$ has an initial algebra $f : Fx \to x$, then f is an isomorphism.*

3.4.3 Lifts

Definition 3.4.28 (Lifts and extensions). A lift of a morphism $f : x \to y$ along an epi $e : z \to y$ is a morphism $g : x \to z$ satisfying $f = e \circ g$. Dualizing this definition gives

the notion of extensions. (The epi/mono condition is often dropped in the literature.)

Definition 3.4.29 (Lifting property). A morphism $f : x \to y$ has the left lifting property with respect to a morphism $g : x' \to y'$ (or g has the right lifting property with respect to f) if for every commutative diagram

$$\begin{tikzcd} x \arrow[r] \arrow[d, "f"'] & x' \arrow[d, "g"] \\ y \arrow[r] \arrow[ur, dashed, "\exists\psi"] & y' \end{tikzcd}$$

there exists a morphism $\psi : y \to x'$ such that the triangles commute. If the morphism ψ is unique, then f and g are said to be **orthogonal**.

Definition 3.4.30 (Injective / projective morphisms). Consider a class of morphisms $I \subseteq \mathsf{hom}(\mathbf{C})$. A morphism $f \in \mathsf{hom}(\mathbf{C})$ is said to be I-injective (resp. I-projective) if it has the right (resp. left) lifting property with respect to all morphisms in I.

Given a set of morphisms I, the sets of I-injective and I-projective morphisms are denoted by $\mathsf{rlp}(I)$ and $\mathsf{llp}(I)$, respectively.

Definition 3.4.31 (Injective and projective objects). If $\mathbf{C}$ has a terminal object 1, an object x is called I-injective if its terminal morphism is I-injective. If $\mathbf{C}$ has an initial object, I-projective objects can be defined dually. (See Fig. 3.2.)

$$\begin{tikzcd} x \arrow[r, "g"] \arrow[d, "\forall f \in I"'] & i \\ y \arrow[ur, dashed, "\exists\psi"'] & \end{tikzcd}$$

(a) Injective object i.

$$\begin{tikzcd} & x \arrow[d, "\forall f \in I"] \\ p \arrow[r, "g"'] \arrow[ur, dashed, "\exists\psi"] & y \end{tikzcd}$$

(b) Projective object p.

Figure 3.2: Injective and projective objects.

If I is the class of monomorphisms (resp. epimorphisms), the terminology is simplified to **injective** (resp. **projective**) objects. For projective objects, this is also equivalent to requiring that the (covariant) hom-functor preserves epimorphisms.

A category $\mathbf{C}$ is said to **have enough injectives** if, for every object, there exists a monomorphism into an injective object. The category is said to **have enough projectives** if, for every object, there exists an epimorphism from a projective object.

Definition 3.4.32 (Fibrations and cofibrations). Consider a category $\mathbf{C}$ together with a class $I \subseteq \mathsf{hom}(\mathbf{C})$ of morphisms. A morphism $f \in \mathsf{hom}(\mathbf{C})$ is called an I-fibration (resp. I-cofibration) if it has the right (resp. left) lifting property with respect to all I-projective (resp. I-injective) morphisms.

3.4.4 Limits and colimits

Definition 3.4.33 (Diagram). A diagram in $\mathbf{C}$ with index category $\mathbf{I}$ is a (covariant) functor $D : \mathbf{I} \to \mathbf{C}$.

Definition 3.4.34 (Cone). Let $D : \mathbf{I} \to \mathbf{C}$ be a diagram. A cone from $c \in \mathsf{ob}(\mathbf{C})$ to D consists of a family of morphisms $\psi_i : c \to Di$ indexed by $\mathbf{I}$ such that $\psi_j = Df \circ \psi_i$ for all morphisms $f : i \to j \in \mathsf{hom}(\mathbf{I})$. (This is depicted in Fig. 3.3a.)

$$\begin{tikzcd} & c \arrow[dl, "\psi_i"'] \arrow[dr, "\psi_{f(i)}"] & \\ Di \arrow[rr, "Df"'] & & Df(i) \end{tikzcd}$$

(a) Component of cone over D.

$$\begin{tikzcd} a \arrow[rr, "f"] \arrow[dr, "\psi_i"'] & & b \arrow[dl, "\phi_i"] \\ & Di & \end{tikzcd}$$

(b) Morphism of cones.

Figure 3.3: Category of cones.

Alternative Definition 3.4.35. The above definition can be reformulated by defining an additional functor $\Delta_x : \mathbf{I} \to \mathbf{C}$ that maps every element $i \in \mathsf{ob}(\mathbf{I})$ to x and every morphism $g \in \mathsf{hom}(\mathbf{I})$ to $\mathbb{1}_x$, i.e. $\Delta : C \to [\mathbf{I}, \mathbf{C}]$ is the **diagonal functor**. The morphisms ψ_i can then be seen to be the components of a natural transformation $\psi : \Delta_x \Rightarrow D$. Hence, a cone (x, ψ) is an element of $[\mathbf{I}, \mathbf{C}](\Delta_x, D)$.

Definition 3.4.36 (Morphism of cones). Let $D : \mathbf{I} \to \mathbf{C}$ be a diagram and let (x, ψ) and (y, ϕ) be two cones over D. A morphism between these cones is a morphism of the apexes $f : x \to y$ such that the diagrams of the form 3.3b commute for all $i \in \mathsf{ob}(\mathbf{I})$. The cones over D together with these morphisms form a category $\mathbf{Cone}(D)$. In fact this can easily be seen to be the comma category $\Delta \downarrow D$.

Definition 3.4.37 (Limit). Consider a diagram $D : \mathbf{I} \to \mathbf{C}$. The limit of this diagram, denoted by $\lim D$, is (if it exists) the terminal object of the category $\mathbf{Cone}(D)$.

Remark 3.4.38. In the older literature, the name **projective limit** was sometimes used. The dual notion, a **colimit**, was often called an **inductive limit** in the older literature.

This definition leads to the following universal property.

Universal Property 3.4.39. Let $D : \mathbf{I} \to \mathbf{C}$ be a diagram. For every cone $(x, \psi) \in \mathbf{Cone}(D)$, there exists a unique morphism $f : x \to \lim D$. This defines a bijection

$$[\mathbf{I}, \mathbf{C}](\Delta_x, D) \cong \mathbf{C}(x, \lim D)\,. \tag{3.22}$$

If all (small) limits exist, the limit functor $\lim : [\mathbf{I}, \mathbf{C}] \to \mathbf{C}$ can be defined. The universal property of limits then implies that it is right adjoint to the constant functor Δ.

For diagrams in $\mathbf{Set}$, one can use the fully faithfulness of the Yoneda embedding to

obtain the following expression:

$$\lim D \cong [\mathbf{I}, \mathbf{Set}](\Delta_*, D)\,. \tag{3.23}$$

Remark 3.4.40. In Section 4.2 on enriched category theory, a generalization of the above construction (the so-called *weighted limits*) will be given that is better suited to the enriched setting and allows to express a wide variety of constructions as (weighted) limits.

Example 3.4.41 (Terminal object). The terminal object 1 is the limit of the empty diagram.

Definition 3.4.42 (Finitely complete category). A category is said to be finitely complete if it has all finite limits. If all (small) limits exist, the category is said to be **complete**. The dual notion for colimits is called **(finite) cocompleteness**.

Example 3.4.43 (Presheaf categories). All presheaf categories are both complete and cocomplete.

Definition 3.4.44 (Continuous functor). A functor that preserves all small limits.

A more restricted form is also common in the literature.

Definition 3.4.45 (Exact functor). A functor that preserves all finite limits is said to be **left exact**. Analogously, a functor that preserves all finite colimits is said to be **right exact**.

Example 3.4.46 (Hom-functors). In a locally small category every hom-functor is continuous (in fact these functors even preserve limits that are not necessarily small). This implies for example that

$$\mathbf{C}(x, \lim D) \cong \lim \mathbf{C}(x, D)\,. \tag{3.24}$$

In the case where $\mathbf{C}$ is small, one can characterize the Yoneda embedding through a universal property:

Universal Property 3.4.47 (Free cocompletion). The Yoneda embedding $\mathbf{C} \hookrightarrow \widehat{\mathbf{C}}$ turns the presheaf category $\widehat{\mathbf{C}}$ into the **free cocompletion** of $\mathbf{C}$, i.e. there exists an equivalence of categories between the functor category of cocontinuous functors $[\widehat{\mathbf{C}}, \mathbf{D}]_{\text{cont}}$ and the ordinary functor category $[\mathbf{C}, \mathbf{D}]$.

Definition 3.4.48 (Tiny object). An object in a locally small category for which the covariant hom-functor preserves small colimits. This is sometimes called a **small-projective** object since it is, in particular, projective[9].

Definition 3.4.49 (Cauchy completion). Let $\mathbf{C}$ be a small category. An important (small and full) subcategory of the free cocompletion of $\mathbf{C}$ is given by the Cauchy completion,

[9] Epimorphisms are characterized by a *pushout* (see Definition 3.4.72 further below).

i.e. the subcategory of $\widehat{\mathbf{C}}$ on the tiny objects.[10] It can be shown that the free cocompletion of the Cauchy completion coincides with the one on $\mathbf{C}$ (up to equivalence).

A category is said to be **Cauchy-complete** if it is equivalent to its Cauchy completion. It can be shown that a category is Cauchy-complete if and only if it has all small absolute colimits.

Definition 3.4.50 (Filtered category). A category in which every finite diagram admits a cocone. For regular cardinals κ, this notion can be generalized. A category is said to be κ-filtered if every diagram with less than κ arrows admits a cocone. (In this terminology, filtered categories are the same as ω-filtered categories.)

Definition 3.4.51 (Directed limit). Consider a diagram $D : \mathbf{I} \to \mathbf{C}$. The limit (resp. colimit) of D is said to be (co)directed (resp. directed) if $\mathbf{I}$ is a downward (resp. upward) directed set 1.7.15.

The following definition is a categorification of the previous one.

Definition 3.4.52 (Filtered limit). Consider a diagram $D : \mathbf{I} \to \mathbf{C}$. The limit (resp. colimit) of D is said to be (co)filtered (resp. filtered) if $\mathbf{I}$ is a cofiltered (resp. filtered) category.

Property 3.4.53. A category has all directed limits if and only if it has all filtered limits. (A dual statement holds for colimits.)

Definition 3.4.54 (Finitary functor). A functor that preserves all filtered colimits.

Definition 3.4.55 (Pro-object). A functor $F : \mathbf{I} \to \mathbf{C}$ from a cofiltered category. By composing these functors with the Yoneda embedding $\mathcal{Y} : \mathbf{C} \to [\mathbf{C}^{\mathrm{op}}, \mathbf{Set}]$, pro-objects can also be identified with cofiltered limits of representable presheaves. In conjunction with Remark 3.4.38, this clarifies the terminology.

Universal Property 3.4.56. The **procategory** $\mathbf{Pro}(\mathbf{C})$ is the universal completion of $\mathbf{C}$ under cofiltered limits. $\mathbf{Pro}(\mathbf{C})$ satisfies (cf. Universal Property 3.4.47):

- it admits all cofiltered limits, and
- if $\mathbf{D}$ admits all cofiltered limits, there is an equivalence of functor categories

$$[\mathbf{C}, \mathbf{D}] \cong \mathbf{Fin}\big(\mathbf{Pro}(\mathbf{C}), \mathbf{D}\big)\,, \tag{3.25}$$

  where the category on the right-hand side is the category of finitary functors.

Remark 3.4.57 (Ind-objects). By dualizing the above definitions, i.e. by replacing cofiltered limits by filtered colimits, the category of ind-objects $\mathbf{Ind}(\mathbf{C})$ is obtained.

Definition 3.4.58 (Compact object). An object for which the covariant hom-functor

[10] A generalization in the context of enriched categories is given by the *Karoubi envelope*.

preserves all filtered colimits. These objects are also said to be **finitely presentable**.[11]

Definition 3.4.59 (Product). Let $\mathbf{I}$ be a discrete category. The (co)limit over a diagram $D : \mathbf{I} \to \mathbf{C}$ is called a (co)product in $\mathbf{C}$.

Example 3.4.60 (Equalizer). Consider a diagram of the form

$$x \underset{g}{\overset{f}{\rightrightarrows}} y\,.$$

The limit of this diagram is called the equalizer of f and g. It consists of an object e and a morphism $\varepsilon : e \to x$ such that the following **fork** diagram

$$e \xrightarrow{\varepsilon} x \underset{g}{\overset{f}{\rightrightarrows}} y \tag{3.26}$$

is universal with respect to (e, ε). By dualizing one obtains **cofork** diagrams $x \rightrightarrows y \to z$ and their universal versions, the **coequalizers**.

Example 3.4.61 (Split coequalizer). A cofork diagram

$$x \underset{g}{\overset{f}{\rightrightarrows}} y \xrightarrow{\tau} z$$

together with a section φ of f and a section σ of τ such that $\sigma \circ \tau = g \circ \varphi$.

Definition 3.4.62 (Regular morphisms). A mono (resp. epi) is said to be regular if it arises as an equalizer (resp. coequalizer) of two parallel morphisms.

Although not all categories are balanced (Definition 3.4.5), the following property does hold in any category.

Property 3.4.63 (Regular bimorphism). Both monic regular epimorphisms and epic regular monomorphisms are isomorphisms.

Alternative Definition 3.4.64 (Finitely complete category). A category is said to be finitely complete if it has a terminal object and if all binary equalizers and products exist.

Definition 3.4.65 (Span). A span in a category C is a diagram of the form 3.4a. By definition of a diagram, a span in C is equivalent to a functor $S : \mathbf{\Lambda} \to \mathbf{C}$, where $\mathbf{\Lambda}$ is the category with three objects $\{-1, 0, 1\}$ and two morphisms $i : 0 \to -1$ and $j : 0 \to 1$. For this reason $\mathbf{\Lambda}$ is sometimes called the walking or universal span.

Definition 3.4.66 (Pullback). The pullback of two morphisms $f : x \to z$ and $g : y \to z$ is defined as the limit of cospan 3.4b. The full diagram characterizing the pullback, which has the form of a square, is sometimes called a **Cartesian square**.

Notation 3.4.67 (Pullback). The pullback of two morphisms $f : x \to z$ and $g : y \to z$ is often denoted by $x \times_z y$. The associated pullback square is sometimes written as in Fig. 3.5a.

[11] This name derives from the fact that modules are finitely presented if and only if their covariant hom-functor preserves direct limits (i.e. directed colimits in the context of algebra).

$$\begin{tikzcd} & s \arrow[dl, "f"'] \arrow[dr, "g"] & \\ x & & y \end{tikzcd}$$

(a) Span (category theory).

$$\begin{tikzcd} x \arrow[dr, "f"'] & & y \arrow[dl, "g"] \\ & c & \end{tikzcd}$$

(b) Cospan.

Figure 3.4: (Co)span diagrams.

$$\begin{tikzcd} x \times_z y \arrow[r] \arrow[d] \arrow[dr, phantom, "\text{pb}"] & y \arrow[d, "g"] \\ x \arrow[r, "f"'] & z \end{tikzcd}$$

(a) Pullback square.

$$\begin{tikzcd} z \arrow[r, "g"] \arrow[d, "f"'] \arrow[dr, phantom, "\text{po}"] & y \arrow[d] \\ x \arrow[r] & x \sqcup_z y \end{tikzcd}$$

(b) Pushout square.

Figure 3.5: Pullback and pushout diagrams.

Example 3.4.68 (Product). If a terminal object 1 exists, the pullback $x \times_1 y$ is equal to the product $x \times y$.

In fact, pullbacks are sometimes also called **fibred products**. The reason for this terminology is not only that products reduce to a particular case, but also that in the case of **Set** the pullbacks have a fibrewise product structure:

$$x \times_y z \cong \bigsqcup_{a \in x} f^{-1}(a) \times g^{-1}(a)\,. \tag{3.27}$$

Example 3.4.69 (Kernel pair). Consider a morphism $f : x \to y$. Its kernel pair is defined as the pullback of f along itself.

Definition 3.4.70 (Pushout). The dual notion of a pullback, i.e. the colimit of a span. See Fig. 3.5b.

Property 3.4.71. Pullbacks preserve monos and pushouts preserve epis.

Alternative Definition 3.4.72 (Epimorphism). A morphism whose cokernel pair is the identity.

Property 3.4.73 (Pasting law). Consider a diagram of the form

$$\begin{tikzcd} A \arrow[r] \arrow[d] & B \arrow[r] \arrow[d] & C \arrow[d] \\ X \arrow[r] & Y \arrow[r] & Z \end{tikzcd}$$

If the right square is a pullback diagram, the left square is a pullback diagram if and

only if the total diagram is. Dually, if the left square is a pushout diagram, the right square is a pushout diagram if and only if the total diagram is.

Property 3.4.74 (Span category ♣). Consider a category $\mathbf{C}$ with pullbacks. The category $\mathbf{Span}(\mathbf{C})$ is defined as the category with the same objects as $\mathbf{C}$ but with spans as morphisms. Composition of spans is given by pullbacks. By including morphisms of spans, $\mathbf{Span}(\mathbf{C})$ can be refined to a bicategory.

Definition 3.4.75 (Wedge). Consider a profunctor $F : \mathbf{C} \nrightarrow \mathbf{C}$. A wedge $e : w \to F$ is an object $w \in \mathsf{ob}(\mathbf{Set})$ together with a collection of morphisms $e_x : w \to F(x,x)$ indexed by $\mathbf{C}$ such that for every morphism $f : x \to y$ the following diagram commutes:

$$\begin{tikzcd}
 & w \arrow[dl, "e_x"'] \arrow[dr, "e_y"] & \\
F(x,x) \arrow[dr, "F(\mathbb{1}_x, f)"'] & & F(y,y) \arrow[dl, "F(f, \mathbb{1}_y)"] \\
 & F(x,y) &
\end{tikzcd}$$

As was the case for cones, this can be reformulated in terms of (di)natural transformations. A wedge (w, e) of a profunctor $F : \mathbf{C} \nrightarrow \mathbf{C}$ is a dinatural transformation from the constant profunctor Δ_w to F.

Definition 3.4.76 (End). The end of a profunctor $F : \mathbf{C} \nrightarrow \mathbf{C}$ is defined as the universal wedge of F. The components of the wedge are called the **projection maps** of the end. This stems from the fact that for a discrete category the end coincides with the product $\prod_{x \in \mathsf{ob}(\mathbf{C})} F(x,x)$.

This is equivalent to a definition in terms of equalizers. Consider the two canonical maps

$$\prod_{x \in \mathsf{ob}(\mathbf{C})} \mathbf{C}(x,x) \rightrightarrows \prod_{f : x \to y} \mathbf{C}(x,y)\,. \tag{3.28}$$

This diagram can be interpreted as the product of all lower halves of the wedge diagrams above. It is not hard to see that its equalizer (universally) satisfies the wedge condition for all $f \in \mathsf{hom}(\mathbf{C})$.

Notation 3.4.77 (End). The end of a profunctor $F : \mathbf{C} \nrightarrow \mathbf{C}$ is often denoted using an integral sign with subscript:

$$\int_{x \in \mathbf{C}} F(x,x)\,.$$

For the dual construction, called a **coend**, an integral sign with superscript is used.

Example 3.4.78 (Natural transformations). Consider two functors $F, G : \mathbf{C} \to \mathbf{D}$. The map $(x,y) \mapsto \mathbf{D}(Fx, Gy)$ gives a profunctor $H : \mathbf{C} \nrightarrow \mathbf{C}$. By looking at the wedge con-

dition for this profunctor, the following equality for all morphisms $f : x \to y$ can be derived:

$$\tau_y \circ Ff = Gf \circ \tau_x \,, \tag{3.29}$$

where τ is the wedge projection. Comparing this equality to Definition 3.2.15 gives

$$\mathsf{Nat}(F, G) = \int_{x \in \mathbf{C}} \mathbf{D}(Fx, Gx) \,. \tag{3.30}$$

Property 3.4.79. Using the continuity of the hom-functor (Definition 3.4.44), one can prove the following equality which can be used to turn ends into coends and vice versa:

$$\mathbf{Set}\left(\int^{x \in \mathbf{C}} F(x,x), y \right) = \int_{x \in \mathbf{C}} \mathbf{Set}\big(F(x,x), y\big) \,. \tag{3.31}$$

Using the above properties and definitions, one obtains the following two statements, called the **Yoneda reduction** and **co-Yoneda lemma**:

Property 3.4.80 (Ninja Yoneda lemma). Let $F : \mathbf{C} \to \mathbf{Set}$ be a covariant functor (similar statements hold for contravariant functors).

$$\int_{x \in \mathbf{C}} \mathbf{Set}\big(\mathbf{C}(-, x), Fx\big) \cong F \tag{3.32}$$

$$\int^{x \in \mathbf{C}} \mathbf{C}(x, -) \times Fx \cong F \tag{3.33}$$

For a generalization to the enriched setting see Property 4.2.16.

Remark 3.4.81. A common remark at this point is the comparison with the Dirac distribution (19.1.21):

$$\int \delta(x - y) f(x) = f(y) \,. \tag{3.34}$$

By interpreting the functor F as a function, the representable functors can be seen to behave as Dirac distributions.

Property 3.4.82.

$$\int_{F \in \mathbf{coPsh}(\mathbf{C})} \mathbf{Set}(Fx, Fy) \cong \mathbf{C}(x, y) \tag{3.35}$$

Definition 3.4.83 (Kan extension). Consider two functors $F : \mathbf{A} \to \mathbf{B}$ and $G : \mathbf{A} \to \mathbf{C}$. The right Kan extension of F along G is given by the universal functor $\mathsf{Ran}_G F : \mathbf{C} \to \mathbf{B}$ and natural transformation $\eta : \mathsf{Ran}_G F \circ G \Rightarrow F$:

$$\begin{tikzcd}
\mathbf{C} \arrow[dr, dashed, "\mathsf{Ran}_G F"] \arrow[dr, phantom, shift right=3ex, "\Downarrow \eta"] & \\
\mathbf{A} \arrow[u, "G"] \arrow[r, "F"'] & \mathbf{B}
\end{tikzcd}$$

The left Kan extension $\mathsf{Lan}_G F$ is obtained by dualizing this construction.

Property 3.4.84 (Complete categories). Complete (resp. cocomplete) categories admit all right (resp. left) Kan extensions.

Definition 3.4.85 (Preservation of Kan extension). A Kan extension $\mathsf{Lan}_G F$ is said to be **absolute** if every functor with the same codomain as preserves the Kan extension, i.e. a Kan extension is absolute if right whiskering it by another functor defines the Kan extension of the composition. If it is only preserved by all representable functors, the Kan extension is said to be **pointwise** or **strong**. Pointwise Kan extensions can be expressed as (co)limits, an expression is provided in Definition 4.2.17 in the *enriched setting* using (co)ends.

Alternative Definition 3.4.86 (Kan extension). The definition above gives a natural isomorphism (here given for left extension):

$$[\mathbf{A}, \mathbf{B}](F, G^* -) \cong [\mathbf{C}, \mathbf{B}](\mathsf{Lan}_G F, -) \,. \tag{3.36}$$

In the spirit of partial adjoints or partial limits, this construction defines so-called **local Kan extensions**. If local Kan extensions exist for all functors $F \in [\mathbf{A}, \mathbf{B}]$, a right adjoint $\mathsf{Ran}_G : [\mathbf{A}, \mathbf{B}] \to [\mathbf{C}, \mathbf{B}]$ to the pullback functor $G^* : F \mapsto F \circ G$ is obtained. Similarly, left Kan extension can be defined as the left adjoint to the pullback functor.

Remark 3.4.87. Using this equivalence of hom-spaces, Kan extensions can be generalized from **Cat** to any *2-category*.

Example 3.4.88 (Limit). Denote the terminal category by $\mathbf{1}$. By choosing the functor G in the definition of a right Kan extension to be the unique functor $!_{\mathbf{C}} : \mathbf{C} \to \mathbf{1}$, one obtain the universal property characterizing limits (Universal Property 3.4.39):

$$\lim F \cong \mathsf{Ran}_{!_C} F \,. \tag{3.37}$$

Similarly, colimits can be obtained as left Kan extensions.

The existence of Kan extensions can also be used to determine the existence of adjoints.

Property 3.4.89 (Adjoint functors). A functor $F : \mathbf{C} \to \mathbf{D}$ admits a left (resp. right) adjoint if and only if the right (resp. left) Kan extension of the identity functor $\mathbb{1}_{\mathbf{C}}$ along F exists. If it exists as an absolute extension, the left adjoint is given exactly by this Kan extension.

Definition 3.4.90 (Codensity monad). Consider a general functor $F : \mathbf{C} \to \mathbf{D}$. If the right Kan extension $\mathsf{Ran}_F F$ exists, it defines a monad. Functors for which this monad is the identity are said to be **codense**.[12] Left Kan extensions give, by duality, rise to *density comonads*.

Property 3.4.91 (Faithfulness). Kan extension along a fully faithful functor is itself a fully faithful functor.

[12] Codense functors are usually defined in a different way, but one can show that this is an equivalent definition (hence the name).

Property 3.4.92 (Representability). (Left) Kan extension of a representable along a functor F is equivalent to the representable of the image of F.

Property 3.4.93 (Adjoint quadruple). Consider an adjunction $F \dashv G$ and consider Kan extension of presheaves taking values in a bicomplete category. In this case precomposition with (the opposite of) one of the adjoints coincides with Kan extension along (the opposite) of the other:

$$(F^{\mathrm{op}})^* \cong \mathsf{Lan}_{G^{\mathrm{op}}}\,, \tag{3.38}$$

$$(G^{\mathrm{op}})^* \cong \mathsf{Ran}_{F^{\mathrm{op}}}\,. \tag{3.39}$$

This implies that every adjunction induces an adjoint quadruple

$$\mathsf{Lan}_{F^{\mathrm{op}}} \dashv \mathsf{Lan}_{G^{\mathrm{op}}} \dashv \mathsf{Ran}_{F^{\mathrm{op}}} \dashv \mathsf{Ran}_{G^{\mathrm{op}}}\,. \tag{3.40}$$

3.5 Internal structures

Property 3.5.1 (Eckmann–Hilton argument). A monoid internal to **Mon**, the category of monoids, is the same as a commutative monoid. (See also Property 2.1.4.)

Definition 3.5.2 (Internal category). Let $\mathcal{E}$ be a category with pullbacks. A category $\mathbf{C}$ internal to $\mathcal{E}$ consists of the following data:

- an object $C_0 \in \mathsf{ob}(\mathcal{E})$ of objects;
- an object $C_1 \in \mathsf{ob}(\mathcal{E})$ of morphisms;
- source and target morphisms $s, t \in \mathcal{E}(C_1, C_0)$;
- an 'identity-assigning' morphism $e \in \mathcal{E}(C_0, C_1)$ such that

$$s \circ e = \mathbb{1}_{C_0} \qquad\qquad t \circ e = \mathbb{1}_{C_0}\,; \tag{3.41}$$

  and

- a composition morphism $c : C_1 \times_{C_0} C_1 \to C_1$ such that the following equations hold:

$$\begin{gathered} s \circ c = s \circ \pi_1 \qquad\qquad t \circ c = t \circ \pi_2 \\ \pi_1 = c \circ (e \times_{C_0} \mathbb{1}) \qquad\qquad c \circ (\mathbb{1} \times_{C_0} e) = \pi_2 \\ c \circ (c \times_{C_0} \mathbb{1}) = c \circ (\mathbb{1} \times_{C_0} c)\,, \end{gathered} \tag{3.42}$$

  where π_1, π_2 are the canonical projections associated with the pullback $C_1 \times_{C_0} C_1$ of (s, t).

Morphisms between these categories, suitably called **internal functors**, are given by a pair of morphisms (in $\mathcal{E}$) between internal objects and morphisms, that preserve composition and identities. Internal natural transformations are defined in a similar way.

Notation 3.5.3. The (*bi*)*category* of internal categories in $\mathcal{E}$ is denoted by $\mathbf{Cat}(\mathcal{E})$. It should be noted that for $\mathcal{E} = \mathbf{Set}$, the ordinary category of small categories $\mathbf{Cat}(\mathbf{Set}) = \mathbf{Cat}$ is obtained.

Alternative Definition 3.5.4. The above definition can be reformulated in a very elegant way. An internal category in $\mathcal{E}$ is a monad in the bicategory $\mathbf{Span}(\mathcal{E})$ as shown in Fig. 3.6.

Span gives source and target maps

$$\begin{tikzcd} & C_1 \arrow[dl, "s"'] \arrow[dr, "t"] & \\ C_0 & & C_0 \end{tikzcd}$$

Multiplication gives composition

$$\begin{tikzcd} & C_1 \times_{C_0} C_1 \arrow[dl, "s \circ \pi_0"'] \arrow[dr, "t \circ \pi_1"] \arrow[dd, "c"] & \\ C_0 & & C_0 \\ & C_1 \arrow[ul, "s"] \arrow[ur, "t"'] & \end{tikzcd}$$

Unit gives identity

$$\begin{tikzcd} & C_0 \arrow[dl, "\mathbb{1}_a"'] \arrow[dr, "\mathbb{1}_a"] \arrow[dd, "e"] & \\ C_0 & & C_0 \\ & C_1 \arrow[ul, "s"] \arrow[ur, "t"'] & \end{tikzcd}$$

Figure 3.6: Internal category as a monad in $\mathbf{Span}(\mathcal{E})$.

Functors between internal categories are not the only relevant morphisms. However, when defining (co)presheaves such as the hom-functor, a problem occurs. In **Cat** there exist, by definition, maps to the ambient category **Set** (ordinary category theory has a set-theoretic foundation). However, for internal categories there does not necessarily exist a morphism $\mathbf{C} \to \mathcal{E}$. To solve this problem, one can consider a more general structure.

Definition 3.5.5 (Internal diagram). A left module over a monad in $\mathbf{Span}(\mathcal{E})$. The dual notion is better known as an **internal presheaf**. The category of internal diagrams on an internal category $\mathbf{C} \in \mathbf{Cat}(\mathcal{E})$ is denoted by $\mathcal{E}^{\mathbf{C}}$.

This can be spelled out more explicitly. An internal diagram in an internal category $\mathbf{C} \in \mathbf{Cat}(\mathcal{E})$ consists of:

1. a morphism $\gamma_0 : F_0 \to C_0$, and
2. a morphism $\mathsf{ap} : F_0 \times_{C_0} C_1 \to F_0$

satisfying:

$$\begin{aligned} \gamma_0 \circ \mathsf{ap} &= d_1 \circ \pi_{C_1} \\ \mathsf{ap} \circ (\mathbb{1}_{F_0} \times e) &= \pi_{F_0} \\ \mathsf{ap} \circ (\mathsf{ap} \times \mathbb{1}_{C_0}) &= \mathsf{ap} \circ (\mathbb{1}_{F_0} \times c) \end{aligned} \tag{3.43}$$

Alternative Definition 3.5.6 (Internal diagram). An object of the slice category $\mathbf{Cat}(\mathcal{E})_{/\mathsf{C}}$ satisfying the following pullback condition:

$$\begin{tikzcd} F_1 \arrow[r, "d_1"] \arrow[d, "\gamma_1 \equiv \pi_{C_1}"'] \arrow[dr, phantom, "\mathrm{pb}"] & F_0 \arrow[d, "\gamma_0"] \\ C_1 \arrow[r, "d_0"'] & C_0 \end{tikzcd} \tag{3.44}$$

In fact, this is a specific instance of an even more general concept. For more information on the definitions and applications, see Johnstone (2014); Mac Lane (2013).

Definition 3.5.7 (Internal profunctor). A bimodule between monads in $\mathbf{Span}(\mathcal{E})$. Together with the above definitions, this gives rise to an equivalence

$$\mathbf{Mod}(\mathbf{Span}(\mathcal{E})) \cong \mathbf{Prof}(\mathcal{E})\,. \tag{3.45}$$

Construction 3.5.8 (Internal Yoneda profunctor). Consider an internal functor $F : \mathbf{C} \to \mathbf{D}$. This functor induces two internal profunctors $F_* : \mathbf{D} \nrightarrow \mathbf{C}$ and $F^* : \mathbf{C} \nrightarrow \mathbf{D}$. For F_* (the profunctor F^* is defined similarly) the object span is defined as

$$C_0 \xleftarrow{\pi_0} C_0 \times_{D_0} D_1 \xrightarrow{t \circ \pi_1} D_0\,. \tag{3.46}$$

The action of $f \in D_1$ is given by postcomposition with f in the second factor, while the action of $g \in C_1$ is given by precomposition with Fg in the second factor and changing to the domain of g in the first factor.

It can easily be shown that the profunctors induced by an identity functor $\mathbb{1}_{\mathsf{C}}$ have an object span that corresponds to the internal category $\mathbf{C}$ with the actions given by (internal) composition. In the case of $\mathcal{E} = \mathbf{Set}$, this boils down to the hom-functor. The fact that the object span is equivalent to the category $\mathbf{C}$ is essentially the Yoneda embedding. For this reason, this profunctor is in general called the (internal) Yoneda profunctor $\mathcal{Y}(\mathbf{C})$.

3.5.1 Groupoids

Definition 3.5.9 (Groupoid). A (small) groupoid $\mathcal{G}$ is a (small) category in which all morphisms are invertible.

Example 3.5.10 (Action groupoid). Consider a set X with an action of a group G. The action groupoid $X /\!\!/ G$ is defined as the following category:

1. **Objects**: X,
2. **Morphisms**: An arrow $x \to y$ for every $g \in G$ such that $g \cdot x = y$.

Example 3.5.11 (Delooping). Consider a group G. Its delooping $\mathbf{B}G$ is defined as the one-object groupoid for which $\mathbf{B}G(*,*) = G$.

Property 3.5.12 (Representations). Consider a group G together with its delooping $\mathbf{B}G$. When considering *representations* as functors $\rho : \mathbf{B}G \to \mathbf{FinVect}$, one can see that the intertwiners (Definition 2.3.6) are exactly the natural transformations. More generally, all G-sets (Definition 2.3.1) can be obtained as functors $\mathbf{B}G \to \mathbf{Set}$.

Definition 3.5.13 (Core). Let $\mathbf{C}$ be a (small) category. The core $\mathsf{Core}(\mathbf{C}) \in \mathbf{Grpd}$ of $\mathbf{C}$ is defined as the maximal subgroupoid of $\mathbf{C}$.

Definition 3.5.14 (Orbit). Let $\mathcal{G}$ be a groupoid with O, M respectively the sets of objects and morphisms. On O one can define an equivalence $x \sim y \iff \exists \phi : x \to y$. The equivalence classes are called orbits and the set of orbits is denoted by O/M.

Definition 3.5.15 (Transitive component). Let $\mathcal{G}$ be a groupoid with O, M respectively the sets of objects and morphisms and let s, t denote the source and target maps on M. Given an orbit $o \in O/M$, the transitive component of M associated to o is defined as $s^{-1}(o)$, or equivalently, as $t^{-1}(o)$.

Property 3.5.16. Every groupoid is a (disjoint) union of its transitive components.

Definition 3.5.17 (Transitive groupoid). A groupoid $\mathcal{G}$ is said to be transitive if for all objects $x \neq y \in \mathsf{ob}(\mathcal{G})$, the set $\mathcal{G}(x, y)$ is not empty.

3.6 Lawvere theories ♣

Definition 3.6.1 (Lawvere theory). Let $\mathbf{F}$ denote the skeleton of $\mathbf{FinSet}$. A Lawvere theory consists of a small category $\mathbf{L}$ and a strict (finite) product-preserving *identity-on-objects* functor $\mathcal{L} : \mathbf{F}^{\mathrm{op}} \to \mathbf{L}$.

Equivalently, a Lawvere theory is a small category $\mathbf{L}$ with a **generic object** c_0 such that every object $c \in \mathsf{ob}(\mathbf{L})$ is a finite power of c_0.

Property 3.6.2. Lawvere theories $(\mathbf{L}, \mathcal{L})$ form a category $\mathbf{Law}$. Morphisms between Lawvere theories are (finite) product-preserving functors.

Definition 3.6.3 (Model). A model or **algebra** over a Lawvere theory $\mathbf{L}$ is a (finite) product-preserving functor $A : \mathbf{L} \to \mathbf{Set}$.

3.7 Operad theory ♣

3.7.1 Operads

Definition 3.7.1 (Plain operad[13]). Let $\mathcal{O} = \{P(n)\}_{n\in\mathbb{N}}$ be a collection of sets, called **n-ary operations** (where n is called the **arity**). The collection $\mathcal{O}$ is called a plain operad if it satisfies following axioms:

1. $P(1)$ contains an identity element $\mathbb{1}$.
2. For all positive integers $n, k_1, \dots, k_n$ there exists a composition map

$$\begin{aligned} \circ : P(n) \times P(k_1) \times \cdots \times P(k_n) &\to P(k_1 + \cdots + k_n) \\ : (\psi, \theta_1, \dots, \theta_n) &\mapsto \psi \circ (\theta_1, \dots, \theta_n) \end{aligned} \tag{3.47}$$

that satisfies two additional axioms:

- **identity**:

$$\theta \circ (\mathbb{1}, \dots, \mathbb{1}) = \mathbb{1} \circ \theta = \theta, \tag{3.48}$$

and

- **associativity**:

$$\begin{aligned} \psi \circ \Big(\theta_1 \circ (\theta_{1,1}, \dots, \theta_{1,k_1}), \dots, \theta_n \circ (\theta_{n,1}, \dots, \theta_{n,k_n})\Big) \\ = \Big(\psi \circ (\theta_1, \dots, \theta_n)\Big) \circ (\theta_{1,1}, \dots, \theta_{1,k_1}, \theta_{2,1}, \dots, \theta_{n,k_n}). \end{aligned} \tag{3.49}$$

If the operad is represented using planar tree diagrams, the associativity obtains a nice intuitive form. When combining planar tree diagrams in three layers, the associativity axiom says that one can either first glue the first two layers together or one can first glue the last two layers together.

Remark 3.7.2. Plain operads can be defined in any monoidal category. In the same way symmetric operad can be defined in any symmetric monoidal category.

Example 3.7.3 (Endomorphism operad). Consider a vector space V. For every $n \in \mathbb{N}$, one can define the endomorphism algebra $\mathsf{End}(V^{\otimes n}, V)$. The endomorphism operad $\mathcal{E}\mathrm{nd}(V)$ is defined as $\{\mathsf{End}(V^{\otimes n}, V)\}_{n\in\mathbb{N}}$.

Definition 3.7.4 (O-algebra). An object X is called an algebra over an operad O if there exist morphisms

$$O(n) \times X^n \to X$$

for every $n \in \mathbb{N}$ satisfying the usual composition and identity laws. Alternatively, this can be rephrased as the existence of a (plain) operad morphism $O(n) \to \mathcal{E}\mathrm{nd}(X)$.

Example 3.7.5 (Categorical O-algebra). An O-algebra in the category **Cat**.

[13] Also called a **nonsymmetric operad** or **non-Σ operad**.

3.7.2 Algebraic topology

Definition 3.7.6 (Stasheff operad). A topological operad $\mathcal{K}$ such that $\mathcal{K}(n)$ is given by the nth *Stasheff polytope/associahedron*. Composition is given by the inclusion of faces.

Definition 3.7.7 (A_∞-space). An algebra over the Stasheff operad. This induces the structure of a multiplication that is associative up to a coherent homotopy.

Definition 3.7.8 (Little k-cubes operad). A topological operad for which every topological space $\mathcal{P}(n)$ consists of all possible configurations of n embedded k-cubes in a (unit) k-cube. Composition is given by the obvious way of inserting one unit k-cube in one of the smaller embedded k-cubes.

Property 3.7.9 (Recognition principle). If a connected topological space X forms an algebra over the little k-cubes operad, it is (weakly) homotopy equivalent to the k-fold loop space $\Omega^k Y$ of another pointed topological space Y. For $k = 1$, one should technically use the Stasheff operad, but it can be shown that this is related to the little interval operad.

Chapter 4

Higher-dimensional Algebra ♣

The main reference for this chapter is a series of eponymous papers (Baez & Crans, 2003; Baez & Lauda, 2003). For Kapranov-Voevodsky 2-vector spaces, the reader is referred to the original paper by Kapranov and Voevodsky (1994). References for the section on Berezin calculus are Choquet-Bruhat and DeWitt-Morette (2000); Losev (2007). The section about higher Lie theory is mainly based on Fiorenza (2004). For fusion and modular categories, the main reference is Etingof, Gelaki, Nikshych, and Ostrik (2016).

Chapter Contents

4.1 Monoidal categories **118**
4.1.1 Braided categories 120
4.1.2 Monoidal functors 120
4.1.3 Closed categories 122
4.2 Enriched category theory **123**
4.2.1 Enriched constructions 125
4.2.2 Weighted (co)limits 128
4.3 Abelian categories **130**
4.3.1 Additive categories 130
4.3.2 Exact functors 132
4.3.3 Finiteness 133
4.4 Duality **135**
4.5 Tensor and fusion categories **138**
4.6 Ribbon and modular categories **140**
4.7 Module categories **141**
4.8 Higher vector spaces **142**
4.8.1 Kapranov–Voevodksy 2-vector spaces 142

4.8.2 Baez–Crans 2-vector spaces 142
4.9 Higher Lie theory **143**
4.9.1 Lie superalgebras 143
4.9.2 Lie n-algebras 145
4.10 Monoidal n-categories **148**
4.10.1 Relation with group cohomology 149

4.1 Monoidal categories

Definition 4.1.1 (Monoidal category). A category $\mathbf{C}$ equipped with a bifunctor

$$- \otimes - : \mathbf{C} \times \mathbf{C} \to \mathbf{C}, \tag{4.1}$$

called the **tensor product** or **monoidal product**, a distinct object $\mathbf{1}$, called the **(monoidal) unit**, and the following three natural isomorphisms, called the **coherence maps**:

- **Associator**: $\alpha_{x,y,z} : (x \otimes y) \otimes z \cong x \otimes (y \otimes z)$,
- **Left unitor**: $\lambda_x : \mathbf{1} \otimes x \cong x$, and
- **Right unitor**: $\rho_x : x \otimes \mathbf{1} \cong x$.

These natural transformations are required make the **triangle** and **pentagon** diagrams in Fig. 4.1 and Fig. 4.2, respectively, commute. A monoidal category for which the associator and the unitors are identity transformations is often said to be **strict**.

$$\begin{tikzcd}
(x \otimes \mathbf{1}) \otimes y \arrow[rr, "\alpha_{x,\mathbf{1},y}"] \arrow[dr, "\rho_x \otimes \mathbb{1}_y"'] & & x \otimes (\mathbf{1} \otimes y) \arrow[dl, "\mathbb{1}_x \otimes \lambda_y"] \\
& x \otimes y &
\end{tikzcd}$$

Figure 4.1: Triangle diagram.

Example 4.1.2 (Cartesian category). A monoidal category where the monoidal product is given by the ordinary product (Definition 3.4.59). If the monoidal product is not the ordinary product, but the monoidal unit is still terminal, the category is said to be **semicartesian**.

Definition 4.1.3 (Scalar). In a monoidal category, the scalars are defined as the endomorphisms $\mathbf{1} \to \mathbf{1}$. The set of scalars forms a commutative monoid.

$$\begin{tikzcd}
 & \big((w \otimes x) \otimes y\big) \otimes z \arrow[rr, "\alpha_{w,x,y} \otimes \mathbb{1}_z"] \arrow[dl, "\alpha_{w\otimes x,y,z}"'] & & \big(w \otimes (x \otimes y)\big) \otimes z \arrow[dr, "\alpha_{w,x\otimes y,z}"] & \\
(w \otimes x) \otimes (y \otimes z) \arrow[drr, "\alpha_{w,x,y\otimes z}"'] & & & & w \otimes \big((x \otimes y) \otimes z\big) \arrow[dll, "\mathbb{1}_w \otimes \alpha_{x,y,z}"] \\
 & & w \otimes \big(x \otimes (y \otimes z)\big) & &
\end{tikzcd}$$

Figure 4.2: Pentagon diagram.

Property 4.1.4. Every scalar $s : \mathbf{1} \to \mathbf{1}$ induces a natural transformation $s : \mathbb{1}_{\mathsf{C}} \Rightarrow \mathbb{1}_{\mathsf{C}}$ with components

$$s_x : x \cong \mathbf{1} \otimes x \xrightarrow{s \otimes \mathbb{1}_x} \mathbf{1} \otimes x \cong x\,. \tag{4.2}$$

For every morphism $f \in \mathsf{hom}(\mathbf{C})$, the naturality square $f \circ s_x = s_y \circ f$ alo defines a morphism $s \diamond f$ that is equivalently given by $\rho_y \circ (f \otimes s) \circ \rho_x^{-1}$ (one could have used the left unitors as well). These morphisms satisfy the following well-known rules of scalar multiplication from linear algebra:

- $s \diamond (s' \diamond f) = (s \circ s') \diamond f$,
- $(s \diamond f) \circ (s' \diamond g) = (s \circ s') \diamond (f \circ g)$, and
- $(s \diamond f) \otimes (s' \diamond g) = (s \circ s') \diamond (f \otimes g)$.

Definition 4.1.5 (Weak inverse). Let $(\mathbf{C}, \otimes, \mathbf{1})$ be a monoidal category. An object $y \in \mathsf{ob}(\mathbf{C})$ is called a weak inverse of an object $x \in \mathsf{ob}(\mathbf{C})$ if it satisfies $x \otimes y \cong \mathbf{1}$.

Remark 4.1.6. One can show that the existence of a one-sided weak inverse (as in the definition above) implies that it is in fact a two-sided weak inverse, i.e. $y \otimes x \cong \mathbf{1}$ also holds.

Theorem 4.1.7 (MacLane's coherence theorem). *Consider two functors $F, G : \mathbf{A} \to \mathbf{B}$ between two monoidal categories $\mathbf{A}, \mathbf{B}$. Any two natural transformations $\eta, \varepsilon : F \Rightarrow G$, constructed solely from the associator and the unitors, coincide.*

4.1.1 Braided categories

Definition 4.1.8 (Braided monoidal category). A monoidal category $(\mathbf{C}, \otimes, \mathbf{1})$ equipped with a natural isomorphism

$$\sigma_{x,y} : x \otimes y \cong y \otimes x \tag{4.3}$$

that makes the two **hexagon** diagrams 4.3a and 4.3b commute for all $x, y, z \in \mathrm{ob}(\mathbf{C})$. The isomorphism σ is called the **braiding** (morphism).

$$\begin{tikzcd}
 & (x \otimes y) \otimes z \arrow[dl, "\sigma_{x,y} \otimes \mathbb{1}_z"'] \arrow[dr, "\alpha_{x,y,z}"] & \\
(y \otimes x) \otimes z \arrow[d, "\alpha_{y,x,z}"'] & & x \otimes (y \otimes z) \arrow[d, "\sigma_{x,y\otimes z}"] \\
y \otimes (x \otimes z) \arrow[dr, "\mathbb{1}_y \otimes \sigma_{x,z}"'] & & (y \otimes z) \otimes x \arrow[dl, "\alpha_{y,z,x}"] \\
 & y \otimes (z \otimes x) &
\end{tikzcd}$$

(a) Hexagon diagram 1.

$$\begin{tikzcd}
 & x \otimes (y \otimes z) \arrow[dl, "\mathbb{1}_x \otimes \sigma_{y,z}"'] \arrow[dr, "\alpha^{-1}_{x,y,z}"] & \\
x \otimes (z \otimes y) \arrow[d, "\alpha^{-1}_{x,z,y}"'] & & (x \otimes y) \otimes z \arrow[d, "\sigma_{x\otimes y,z}"] \\
(x \otimes z) \otimes y \arrow[dr, "\sigma_{x,z} \otimes \mathbb{1}_y"'] & & z \otimes (x \otimes y) \arrow[dl, "\alpha^{-1}_{z,x,y}"] \\
 & (z \otimes x) \otimes y &
\end{tikzcd}$$

(b) Hexagon diagram 2.

Figure 4.3: Hexagon diagram.

Property 4.1.9 (Yang–Baxter equation). The components $\sigma_{x,x}$ of a braiding satisfy the *Yang–Baxter equation*. More generally, the braiding σ satisfies the following equation for all objects $x, y, z \in \mathrm{ob}(\mathbf{C})$:

$$(\sigma_{y,z} \otimes \mathbb{1}_x) \circ (\mathbb{1}_y \otimes \sigma_{x,z}) \circ (\sigma_{x,y} \otimes \mathbb{1}_z) = (\mathbb{1}_z \otimes \sigma_{x,y}) \circ (\sigma_{x,z} \otimes \mathbb{1}_y) \circ (\mathbb{1}_x \otimes \sigma_{y,z}) . \tag{4.4}$$

Remark 4.1.10. When drawing the above equality using string diagrams, it can be seen that the Yang–Baxter equation corresponds to the invariance of string diagrams under a *Reidemeister-III move*.

Definition 4.1.11 (Symmetric monoidal category). A braided monoidal category where the braiding σ satisfies

$$\sigma_{x,y} \circ \sigma_{y,x} = \mathbb{1}_{x\otimes y} . \tag{4.5}$$

4.1.2 Monoidal functors

Definition 4.1.12 (Monoidal functor). Let $(\mathbf{A}, \otimes, \mathbf{1}_\mathbf{A}), (\mathbf{B}, \circledast, \mathbf{1}_\mathbf{B})$ be two monoidal categories. A functor $F : \mathbf{A} \to \mathbf{B}$ is said to be monoidal if there exists:

1. a natural isomorphism $\psi_{x,y} : Fx \circledast Fy \Rightarrow F(x \otimes y)$ that makes the diagram in Fig. 4.4 commute, and

$$\begin{tikzcd}
(Fx \circledast Fy) \circledast Fz \arrow[r, "\alpha_{\mathbf{B}}"] \arrow[d, "\psi_{x,y} \circledast \mathbb{1}_{Fz}"'] & Fx \circledast (Fy \circledast Fz) \arrow[d, "\mathbb{1}_{Fx} \circledast \psi_{y,z}"] \\
F(x \otimes y) \circledast Fz \arrow[d, "\psi_{x \otimes y, z}"'] & Fx \circledast F(y \otimes z) \arrow[d, "\psi_{ax, y \otimes z}"] \\
F\big((x \otimes y) \otimes z\big) \arrow[r, "F\alpha_{\mathbf{A}}"'] & F\big(x \otimes (y \otimes z)\big)
\end{tikzcd}$$

Figure 4.4: Monoidal functor.

2. an isomorphism $\phi : \mathbf{1}_{\mathbf{B}} \to F\mathbf{1}_{\mathbf{A}}$ that makes the two diagrams in Fig. 4.5 commute.

The maps ψ and ϕ are also called **coherence maps** or **structure morphisms**.

$$\begin{tikzcd}
Fx \circledast \mathbf{1}_{\mathbf{B}} \arrow[r, "\mathbb{1}_{Fx} \circledast \phi"] \arrow[d, "\rho_{\mathbf{B}}"'] & Fx \circledast F\mathbf{1}_{\mathbf{A}} \arrow[d, "\psi_{x, \mathbf{1}_{\mathbf{A}}}"] \\
Fx & F(x \otimes \mathbf{1}_{\mathbf{A}}) \arrow[l, "F\rho_{\mathbf{A}}"]
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathbf{1}_{\mathbf{B}} \circledast Fy \arrow[r, "\phi \circledast \mathbb{1}_{Fy}"] \arrow[d, "\lambda_{\mathbf{B}}"'] & F\mathbf{1}_{\mathbf{A}} \circledast Fy \arrow[d, "\psi_{\mathbf{1}_{\mathbf{A}}, y}"] \\
Fy & F(\mathbf{1}_{\mathbf{A}} \otimes y) \arrow[l, "F\lambda_{\mathbf{A}}"]
\end{tikzcd}$$

Figure 4.5: Unitality diagrams.

Property 4.1.13 (Canonical unit). For every monoidal functor F, there exists a canonical isomorphism $\phi : \mathbf{1}_{\mathbf{B}} \to F\mathbf{1}_{\mathbf{A}}$ defined by the commutative diagram in Fig. 4.6.

$$\begin{tikzcd}
\mathbf{1}_{\mathbf{B}} \circledast F\mathbf{1}_{\mathbf{A}} \arrow[r, "\lambda_{\mathbf{B}}"] \arrow[d, "\phi \circledast \mathbb{1}_{F\mathbf{1}_{\mathbf{A}}}"'] & F\mathbf{1}_{\mathbf{A}} \arrow[d, "F\lambda_{\mathbf{A}}"] \\
F\mathbf{1}_{\mathbf{A}} \circledast F\mathbf{1}_{\mathbf{A}} \arrow[r, "\psi_{\mathbf{1}_{\mathbf{A}}, \mathbf{1}_{\mathbf{A}}}"'] & F(\mathbf{1}_{\mathbf{A}} \otimes \mathbf{1}_{\mathbf{A}})
\end{tikzcd}$$

Figure 4.6: Canonical unit isomorphism.

Definition 4.1.14 (Lax monoidal functor). A monoidal functor for which the coherence maps are merely morphisms and not isomorphisms.

Definition 4.1.15 (Monoidal natural transformation). A natural transformation η between (lax) monoidal functors (F, ψ, ϕ_F) and $(G, \widetilde{\psi}, \phi_G)$ that makes the diagrams in Fig. 4.7 commute.

Definition 4.1.16 (Monoidal equivalence). An equivalence of monoidal categories consisting of monoidal functors and monoidal natural isomorphisms.

$$\begin{tikzcd} & \mathbf{1_B} \arrow[dl, "\phi_F"'] \arrow[dr, "\phi_G"] & \\ F\mathbf{1_A} \arrow[rr, "\eta_{\mathbf{1_A}}"'] & & G\mathbf{1_A} \end{tikzcd} \qquad \begin{tikzcd} Fx \circledast Fy \arrow[r, "\psi_{a,b}"] \arrow[d, "\eta_a \circledast \eta_b"'] & F(x \otimes y) \arrow[d, "\eta_{a\otimes b}"] \\ Gx \circledast Gy \arrow[r, "\widetilde{\psi}_{a,b}"'] & G(x \otimes y) \end{tikzcd}$$

Figure 4.7: Monoidal natural transformation.

Theorem 4.1.17 (MacLane's strictness theorem). *Every monoidal category is monoidally equivalent to a strict monoidal category.*

4.1.3 Closed categories

Definition 4.1.18 (Internal hom). Let $(\mathbf{M}, \otimes, \mathbf{1})$ be a monoidal category. In this setting, one can generalize the *currying* procedure, i.e. the identification of maps $x \times y \to z$ with maps $x \to (y \to z)$. The internal hom-functor $\underline{\mathsf{Hom}}$ is defined by the following natural isomorphism:

$$\mathbf{M}(x \otimes y, z) \cong \mathbf{M}\big(x, \underline{\mathsf{Hom}}(y, z)\big). \tag{4.6}$$

The existence of all internal homs is equivalent to the existence of a right adjoint to the tensor functor.

Notation 4.1.19. The internal hom $\underline{\mathsf{Hom}}(x, y)$ is also often denoted by $[x, y]$ (or $x \multimap y$). From now on, this convention will be followed (unless otherwise specified).

Definition 4.1.20 (Closed monoidal category). A monoidal category that admits all internal homs. If the monoidal structure is induced by a (Cartesian) product structure, the category is often said to be **Cartesian closed**. A category for which all slice categories are Cartesian closed is said to be **locally Cartesian closed**.

Property 4.1.21. A locally Cartesian closed category with a terminal object is Cartesian closed. Moreover, it is finitely complete. More generally, a locally Cartesian closed category has all pullbacks.

Definition 4.1.22 (Exponential object). In the case of Cartesian (monoidal) categories, the internal hom $[x, y]$ is called the exponential object. This object is often denoted by y^x.

In Cartesian closed categories, a different, but frequently used, notation is $x \Rightarrow y$. However, this notation will not be used as it might be confused with the notation for *2-morphisms*.

Definition 4.1.23 (Cartesian closed functor). A functor between Cartesian closed categories that preserves products and exponential objects. As such, it is the natural notion of functor between Cartesian closed categories.

Property 4.1.24 (Frobenius reciprocity). A functor R between Cartesian closed categories that admits a left adjoint L is Cartesian closed if and only if the natural transformation

$$L(y \times Rx) \to Ly \times x \tag{4.7}$$

is a natural isomorphism.

Property 4.1.25 (Global elements). The following isomorphism is natural in both $x, y \in \mathrm{ob}(\mathbf{M})$:

$$\mathbf{M}\big(\mathbf{1}, [x, y]\big) \cong \mathbf{M}(x, y)\,. \tag{4.8}$$

It is this relation that gives the best explanation for the term 'internal hom'. One also immediately obtains the following natural isomorphism:

$$\mathbf{M}\big(x, [\mathbf{1}, y]\big) \cong \mathbf{M}(x, y)\,. \tag{4.9}$$

Because the Yoneda embedding is fully faithful, this implies that $[\mathbf{1}, y] \cong y$. Although the global elements $\mathbf{M}(\mathbf{1}, y)$ do not fully specify an object y, this does hold internally.

Property 4.1.26 (Symmetry). Let $\mathbf{M}$ be a closed monoidal category. The definition of an internal hom can also be internalized, i.e. there exists a natural isomorphism of the form

$$[x \otimes y, z] \cong [x, [y, z]]\,. \tag{4.10}$$

Furthermore, if $\mathbf{M}$ is also symmetric, there exists an internal isomorphism of the form

$$[x, [y, z]] \cong [y, [x, z]]\,. \tag{4.11}$$

Definition 4.1.27 (Strong adjunction). Consider a monoidal category $\mathbf{M}$ together with two endofunctors $L, R : \mathbf{M} \to \mathbf{M}$. These functors are said to form a strong adjunction if there exists a natural isomorphism

$$[Lx, y] \cong [x, Ry]\,. \tag{4.12}$$

Property 4.1.25 above implies that every strong adjunction is, in particular, an ordinary adjunction in the sense of Section 3.2.4.

4.2 Enriched category theory

The following definition is due to *Bénabou*. It should represent the 'ideal place in which to do category theory'.

Definition 4.2.1 (Cosmos). A complete and cocomplete, closed symmetric monoidal category.

Definition 4.2.2 (Enriched category). Let $(\mathcal{V}, \otimes, \mathbf{1})$ be a monoidal category. A $\mathcal{V}$-enriched category, also called a $\mathcal{V}$-category[1], consists of the following elements:

- a collection of objects $\mathsf{ob}(\mathbf{C})$, and
- for every pair of objects $x, y \in \mathsf{ob}(\mathbf{C})$, an object $\mathbf{C}(x, y) \in \mathsf{ob}(\mathcal{V})$ for which the following morphisms exist:
  1. $\mathsf{id}_x : \mathbf{1} \to \mathbf{C}(x, x)$ giving the (enriched) identity morphism, and
  2. $\circ_{xyz} : \mathbf{C}(y, z) \otimes \mathbf{C}(x, y) \to \mathbf{C}(x, z)$ replacing the usual composition.

The associativity and unity properties are given by commutative diagrams for the id and $\circ$ morphisms together with the associators and unitors in $\mathcal{V}$.

Definition 4.2.3 (Change of base). Consider a monoidal functor $F : \mathcal{V} \to \mathcal{W}$. This induces a change-of-base functor $F_* : \mathcal{V}\mathbf{Cat} \to \mathcal{W}\mathbf{Cat}$ by applying F to every hom-object.

Definition 4.2.4 (Underlying category). Given a $\mathcal{V}$-enriched category $\mathbf{C}$, the underlying category $\mathbf{C}_0$ is defined as follows:

- **Objects**: $\mathsf{ob}(\mathbf{C})$, and
- **Morphisms**: $\mathcal{V}\big(\mathbf{1}, \mathbf{C}(x, y)\big)$,

where $\mathbf{1}$ is the monoidal unit in $\mathcal{V}$. This construction can be obtained as the functor $\mathcal{V}\mathbf{Cat}(\mathcal{I}, -)$, where $\mathcal{I}$ is the one-object $\mathcal{V}$-category with $\mathcal{I}(*, *) := \mathbf{1}$.

Property 4.2.5 ($\mathcal{V}$ as a $\mathcal{V}$-category). Consider a closed monoidal category $\mathcal{V}$. This category can be given the structure $\widetilde{\mathcal{V}}$ of a $\mathcal{V}$-category by taking the hom-objects to be the internal homs, i.e. $\widetilde{\mathcal{V}}(x, y) := [x, y]$ for all $x, y \in \mathcal{V}$. Property 4.1.25 then implies that there exists an isomorphism between the underlying category $\widetilde{\mathcal{V}}_0$ and the original category $\mathcal{V}$.

Given two $\mathcal{V}$-enriched categories, one can define suitable functors between them.

Definition 4.2.6 (Enriched functor). A $\mathcal{V}$-enriched functor $F : \mathbf{A} \to \mathbf{B}$ consists of the following data:

- a function $F_0 : \mathsf{ob}(\mathbf{A}) \to \mathsf{ob}(\mathbf{B})$ (as for ordinary functors), and
- for every two objects $x, y \in \mathsf{ob}(\mathbf{A})$, a morphism $F_{x,y} : \mathbf{A}(x, y) \to \mathbf{B}(Fx, Fy)$ in $\mathcal{V}$.

These have to satisfy the 'usual' composition and unit conditions.

By extending Eq. (3.30) using enriched ends, one obtains a definition of enriched nat-

[1] Not to be confused with the notation for fibre categories (Definition 3.3.7).

$$\begin{tikzcd}
 & \mathbf{A}(x,y) \arrow[dl, "\lambda^{-1}"'] \arrow[dr, "\rho^{-1}"] & \\
\mathbf{1} \otimes \mathbf{A}(x,y) \arrow[d, "\eta_y \otimes F_{x,y}"'] & & \mathbf{A}(x,y) \otimes \mathbf{1} \arrow[d, "G_{x,y} \otimes \eta_x"] \\
\mathbf{B}(Fy,Gy) \otimes \mathbf{B}(Fx,Fy) \arrow[dr, "\circ"'] & & \mathbf{B}(Gx,Gy) \otimes \mathbf{B}(Fx,Gx) \arrow[dl, "\circ"] \\
 & \mathbf{B}(Fx,Gy) &
\end{tikzcd}$$

Figure 4.8: $\mathcal{V}$-naturality diagram.

ural transformations and, therefore, also a definition of enriched functor categories:

$$[\mathbf{A},\mathbf{B}](F,G) := \int_{x\in\mathbf{A}} \mathbf{B}(Fx,Gx)\,. \tag{4.13}$$

Given two $\mathcal{V}$-enriched functors $F,G : \mathbf{A} \to \mathbf{B}$, one can also try to define $\mathcal{V}$-natural transformations by extending the usual definition of natural transformations (Definition 3.2.15).

Definition 4.2.7 (Enriched natural transformation). An ordinary natural transformation consists of an $\mathrm{ob}(\mathbf{A})$-indexed family of morphism $\eta_x : Fx \to Gx$. This can also be interpreted as an $\mathrm{ob}(\mathbf{A})$-indexed family of morphisms $\eta_x : 1 \to \mathbf{B}(Fx,Gx)$ from the initial object (one-element set). By analogy, a $\mathcal{V}$-natural transformation is defined as an $\mathrm{ob}(\mathbf{A})$-indexed family of morphisms $\eta_x : \mathbf{1} \to \mathbf{B}(Fx,Gx)$ from the monoidal unit. The usual naturality square is replaced by the naturality hexagon in Fig. 4.8.

The question then becomes how these two definitions are related. The end in Eq. (4.13) comes equipped with a projection $\varepsilon_x : [\mathbf{A},\mathbf{B}](F,G) \to \mathbf{B}(Fx,Gx)$. Precomposing this morphism with a morphism in the underlying category $[\mathbf{A},\mathbf{B}]_0$ exactly gives a $\mathcal{V}$-natural transformation. So, the underlying category of $[\mathbf{A},\mathbf{B}]$ is the ordinary category of $\mathcal{V}$-functors and $\mathcal{V}$-natural transformations.

4.2.1 Enriched constructions

Definition 4.2.8 (Functor tensor product). Consider a covariant functor $G : \mathbf{C} \to \mathcal{V}$ and a contravariant functor $F : \mathbf{C}^{\mathrm{op}} \to \mathcal{V}$ into a monoidal category $\mathcal{V}$, where $\mathbf{C}$ does not have to be enriched over $\mathcal{V}$. The tensor product of F and G is defined as the following coend:

$$F \otimes_{\mathbf{C}} G := \int^{x\in\mathbf{C}} Fx \otimes Gx\,. \tag{4.14}$$

It should be noted that the above tensor product does not produce a new functor, instead it only gives an object in $\mathcal{V}$. A different type of tensor product, one that does give

a functor, exists in the enriched setting (note that there is no relation between these two definitions).

Definition 4.2.9 (Day convolution). Consider a monoidally cocomplete category $\mathcal{V}$, i.e. cocomplete monoidal category for which the tensor product bifunctor is cocontinuous in each argument, together with a $\mathcal{V}$-enriched category $\mathbf{C}$. The convolution or tensor product (if it exists) of two $\mathcal{V}$-enriched functors $F, G : \mathbf{C} \to \mathcal{V}$ is defined as the following coend:

$$F \otimes_{\mathsf{Day}} G := \iint^{x,y \in \mathbf{C}} \mathbf{C}(x \otimes y, -) \otimes Fx \otimes Gy \,. \tag{4.15}$$

Property 4.2.10 (Monoidal structure). When $\mathbf{M}$ is closed symmetric monoidal, Day convolution is associative and, hence, defines a monoidal structure on the functor category $[\mathbf{C}, \mathbf{M}]$. The tensor unit is given by the functor (co)represented by the tensor unit in $\mathbf{C}$.

Definition 4.2.11 (Copower). Consider a $\mathcal{V}$-enriched category $\mathbf{C}$. The copower (or **tensor**) functor $\cdot : \mathcal{V} \times \mathbf{C} \to \mathbf{C}$ is defined by the following natural isomorphism:

$$\mathbf{C}(v \cdot x, y) \cong \mathcal{V}\big(v, \mathbf{C}(x, y)\big) \,. \tag{4.16}$$

Dually, the **power** (or **cotensor**) functor $[-,-] : \mathcal{V} \times \mathbf{C} \to \mathbf{C}$ is defined by the following natural isomorphism:

$$\mathbf{C}\big(x, [v, y]\big) \cong \mathcal{V}\big(v, \mathbf{C}(x, y)\big) \,. \tag{4.17}$$

If an enriched category admits all (co)powers, it is said to be **(co)powered** (over its enriching category).

Remark 4.2.12. Property 4.1.26 says that every (closed) symmetric monoidal category $\mathbf{M}$ is powered over itself, the power just being the internal hom. The same holds for the copower, which is just the usual tensor product functor.

Example 4.2.13 (Disjoint unions). Every (co)complete (locally) small category $\mathbf{C}$ admits the structure of a $\mathbf{Set}$-(co)powered category:

$$\begin{aligned} x^S &:= \prod_{s \in S} x \,, \\ S \cdot x &:= \bigsqcup_{s \in S} x \,. \end{aligned} \tag{4.18}$$

The definition and properties of internal hom-functors and (co)powers can be formalized as follows.

Definition 4.2.14 (Two-variable adjunction). Consider three categories $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$. A two-variable adjunction $\mathbf{A} \times \mathbf{B} \to \mathbf{C}$ consists of three bifunctors:

- $- \otimes - : \mathbf{A} \times \mathbf{B} \to \mathbf{C}$,

- $\mathsf{Hom}_L : \mathbf{A}^{\mathrm{op}} \times \mathbf{C} \to \mathbf{B}$, and
- $\mathsf{Hom}_R : \mathbf{B}^{\mathrm{op}} \times \mathbf{C} \to \mathbf{A}$

admitting the following natural isomorphisms:

$$\mathbf{C}(x \otimes y, z) \cong \mathbf{A}\big(x, \mathsf{Hom}_R(y, z)\big) \cong \mathbf{B}\big(y, \mathsf{Hom}_L(x, z)\big)\,. \tag{4.19}$$

It should be noted that fixing any of the variables gives rise to ordinary adjunctions in the sense of Section 3.2.4.

Property 4.2.15 (Powers and copowers). A category $\mathbf{C}$ enriched over a monoidal category $\mathcal{V}$ is powered and copowered over $\mathcal{V}$ exactly if the hom-functor $\mathbf{C}^{\mathrm{op}} \times \mathbf{C} \to \mathcal{V}$ is the right adjoint in an enriched two-variable adjunction. The power and copower functors are then given by the other two adjoints.

The co-Yoneda lemma 3.4.80 can be generalized to the enriched setting.

Property 4.2.16 (Ninja Yoneda lemma). Let $F : \mathbf{C}^{\mathrm{op}} \to \mathcal{V}$ be a contravariant functor (similar statements hold for covariant functors).

$$\begin{aligned} \int^{x \in \mathbf{C}} \mathbf{C}(-, x) \otimes Fx &\cong F \\ \int_{x \in \mathbf{C}} \mathcal{V}\big(\mathbf{C}(x, -), Fx\big) &\cong F\,. \end{aligned} \tag{4.20}$$

The following definition constructs Kan extensions in the enriched setting (these can be shown to reduce to Definition 3.4.83 when enriching over $\mathbf{Set}$).

Alternative Definition 4.2.17 (Kan extension). Let $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$ be categories enriched over a monoidal category $\mathcal{V}$. If $\mathbf{B}$ is assumed to be copowered over $\mathcal{V}$, left Kan extension of $F : \mathbf{A} \to \mathbf{B}$ along $G : \mathbf{A} \to \mathbf{C}$ can be defined through a coend:

$$\mathsf{Lan}_G F := \int^{x \in \mathbf{A}} \mathbf{C}(Gx, -) \cdot Fx\,. \tag{4.21}$$

If $\mathbf{B}$ is assumed to be powered over $\mathcal{V}$, right Kan extension can be defined through an end:

$$\mathsf{Ran}_G F := \int_{x \in \mathbf{A}} \big[\mathbf{C}(-, Gx), Fx\big]\,. \tag{4.22}$$

If F is contravariant, the arguments of the hom-functors should simply be interchanged.

Remark 4.2.18. By choosing $\mathcal{V} = \mathbf{Set}, \mathbf{C} = \mathbf{A}$ and $G = \mathbb{1}_{\mathbf{A}}$ in the previous definition, one obtains the ninja Yoneda lemma 3.4.80.

Property 4.2.19. As already remarked in Definition 3.4.85, Kan extensions computed using (co)ends as above are strong.

Alternative Definition 4.2.20 (Functor tensor product). Let $\mathbf{B}$ be a $\mathcal{V}$-enriched category. Consider a covariant functor $G : \mathbf{A} \to \mathbf{B}$ and a contravariant functor $F : \mathbf{A}^{\mathrm{op}} \to \mathcal{V}$.

The tensor product (Definition 4.2.8) can be generalized whenever $\mathbf{B}$ is copowered over $\mathcal{V}$:

$$F \otimes_{\mathbf{A}} G := \int^{x \in \mathbf{A}} Fx \cdot Gx \,. \tag{4.23}$$

4.2.2 Weighted (co)limits

In this section, the definition of ordinary limits and, in particular, the defining universal property 3.4.39 is revisited. In this construction, the constant functor Δ_x was one of the main ingredients. This functor can be factorized as $\mathbf{I} \to 1 \to \mathbf{C}$, where 1 denotes the terminal category. At the level of morphisms, this factorization takes the form $\mathbf{I}(i,j) \to * \to \mathbf{C}(x,x)$, where $*$ denotes the terminal one-element set. However, whenever the enriching context is not $\mathbf{Set}$, one does not necessarily have access to a terminal object.

To avoid this issue, limits will first be redefined as representing objects. To this end, consider a general diagram $D : \mathbf{I} \to \mathbf{C}$. By postcomposition with the Yoneda embedding, one obtains the presheaf-valued diagram $\mathbf{C}(-, D-) : \mathbf{I} \to [\mathbf{C}^{\mathrm{op}}, \mathbf{Set}]$. Since presheaf categories are complete (Example 3.4.43), the limit of this diagram exists:

$$\mathbf{Set}\big(S, \lim \mathbf{C}(x, D-)\big) \cong [\mathbf{I}, \mathbf{Set}]\big(\Delta_S, \mathbf{C}(x, D-)\big) \,. \tag{4.24}$$

By restricting to the terminal set $S = *$, one obtains

$$\lim \mathbf{C}(x, D-) \cong [\mathbf{I}, \mathbf{Set}]\big(\Delta_*, \mathbf{C}(x, D-)\big) \,. \tag{4.25}$$

If this presheaf is representable, one can use the continuity of the hom-functor, together with the fact that the Yoneda embedding is fully faithful, to show that the representing object is (isomorphic to) $\lim D$, i.e.

$$[\mathbf{I}, \mathbf{Set}]\big(\Delta_*, \mathbf{C}(x, D-)\big) \cong \mathbf{C}(x, \lim D) \,. \tag{4.26}$$

@@ CLEAN THIS UP (note that continuity and pointwise definition was already mentioned for ordinary limits) @@

Definition 4.2.21 (Weighted limit). The above construction can now be generalized by replacing the constant functor Δ_* by any functor $W : \mathbf{I} \to \mathbf{Set}$. A representing object is then called the W-weighted limit of D. This object is often denoted by $\lim^W D$ or $\{W, D\}$. To distinguish weighted limits from ordinary ones, the latter are sometimes called **conical limits**.

Remark 4.2.22. A motivation for this construction is the following. As was already pointed out in Remark 3.4.18, the mere knowledge of global elements $1 \to x$ is often not enough to characterize an object x. In general, one should look at the collection of generalized elements. When applying this ideology to the case of cones, one sees that replacing the functor Δ_* by a more general functor is the same as replacing the global elements $* \to Di$ by generalized elements $Wi \to Di$.

The generalization to the enriched setting is now evident. There is no reference to the terminal object left, so one can replace **Set** by any enriching category. In the enriched setting, (co)end formulas for (weighted) limits will often be used.

Formula 4.2.23 (Enriched weighted limits). By expressing the natural transformations as an end as in Eq. (3.30) and by using the canonical powering in **Set**, one can express ordinary weighted limits as follows:

$$\lim^{W} D \cong \int_{i \in \mathbf{I}} [Wi, Di] \,. \tag{4.27}$$

The generalization to other enriching categories is now straightforward. Consider a diagram $D : \mathbf{I} \to \mathbf{C}$ and a weight functor $W : \mathbf{I} \to \mathcal{V}$, where $\mathbf{C}$ is $\mathcal{V}$-enriched. If $\mathbf{C}$ is powered over $\mathcal{V}$, the W-weighted limit of D is defined by the same formula as above:

$$\lim^{W} D := \int_{i \in \mathbf{I}} [Wi, Di] \,. \tag{4.28}$$

In a similar way, one can define weighted colimits in copowered $\mathcal{V}$-categories as coends:

$$\operatorname{colim}^{W} D := \int^{i \in \mathbf{I}} Wi \cdot Di \,. \tag{4.29}$$

Here, the weight functor W is required to be contravariant since colimits (and cocones in general) are natural transformations between contravariant functors.

Property 4.2.24 (Weighted limits are Homs). In the case $\mathbf{C} = \mathcal{V}$, the powering becomes the internal hom and, therefore, one sees that weighted limits are given by (enriched) natural transformations (as was the case for ordinary conical limits).

In the following example, the weighted colimit is calculated with respect to the Yoneda embedding.

Example 4.2.25 (Hom-functor). Consider a diagram $D : \mathbf{I} \to \mathbf{C}$. When using the Yoneda embedding $\mathcal{Y}i = \mathbf{I}(-, i)$ as the weight functor, one obtains the following property by virtue of the Yoneda lemma:

$$\operatorname{colim}^{\mathcal{Y}i} D \cong Di \,. \tag{4.30}$$

A similar statement for weighted limits can be obtained with the covariant Yoneda embedding.

Alternative Definition 4.2.26 (Weighted (co)limits). The above property can be used to axiomatize small weighted (co)limits in bicomplete categories:

1. **Yoneda**: For every object $i \in \operatorname{ob}(\mathbf{I})$, there exist isomorphisms

$$\lim^{\mathbf{I}(i,-)} D \cong Di \qquad \text{and} \qquad \operatorname{colim}^{\mathbf{I}(-,i)} D \cong Di \,. \tag{4.31}$$

2. **Cocontinuity**: The weighted (co)limit functors are cocontinuous in the weights.

One can also express Kan extensions as weighted limits (this simply follows from Definition 4.2.17).

Property 4.2.27 (Kan extensions). Consider functors $F : \mathbf{A} \to \mathbf{B}$ and $G : \mathbf{A} \to \mathbf{C}$. If, for every $x \in \mathsf{ob}(\mathbf{C})$, the weighted limit $\lim^{\mathbf{C}(x,G-)} F$ exists, these limits can be combined into a functor that can be shown to be the right Kan extension $\mathsf{Ran}_G F$. The left Kan extension can be obtained as a weighted colimit.

Property 4.2.28 (Category of elements). The weighted (co)limits of a functor (over **Set**) can also be expressed in terms of the category of elements (Definition 3.3.15) of the weight:

$$\lim^W F \cong \lim F \circ \mathbf{C}_W \,, \tag{4.32}$$

where the limit on the right-hand side is a conical limit.

The reason for the (co)end notation for categories of elements can also be explained by noting that these can actually be obtained as weighted (co)limits or (co)ends (here given for covariants functors):

$$\mathsf{El}(F) \cong \int^{c \in \mathbf{C}} Jc \times \mathsf{Disc}\, Fc \,, \tag{4.33}$$

where $J : \mathbf{C}^{\mathsf{op}} \to \mathbf{Cat} : c \mapsto c/\mathbf{C}$ assigns undercategories and $\mathsf{Disc} : \mathbf{Set} \to \mathbf{Cat}$ sends sets to discrete categories.

4.3 Abelian categories

4.3.1 Additive categories

Definition 4.3.1 (Pre-additive category). A (locally small) category enriched over **Ab**, i.e. a category in which every hom-set is an Abelian group and composition is bilinear.

Property 4.3.2. Let $\mathbf{A}$ be a pre-additive category. The following statements are equivalent for an object $x \in \mathsf{ob}(\mathbf{A})$:

- x is initial,
- x is final, or
- $\mathbb{1}_x = 0$.

It follows that every initial/terminal object in a pre-additive category is automatically a zero object (Definition 3.4.12).

Property 4.3.3 (Biproducts). In a pre-additive category, the following isomorphism holds for all finitely indexed sets $\{x_i\}_{i \leq n}$:

$$\prod_{i \leq n} x_i \cong \bigsqcup_{i \leq n} x_i \,. \tag{4.34}$$

Finite (co)products in pre-additive categories are often called **direct sums**. In general, if a product and coproduct exist and are isomorphic, one also speaks of a **biproduct**.

Definition 4.3.4 (Additive category). A pre-additive category admitting all finite products (and, hence, biproducts).

When working with additive categories, it is generally assumed that the associated functors are of a specific type.

Definition 4.3.5 (Additive functor). Let $\mathbf{A}, \mathbf{A}'$ be additive categories. A functor $F : \mathbf{A} \to \mathbf{A}'$ is said to be additive if it preserves finite biproducts:

1. It preserves zero objects: $F\,0_{\mathbf{A}} \cong 0_{\mathbf{A}'}$.
2. There exists a natural isomorphism $F(x \oplus y) \cong Fx \oplus Fy$.

This notion can be generalized to pre-additive categories. A functor between pre-additive categories is said to be additive if it acts by group morphisms on hom-spaces.

Definition 4.3.6 (Grothendieck group). Let $\mathbf{A}$ be an additive category and consider its decategorification (Definition 3.2.24). This set carries the structure of an Abelian monoid and, hence, the Grothendieck construction (Construction 2.2.6) can be applied to obtain an Abelian group $K(\mathbf{A})$. This group is called the Grothendieck group of $\mathbf{A}$.

In a (pre-)additive category, one can define some classical notions from (homological) algebra such as images and kernels.

Definition 4.3.7 (Kernel). Let $f : x \to y$ be a morphism. A[2] kernel of f is a morphism $k : z \to x$ such that:

$$f \circ k = 0 \tag{4.35}$$

and such that it is universal with respect to this property. This implies that a kernel of f could equivalently be defined as the equalizer of f and 0.

Notation 4.3.8 (Kernel). If the kernel of $f : x \to y$ exists, it is denoted by $\mathsf{ker}(f)$.

Definition 4.3.9 (Cokernel). Let $f : x \to y$ be a morphism. A cokernel of f is a morphism $p : y \to z$ such that:

$$p \circ f = 0 \tag{4.36}$$

and such that it is universal with respect to this property. This implies that a cokernel of f could equivalently be defined as the coequalizer of f and 0.

Notation 4.3.10 (Cokernel). If the cokernel of $f : x \to y$ exists, it is denoted by $\mathsf{coker}(f)$.

[2] 'A' since the kernel of a morphism is only determined up to an isomorphism.

Remark 4.3.11. The name and notation of the kernel and the cokernel (in the categorical sense) are explained by remarking that $\mathsf{ker}(f)$ represents the functor

$$F : z \mapsto \mathsf{ker}\big(\mathbf{C}(z,x) \to \mathbf{C}(z,y)\big)\,, \tag{4.37}$$

where ker denotes the algebraic kernel (Definition 2.1.10), and similarly for the cokernel.

Definition 4.3.12 (Pseudo-Abelian category). An additive category in which every projection/idempotent has a kernel.

Definition 4.3.13 (Pre-Abelian category). An additive category in which every morphism has a kernel and cokernel.

Definition 4.3.14 (Abelian category). A pre-Abelian category in which every mono is a kernel and every epi is a cokernel or, equivalently, if for every morphism f there exists an isomorphism

$$\mathsf{coker}\big(\mathsf{ker}(f)\big) \cong \mathsf{ker}\big(\mathsf{coker}(f)\big)\,. \tag{4.38}$$

Property 4.3.15. Every Abelian category is balanced (Definition 3.4.5).

Property 4.3.16 (Injectivity and surjectivity). In Abelian categories, a morphism is monic if and only if it is injective, i.e. its kernel is 0. Analogously, a morphism is epic if and only if it is surjective, i.e. its cokernel is 0.

Definition 4.3.17 (Linear category). Let $\mathbf{Vect}_k$ denote the category of vector spaces over the base field k. A k-linear category is a category enriched over $\mathbf{Vect}_k$. (If the base field is clear, it is often left implicit.)

4.3.2 Exact functors

In the setting of additive categories, Definition 3.4.45 can be characterized more simply.

Definition 4.3.18 (Exact functor). Let $F : \mathbf{A} \to \mathbf{A}'$ be an additive functor between additive categories.

- F is left exact if it preserves kernels.
- F is right exact if it preserves cokernels.
- F is exact if it is both left and right exact.

Corollary 4.3.19. The previous definition implies the following properties (which can, in fact, be used as an alternative definition):

- If F is left exact, it maps an exact sequence of the form

$$0 \longrightarrow x \longrightarrow y \longrightarrow z$$

to an exact sequence of the form

$$0 \longrightarrow Fx \longrightarrow Fy \longrightarrow Fz\,.$$

- If F is right exact, it maps an exact sequence of the form

$$x \longrightarrow y \longrightarrow z \longrightarrow 0$$

to an exact sequence of the form

$$Fx \longrightarrow Fy \longrightarrow Fz \longrightarrow 0\,.$$

- If F is exact, it maps short exact sequences to short exact sequences.

Notation 4.3.20 (Left or right). The category of left modules ${}_R\mathbf{Mod}$ over a ring R is equivalent (as an Abelian category) to the category of right modules $\mathbf{Mod}_{R^{\mathrm{op}}}$ over the opposite ring R. For this reason, one often makes no difference between left and right modules (only bimodules are truly relevant) and 'the category of R-modules' is just denoted by $R\mathbf{Mod}$.

Theorem 4.3.21 (Freyd–Mitchell embedding theorem). *Every small Abelian category admits a fully faithful, exact functor into a category of the form $R\mathbf{Mod}$ for some unital ring R.*

Theorem 4.3.22 (Eilenberg–Watts). *Let R, S be two (not necessarily unital) rings. The tensor product functor induces an equivalence between the category of R-S-bimodules and the category of cocontinuous functors $R\mathbf{Mod} \to S\mathbf{Mod}$.*

4.3.3 Finiteness

Definition 4.3.23 (Simple object). Let $\mathbf{A}$ be an Abelian category. An object $a \in \mathrm{ob}(\mathbf{A})$ is said to be simple if the only subobjects of a are 0 and a itself. An object is said to be semisimple if it is a direct sum of simple obejcts.

Definition 4.3.24 (Semisimple category). A category is said to be semisimple if every object is semisimple, where, in general, the direct sums are taken over finite index sets.

Definition 4.3.25 (Jordan–Hölder series). A filtration (Definition 1.6.9)

$$0 \longrightarrow x_1 \longrightarrow x_2 \longrightarrow \cdots \longrightarrow x_n = x \tag{4.39}$$

of an object x is said to be a Jordan–Hölder series if the quotient objects x_i/x_{i-1} are simple for all $i \leq n$. If the series has finite length, the object x is said to be **finite**.

Theorem 4.3.26 (Jordan–Hölder). *If an object in an Abelian category is finite, all of its Jordan–Hölder series have the same length. In particular, the multiplicities of simple objects are the same for all such series.*

Theorem 4.3.27 (Krull–Schmidt). *Any object in an Abelian category of finite length admits a unique decomposition as a direct sum of indecomposable objects*[3].

Definition 4.3.28 (Locally finite). A $\mathfrak{K}$-linear Abelian category is said to be locally finite if it satisfies the following conditions:

1. every hom-space is finite-dimensional, and
2. every object has finite length.

Definition 4.3.29 (Finite). A $\mathfrak{K}$-linear Abelian category is said to be finite if it satisfies the following conditions:

1. it is locally finite,
2. it has enough projectives or, equivalently, every simple object has a *projective cover*, and
3. the set of isomorphism classes of simple objects is finite.

Theorem 4.3.30 (Schur's lemma). *Let* $\mathbf{A}$ *be an Abelian category. For every two simple objects* $x, y \in \mathsf{ob}(\mathbf{A})$*, all nonzero morphisms* $x \to y$ *are isomorphisms. In particular, if* x, y *are two non-isomorphic simple objects, then* $\mathbf{A}(x, y) = 0$*. Furthermore,* $\mathbf{A}(x, x)$ *is a division ring for every simple object* $x \in \mathsf{ob}(\mathbf{A})$*.*

Corollary 4.3.31. If $\mathbf{A}$ is locally finite and $\mathfrak{K}$ is algebraically closed, then $\mathbf{A}(x, x) \cong \mathfrak{K}$ for all simple objects $x \in \mathsf{ob}(\mathbf{A})$. This follows from the fact that the only finite-dimensional division algebra over an algebraically closed field is the field itself.

The Freyd–Mitchell theorem 4.3.21 can be adapted to the finite linear case as follows.

Theorem 4.3.32 (Deligne). *Every finite* $\mathfrak{K}$*-linear Abelian category is* $\mathfrak{K}$*-linearly equivalent to a category of the form* $A\mathbf{Mod}^{\mathrm{fin}}$ *for* A *a finite-dimensional* $\mathfrak{K}$*-algebra.*

Construction 4.3.33 (Deligne tensor product). Let $\mathbf{A}, \mathbf{B}$ be two Abelian categories. Their Deligne (tensor) product is defined (if it exists) as the category $\mathbf{A} \boxtimes \mathbf{B}$ for which there exists a bijection between right exact functors $\mathbf{A} \boxtimes \mathbf{B} \to \mathbf{C}$ and right exact functors $\mathbf{A} \times \mathbf{B} \to \mathbf{C}$ (the latter being right exact in each argument).

For finite Abelian categories, it can be shown that their Deligne product always exists. By the Deligne embedding theorem, one can find an explicit description. Consider two finite-dimensional $\mathfrak{K}$-algebras A, B. The category $A\mathbf{Mod}^{\mathrm{fin}} \boxtimes B\mathbf{Mod}^{\mathrm{fin}}$ is equivalent to the category $(A \otimes_{\mathfrak{K}} B)\mathbf{Mod}^{\mathrm{fin}}$.

[3] An object is **indecomposable** if it cannot be written as a direct sum of its subobjects.

4.4 Duality

Definition 4.4.1 (Dual object). Let $(\mathbf{C}, \otimes, \mathbf{1})$ be a monoidal category and consider an object $x \in \mathsf{ob}(\mathbf{C})$. A left dual[4] of x is an object $x^* \in \mathsf{ob}(\mathbf{C})$ together with two morphisms $\eta : \mathbf{1} \to x \otimes x^*$ and $\varepsilon : x^* \otimes x \to \mathbf{1}$, called the **unit** and **counit** morphisms[5], such that the diagrams in Fig. 4.9 commute. x is said to be **dualizable** if it admits a left dual.

$$\begin{tikzcd}
 & x & \\
\mathbf{1} \otimes x \arrow[ur, "\lambda_x"] \arrow[d, "\eta \otimes \mathbb{1}_x"'] & & x \otimes \mathbf{1} \arrow[ul, "\rho_x"'] \\
(x \otimes x^*) \otimes x \arrow[rr, "\alpha_{x,x^*,x}"] & & x \otimes (x^* \otimes x) \arrow[u, "\mathbb{1}_x \otimes \varepsilon"']
\end{tikzcd}$$

(a) Dual object I.

$$\begin{tikzcd}
 & x^* & \\
x^* \otimes \mathbf{1} \arrow[ur, "\rho_{x^*}"] \arrow[d, "\mathbb{1}_{x^*} \otimes \eta"'] & & \mathbf{1} \otimes x^* \arrow[ul, "\lambda_{x^*}"'] \\
x^* \otimes (x \otimes x^*) \arrow[rr, "\alpha^{-1}_{x^*,x,x^*}"] & & (x^* \otimes x) \otimes x^* \arrow[u, "\varepsilon \otimes \mathbb{1}_{x^*}"']
\end{tikzcd}$$

(b) Dual object II.

Figure 4.9: Dualizable objects.

Definition 4.4.2 (Rigid category). A monoidal category that has all duals. These categories are also said to be **autonomous**. If only left (resp. right) duals exist, the category is said to be left (resp. right) rigid.

Property 4.4.3 (Braided categories). In general, it is not true that left and right duals coincide. However, in a braided monoidal category this is the case.

Definition 4.4.4 (Compact closed category). A symmetric rigid category.

Example 4.4.5 (FinVect). Consider the category **FinVect** of finite-dimensional vector spaces (the ground field is assumed to be $\mathbb{R}$). The categorical dual of a vector space V is the algebraic dual V^*. The unit morphism is given by the 'resolution of the identity':

$$\eta : \mathbf{1} \to V \otimes V^* : 1 \mapsto \sum_{i=1}^{\dim(V)} e_i \otimes \phi^i \,, \tag{4.40}$$

where $\{e_i\}$ and $\{\phi^i\}$ are dual bases of V and V^*, respectively.

It should be noted that the category **Vect** of all vector spaces is not rigid. By Property 4.4.3 above, left and right duals coincide in any braided monoidal category (such as **Vect**), but for infinite-dimensional vector spaces it is known that $A \cong (A^*)^*$ never holds and, as such, rigidity cannot be extended to **Vect**.

Property 4.4.6 (Tannaka duality). Consider the category $\mathcal{V} = \mathbf{FinVect}_k$. Using coends, one can reconstruct the base field from its modules, i.e. the objects in $\mathcal{V}$:

$$\int^{V \in \mathcal{V}} V^* \otimes V \cong k \,. \tag{4.41}$$

[4] x is called the **right dual** of x^*. The right dual of y is often denoted by ${}^*y$.
[5] Also called the **coevaluation** and **evaluation** morphisms.

This result can be shown to hold for all compact closed categories $\mathcal{V}$. In this context, it is known as **Tannaka reconstruction**. A more general statement goes as follows:

$$\int^{V\in\mathcal{V}} \mathcal{V}(V,-)\otimes V \cong \mathrm{id}_{\mathcal{V}}\,. \tag{4.42}$$

For $\mathcal{V} = \mathbf{FinVect}_k$, the components $\eta_V : \mathcal{V}(V,V) \to k$ of the coend can be shown to coincide with the trace and, as such, the trace obtains a universal property.

Remark 4.4.7. This property can also be generalized by replacing $\mathcal{V}$ by a category of modules $A\mathbf{Mod}$ for some finite-dimensional algebra A. The end and coend give the algebra A and its dual A^*, respectively.

For certain purposes, it is useful to slightly weaken the notion of compact closed categories (e.g. Barr (1991); Kissinger and Uijlen (2019)).

Definition 4.4.8 ($*$-autonomous category). A symmetric monoidal category $(\mathbf{C},\otimes,\mathbf{1})$ with a fully faithful dualization functor $* : \mathbf{C}^{\mathrm{op}} \to \mathbf{C}$ such that

$$\mathbf{C}(x\otimes y, z^*) \cong \mathbf{C}\big(x,(y\otimes z)^*\big) \tag{4.43}$$

for all $x,y,z \in \mathrm{ob}(\mathbf{C})$.

Property 4.4.9 (de Morgan duality). If $(\mathbf{C},\otimes,\mathbf{1},*)$ is a $*$-autonomous category, it admits a second $*$-autonomous structure $(\mathbf{C},\parr,\bot,*)$ given by

$$x \parr y := (x^*\otimes y^*)^* \tag{4.44}$$

for all $x,y \in \mathrm{ob}(\mathbf{C})$ and

$$\bot := \mathbf{1}^*\,. \tag{4.45}$$

Property 4.4.10 (Involution). Note that the definition of $*$-autonomous categories implies that

$$x \cong x^{**} \tag{4.46}$$

for all $x \in \mathrm{ob}(\mathbf{C})$. Moreover, $*$-autonomous categories are closed with the internal hom given by:

$$[x,y] = x^* \parr y \cong (x\otimes y^*)^*\,. \tag{4.47}$$

A $*$-autonomous category is compact closed if and only if duality is compatible with tensor products:

$$(x\otimes y)^* \cong x^*\otimes y^*\,. \tag{4.48}$$

Equivalently, a $*$-autonomous category is compact closed if and only if the de Morgan-dual monoidal structures coincide.

Definition 4.4.11 (Symmetric monoidal dagger category). A symmetric monoidal category $(\mathbf{C}, \otimes, \mathbf{1})$ that also carries the structure of a dagger category (Definition 3.3.1) such that

$$(f \otimes g)^\dagger = f^\dagger \otimes g^\dagger \tag{4.49}$$

and such that the coherence and braiding morphisms are unitary.

Definition 4.4.12 (Dagger-compact category). A symmetric monoidal dagger category that is also a compact closed category such that the following diagram commutes for all objects:

$$\begin{tikzcd} & \mathbf{1} \arrow[dl, "\eta"'] \arrow[dr, "\varepsilon^\dagger"] & \\ x^* \otimes x & & x \otimes x^* \arrow[ll, "\sigma_{x,x^*}"] \end{tikzcd}$$

The trace on **FinVect** can be generalized as follows.

Definition 4.4.13 (Trace). Let $(\mathbf{C}, \otimes, \mathbf{1})$ be a rigid category and consider $f \in \mathbf{C}(x, x^{**})$. The left (**categorical** or **quantum**) trace of f is defined as the following morphism in $\mathsf{End}_{\mathsf{C}}(\mathbf{1})$:

$$\mathsf{tr}^L(f) := \varepsilon_{x^*} \circ (f \otimes \mathbb{1}_{x^*}) \circ \eta_x \,. \tag{4.50}$$

For $f \in \mathbf{C}(x, {}^{**}x)$, the right trace is defined similarly:

$$\mathsf{tr}^R(f) := \varepsilon_{{}^{**}x} \circ (\mathbb{1}_x \otimes f) \circ \eta_{{}^*x} \,. \tag{4.51}$$

Property 4.4.14. The following linear algebra-like properties hold for the categorical trace:

- $\mathsf{tr}^L(f) = \mathsf{tr}^R(f^*)$,
- $\mathsf{tr}^L(f \otimes g) = \mathsf{tr}^L(f)\,\mathsf{tr}^L(g)$, and
- for additive categories: $\mathsf{tr}^L(f \oplus g) = \mathsf{tr}^L(f) + \mathsf{tr}^L(g)$.

The second and third property can be stated analogously for the right trace.

Definition 4.4.15 (Pivotal category). Let $\mathbf{C}$ be a rigid monoidal category. A pivotal structure on $\mathbf{C}$ is a monoidal natural isomorphism $\psi : \mathsf{id}_{\mathsf{C}} \Rightarrow {}^{**}$.

Definition 4.4.16 (Dimension). Let $(\mathbf{C}, \psi)$ be a pivotal category and consider an object $x \in \mathsf{ob}(\mathbf{C})$. The dimension of x is defined as follows:

$$\mathsf{dim}_\psi(x) := \mathsf{tr}^L(\psi_x) \,. \tag{4.52}$$

Definition 4.4.17 (Spherical category). A pivotal category $(\mathbf{C}, \psi)$ in which the left and right dimensions with respect to ψ coincide: $\mathsf{dim}_\psi(x) = \mathsf{dim}_\psi(x^*)$ for all $x \in \mathsf{ob}(\mathbf{C})$.

Definition 4.4.18 (Calabi–Yau category). A $\mathbf{Vect}_k$-enriched category $\mathbf{C}$ equipped with a **trace** functional

$$\mathsf{tr}_x : C(x, x) \to k \tag{4.53}$$

for each object $x \in \mathsf{ob}(\mathbf{C})$ such that the induced pairing

$$\langle \cdot, \cdot \rangle : C(x, y) \otimes C(y, x) \to k : f \otimes g \mapsto \mathsf{tr}_x(g \circ f) \tag{4.54}$$

is symmetric and nondegenerate.

Example 4.4.19 (Frobenius algebra). A one-object Calabi–Yau category is the (pointed) monoid delooping of a *Frobenius algebra* (see Definition 28.1.7).

4.5 Tensor and fusion categories

Some definitions might slightly differ from the ones in the main references and some properties might be stated less generally. K denotes an algebraically closed field (this will often be $\mathbb{C}$).

Definition 4.5.1 (Tensor category). A monoidal category with the following properties:

1. it is rigid,
2. it is Abelian,
3. it is K-linear in a way compatible with the Abelian structure,
4. $\mathsf{End}(\mathbf{1}) \cong K$, and
5. $- \otimes -$ is bilinear on morphisms.

Some authors (such as Etingof et al. (2016)) also add 'locally finite' as a condition (Definition 4.3.28).

Remark 4.5.2. If K is not algebraically closed, one should replace the fourth condition by the condition that $\mathbf{1}$ is a simple object. However, if K is algebraically closed, these statements are equivalent.

Definition 4.5.3 (Tannakian category). A tensor category $\mathbf{C}$ over a field K such that there exists a field extension L/K (Definition 2.5.11) and a K-linear, exact and faithful monoidal functor $F : \mathbf{C} \to \mathbf{FinVect}_L$. If such a structure exists for $L = K$, the category is said to be **neutral**.

Definition 4.5.4 (Pointed tensor category). A tensor category where all simple objects are (weakly) invertible.

Definition 4.5.5 (Fusion category). A semisimple, finite tensor category.

Property 4.5.6. Let $\mathbf{C}$ be a fusion category. There exists a natural isomorphism $\mathbb{1}_{\mathbf{C}} \cong **$.

Remark 4.5.7. Although any fusion category admits a natural isomorphism between an object and its double dual, this morphism does not need to be monoidal. The fact that all fusion categories are pivotal was conjectured by *Etingof, Ostrik* and *Nikshych*. Currently, the best one can do for a general fusion category is a monoidal natural transformation between the identity functor and the fourth dualization functor $\mathbb{1}_{\mathbf{C}} \cong ****$.

Definition 4.5.8 (Categorical dimension). Consider a fusion category $\mathbf{C}$ and choose a natural isomorphism $\psi : \mathbb{1}_{\mathbf{C}} \cong **$. For every simple object $x \in \mathsf{ob}(\mathbf{C})$, one can define a dimension function, sometimes called the **norm squared**, in the following way:

$$|x|^2 := \mathsf{tr}(\psi_x)\,\mathsf{tr}\big((\psi_x^{-1})^*\big)\,. \tag{4.55}$$

If $\mathbf{C}$ is pivotal, this becomes $|x|^2 = \mathsf{dim}_\psi(x)\,\mathsf{dim}_\psi(x^*)$. In particular, when $\mathbf{C}$ is spherical, this becomes $|x|^2 = \mathsf{dim}_\psi(x)^2$. The categorical dimension, sometimes called the **Müger dimension**, is defined as follows:

$$\mathsf{dim}(\mathbf{C}) := \sum_{x\in\mathcal{O}(\mathbf{C})} |x|^2\,, \tag{4.56}$$

where $\mathcal{O}(\mathbf{C})$ denotes the set of isomorphism classes of simple objects.

Remark 4.5.9. It should be noted that the above quantities do not depend on the choice of isomorphism $\psi_x : x \cong x^{**}$ since all of them only differ by a scale factor.

Property 4.5.10 (Nonzero dimension). For any fusion category $\mathbf{C}$, one has that $\mathsf{dim}(\mathbf{C}) \neq 0$. In particular, if $K = \mathbb{C}$, then $\mathsf{dim}(\mathbf{C}) \geq 1$ (since the norm squared of any simple object is then also positive).

Definition 4.5.11 (G-graded fusion category). A semisimple linear category $\mathbf{C}$ is said to have a G-**grading**, where G is a finite group, if it can be decomposed as follows:

$$\mathbf{C} \cong \bigoplus_{g\in G} \mathbf{C}_g\,, \tag{4.57}$$

where every $\mathbf{C}_g$ is linear and semisimple. A fusion category $\mathbf{C}$ is said to be a (G-)graded fusion category if it admits a G-grading such that $\mathbf{C}_g \otimes \mathbf{C}_h \subseteq \mathbf{C}_{gh}$ for all $g, h \in G$.

Example 4.5.12 (G-graded vector spaces). Define the category $\mathbf{Vect}_G^\omega$ as having the same objects and morphisms as $\mathbf{Vect}_G$, the category of G-graded vector spaces, but with the associator given by the the 3-cocycle $\omega \in H^3(G; K^\times)$.

Property 4.5.13. Any pointed fusion category is equivalent to a category of the form $\mathbf{Vect}_G^\omega$ for some G and $\omega \in H^3(G; K^\times)$.

Theorem 4.5.14 (Tannaka duality). *The category of modules of a weak Hopf algebra*[6] *has the structure of a fusion category. Conversely, any fusion category can be obtained as the category of modules of a weak Hopf algebra.*

4.6 Ribbon and modular categories

Definition 4.6.1 (Ribbon structure). Consider a braided monoidal category $(\mathbf{C}, \otimes, \mathbf{1})$ with braiding σ. A **twist** or **balancing** is a natural endomorphism θ such that the following equation is satisfied for all $x, y \in \mathsf{ob}(\mathbf{C})$:

$$\theta_{x\otimes y} = (\theta_x \otimes \theta_y) \circ \sigma_{y,x} \circ \sigma_{x,y} \,. \tag{4.58}$$

If, in addition, $\mathbf{C}$ is rigid and the twist satisfies $\theta_{x^*} = (\theta_x)^*$ for all $x \in \mathsf{ob}(\mathbf{C})$, one speaks of a ribbon or **tortile** category.

Definition 4.6.2 (Drinfel'd morphism). Let $(\mathbf{C}, \otimes, \mathbf{1})$ be a rigid braided monoidal category with braiding σ. This structure admits a canonical natural automorphism $\mathsf{id}_{\mathbf{C}} \cong **$ defined as follows:

$$x \xrightarrow{\mathbb{1}_x \otimes \eta_{x^*}} x \otimes x^* \otimes x^{**} \xrightarrow{\sigma_{x,x^*} \otimes \mathbb{1}_{x^{**}}} x^* \otimes x \otimes x^{**} \xrightarrow{\varepsilon_x \otimes \mathbb{1}_{x^{**}}} x^{**} \,. \tag{4.59}$$

Property 4.6.3. Let $\mathbf{C}$ be a braided monoidal category. Consider the Drinfel'd morphism $u : \mathsf{id}_{\mathbf{C}} \cong **$ defined above. Any natural isomorphism $\psi : \mathsf{id}_{\mathbf{C}} \cong **$ can be written as $u \circ \theta$ where $\theta \in \mathsf{Aut}(\mathbb{1}_{\mathbf{C}})$. It is not hard to see that this natural isomorphism is monoidal (and, hence, pivotal) exactly when θ is a twist. If $\mathbf{C}$ is a fusion category, the pivotal structure is spherical if and only if θ determines a ribbon structure.

Definition 4.6.4 (Premodular category). A ribbon fusion category. Equivalently, a spherical braided fusion category.

Definition 4.6.5 (S-matrix). Given a premodular category $\mathbf{M}$ with braiding σ, the S-matrix is defined as follows:

$$S_{xy} := \mathsf{tr}(\sigma_{y,x} \circ \sigma_{x,y}) \,, \tag{4.60}$$

where $x, y \in \mathcal{O}(\mathbf{M})$ are (isomorphism classes of) simple objects. Since in a premodular category there are only finitely many isomorphism classes of simple objects (denote this number by $\mathcal{I}$), one can see that S is a $\mathcal{I} \times \mathcal{I}$-matrix.

Definition 4.6.6 (Modular category[7]). A premodular category for which the S-matrix is invertible.

[6] A weak version of Definition 28.2.3.

[7] 'Modular tensor category' is often abbreviated as **MTC**.

Property 4.6.7. Let $\mathbf{M}$ be a modular category with S-matrix S and define the following matrix:

$$E_{xy} := \begin{cases} 1 & \text{if } x = y^* \,, \\ 0 & \text{otherwise} \,. \end{cases} \tag{4.61}$$

The following relation with the categorical dimension of $\mathbf{M}$ is obtained:

$$S^2 = \mathsf{dim}(\mathbf{M})E \,. \tag{4.62}$$

Formula 4.6.8 (Verlinde). Consider a modular category $\mathbf{M}$ with S-matrix S. Let $\mathcal{O}(\mathbf{M})$ denote the set of isomorphism classes of simple objects and let dim denote the (pivotal) dimension associated to the spherical structure on $\mathbf{M}$. Using the formula

$$S_{xy}S_{xz} = \mathsf{dim}(x) \sum_{w \in \mathcal{O}(\mathbf{M})} N_{yz}^{w} S_{xw} \tag{4.63}$$

for all $x, y, z \in \mathcal{O}(\mathbf{M})$, one obtains the following important relation:

$$\sum_{w \in \mathcal{O}(\mathbf{M})} \frac{S_{wy}S_{wz}S_{wx^*}}{\mathsf{dim}(w)} = \mathsf{dim}(\mathbf{M})N_{yz}^{x} \,. \tag{4.64}$$

This property implies that the S-matrix of a modular category determines the fusion coefficients of the underlying fusion category.

4.7 Module categories

By categorifying the definition of a module over a ring (Definition 2.4.53), one obtains the notion of a module category.

Definition 4.7.1 (Module category). Let $\mathbf{M}$ be a monoidal category. A left $\mathbf{M}$-module (category) is a category $\mathbf{C}$ equipped with a bilinear functor $\triangleright : \mathbf{M} \times \mathbf{C} \to \mathbf{C}$ together with natural isomorphisms that categorify the associativity and unit conditions of modules (these are also required to be compatible with the associator and unitors of $\mathbf{M}$).

Remark 4.7.2. Similar to how a G-set can be defined as a functor $\mathbf{B}G \to \mathbf{Set}$ (Property 3.5.12), one can define a module category as a 2-functor $\mathbf{BM} \to \mathbf{Cat}$.

Analogous to Definition 4.1.18, one can also define internal homs for module categories.

Definition 4.7.3 (Internal hom). Consider a left $\mathbf{M}$-module $\mathbf{C}$. Given two objects $x, y \in \mathsf{ob}(\mathbf{C})$, one defines their internal hom (if it exists) as the object $\underline{\mathsf{Hom}}(x, y) \in \mathsf{ob}(\mathbf{M})$ satisfying the following condition

$$\mathbf{C}(m \triangleright x, y) \cong \mathbf{M}\big(m, \underline{\mathsf{Hom}}(x, y)\big) \tag{4.65}$$

for all $m \in \mathsf{ob}(\mathbf{M})$.

Property 4.7.4. It should be noted that for the case $\mathbf{C} \equiv \mathbf{M}$, where the action is given by the tensor product in $\mathbf{M}$, one obtains Definition 4.1.18 as a particular case.

4.8 Higher vector spaces

4.8.1 Kapranov–Voevodksy 2-vector spaces

The guiding principle for the definition of 2-vector spaces in this section will be the generalization of certain observations from studying the category **Vect** of ordinary vector spaces. Linear maps between vector spaces can (at least in finite dimensions) be represented as matrices with coefficients in the base field K. Coincidentally, this base field is also the tensor unit in **Vect**. At the same time, all finite-dimensional vector spaces are isomorphic to spaces of the form K^n, where n is given by the dimension of the vector space.

Definition 4.8.1 (2-vector space). To define 2-vector spaces, *Kapranov* and *Voevodsky* lifted these observations to categories by replacing the base field K by the category $\mathbf{Vect}_K$. To wit, $\mathbf{2Vect}_K$ is defined as the 2-category consisting of the following data:

- **Objects**: Finite products of the form $\mathbf{Vect}_K^n$.
- **1-morphisms**: Collections $\|A_{ij}\|$ of finite-dimensional K-vector spaces, called **2-matrices**.
- **2-morphisms**: Collections (f_{ij}) of linear maps between finite-dimensional K-vector spaces.

The multiplication (or composition) of 1-morphisms is defined in analogy to the multiplication of ordinary matrices, but where the usual sum and product are replaced by the direct sum and tensor product.

A seemingly more formal definition uses the concepts of *ring* and module categories.

Alternative Definition 4.8.2. A 2-vector space is a lax module category over **Vect** that is module-equivalent to $\mathbf{Vect}^n$ for some $n \in \mathbb{N}$. The 2-category **2Vect** is then defined as the 2-category with objects these 2-vector spaces, as 1-morphisms the associated **Vect**-module functors and as 2-morphisms the module natural transformations.

4.8.2 Baez–Crans 2-vector spaces

Definition 4.8.3 (2-vector space). A category internal to **Vect**. The morphism are **linear functors**, i.e. functors internal to **Vect**.

Remark 4.8.4. The above definition should not be confused with that of categories and functors enriched over **Vect**.

Example 4.8.5 (Base field). The base field K can be categorified to a 2-vector spaces by taking $K_0 = K_1 := K$ and $s = t = e := \mathbb{1}_K$. This object serves as a unit for the tensor product on $\mathbf{2Vect}_K$.

Property 4.8.6 (Chain complexes). There exists an equivalence between the (2-)categories of 2-vector spaces and 2-term chain complexes.

Proof (Sketch of construction). Given a 2-vector space (V_0, V_1), one can build a chain complex $C_\bullet$ as follows:

- $C_0 := V_0$,
- $C_1 := \ker(s)$, and
- $d := t|_{C_1}$,

where s, t are the source and target morphisms. □

Remark 4.8.7. The equivalence (ar the level of ordinary categories) is an instance of the *Dold–Kan correspondence* (see Theorem 13.1.19).

Definition 4.8.8 (Arrow part). Consider a 2-vector space $V = (V_0, V_1)$. For any morphism $f \in V_1$, one defines the arrow part as follows:

$$\vec{f} := f - e\big(s(f)\big)\,, \tag{4.66}$$

where e, s are the identity and source morphisms in V. Any map can thus be recovered from its arrow part and its source. This allows to identify a map $f \in V_1$ with the pair $\big(s(f), \vec{f}\,\big)$. Using arrow parts, one can rewrite the composition law of morphisms in an intuitive way:

$$g \circ f = \big(s(f), \vec{f} + \vec{g}\,\big)\,. \tag{4.67}$$

Definition 4.8.9 (Antisymmetric morphism). A natural morphism between n-linear functors in **2Vect** is said to be **completely antisymmetric** if its arrow part is completely antisymmetric.

4.9 Higher Lie theory

4.9.1 Lie superalgebras

Definition 4.9.1 (Internal Lie algebra). Let $(\mathbf{C}, \otimes, \mathbf{1})$ be a linear symmetric monoidal category with braiding σ. A Lie algebra internal to $\mathbf{C}$ is an object $L \in \mathrm{ob}(\mathbf{C})$ and a morphism

$$[\cdot,\cdot] : L \otimes L \to L$$

satisyfing the following conditions:

1. **Antisymmetry**: $[\cdot,\cdot] + [\cdot,\cdot] \circ \sigma_{L,L} = 0$, and
2. **Jacobi identity**: $[\cdot,[\cdot,\cdot]] + [\cdot,[\cdot,\cdot]] \circ \tau + [\cdot,[\cdot,\cdot]] \circ \tau^2 = 0\,,$

where $\tau = (\mathbb{1}_L \otimes \sigma_{L,L}) \circ (\sigma_{L,L} \otimes \mathbb{1}_L)$ denotes cyclic permutation.

Example 4.9.2 (Lie superalgebra). When using the braiding

$$\sigma(x \otimes y) = (-1)^{\deg(x)\deg(y)} y \otimes x \tag{4.68}$$

in **sVect**, a Lie superalgebra (also called a super Lie algebra) is obtained. More generally, in $\mathbb{Z}$-**Vect**, a Lie bracket of degree k is induced by the braiding

$$\sigma(x \otimes y) = (-1)^{(\deg(x)-k)(\deg(y)-k)} y \otimes x\,. \tag{4.69}$$

It is simply a Lie bracket on the k-suspension $\Pi^k V$.

Example 4.9.3 (dg-Lie algebras). Lie algebras internal to $\mathbf{Ch}_\bullet(\mathbf{Vect})$ or its generalization to graded vector spaces. Sometimes these are also called strict L_∞-algebras (see further below).

The following notion is a slight modification of the idea of a (graded) *Poisson algebra* (see Definition 32.2.5).

Definition 4.9.4 (Gerstenhaber algebra). A graded-commutative algebra equipped with a degree-1 Lie bracket that acts as a graded derivation:

$$[x, yz] = [x, y]z + (-1)^{\deg(x)(\deg(y)-1)} y[x, z]\,. \tag{4.70}$$

Equivalently, this is a Poisson 2-algebra, i.e. a Poisson algebra internal to $\mathbf{Ch}_\bullet(\mathbf{Vect})$ with bracket of degree -1.

Definition 4.9.5 (Semistrict Lie 2-algebra). A (Baez–Crans) 2-vector space $L \equiv (L_0, L_1)$ equipped with the following morphisms:

- an antisymmetric bilinear functor $[\cdot,\cdot] : L \times L \to L$ (the **bracket**), and
- a completely antisymmetric trilinear natural isomorphism

$$J_{x,y,z} : [[x, y], z] \to [x, [y, z]] + [[x, z], y]\,, \tag{4.71}$$

  called the **Jacobiator**.

These structures are required to satisfy the *Jacobiator identity* (which is just the *Zamolodchikov tetrahedron equation*). If the Jacobiator is trivial, a **strict** Lie 2-algebra is obtained. By further relaxing the antisymmetry, one can obtain the fully weak version of Lie 2-algebras (see for example the work by *Roytenberg*).

From the previous section, it follows that one can define (weak) Lie 2-algebras as 2-term chain complexes equipped with a coherent Lie bracket.

Alternative Definition 4.9.6 (Lie 2-algebra). Consider a 2-term chain complex in the category **FinVect**:

$$0 \longrightarrow L_1 \longrightarrow L_0 \longrightarrow 0\,. \tag{4.72}$$

This complex L is called a Lie 2-algebra if it comes equipped with the following structures:

- a chain map $[\cdot,\cdot] : L \otimes L \to L$ called the **bracket**,
- a chain homotopy $S : [\cdot,\cdot] \Rightarrow -[\cdot,\cdot] \circ \sigma$ called the **alternator**, and
- a chain homotopy

$$J : [\cdot,[\cdot,\cdot]] \Rightarrow [[\cdot,\cdot],\cdot] + [\cdot,[\cdot,\cdot]] \circ (\sigma \otimes \mathbb{1})\,, \tag{4.73}$$

  called the **Jacobiator**.

These chain homotopies are again required to satisfy higher coherence relations. From the previous definition, it follows that the vanishing of the alternator implies that L is semistrict. Analogously, a Lie 2-algebra for which the Jacobiator vanishes is said to be **hemistrict**. Note that this definition of weak Lie 2-algebras, when translated to the 2-vector space setting, would imply that the alternator and Jacobiator are merely natural transformations (and not isomorphisms)!

4.9.2 Lie n-algebras

Definition 4.9.7 (Semistrict Lie ω-algebra). By replacing internal categories by internal *ω-categories* and by relaxing the Jacobiator identity up to coherent homotopy, i.e. up to a completely antisymmetric quadrilinear modification which in turn satisfies an identity up to higher multilinear transfors, one obtains the definition of L_∞-algebras. Similar to A_∞-algebras, these too can be obtained as algebras over a suitable operad (however, in this case the operad is 'slightly' more complex: the cofibrant replacement of the *Lie operad*).

It can be shown that these structures are equivalent to the L_∞-algebras of *Stasheff* defined below.

Definition 4.9.8 (L_∞-algebra[8]). A graded vector space V equipped with a collection of morphisms $l_n : V^{\otimes n} \to V, n \in \mathbb{N}_0$ of degree $n-2$ subject to the relations

$$l_n(v_{\sigma(1)} \dots v_{\sigma(n)}) = \chi(\sigma; v_1, \dots, v_n) l_n(v_1 \dots v_n) \tag{4.74}$$

and

$$\sum_{\substack{i+j=n+1 \\ \sigma \in \mathsf{Unshuff}(i,j-1)}} (-1)^{i(j-1)} \chi(\sigma; v_1, \dots, v_n) l_i\big(l_j(v_{\sigma(1)} \cdots v_{\sigma(j)}) v_{\sigma(j+1)} \cdots v_{\sigma(n)}\big) = 0\,, \tag{4.75}$$

[8] Also called a **strong(ly) homotopy Lie algebra** (abbreviated to **sh Lie algebra**).

where Unshuff denotes the collection of unshuffles (Definition 2.2.36).

The l_1 map turns the L_∞-algebra into a chain complex. The l_2 map is a generalized Lie bracket since it is (graded-)antisymmetric. Higher l_n's can be identified with the Jacobiator and its generalizations. In the next section, a bottom-up approach will be given.

Remark 4.9.9. The definition can be rephrased in terms of graded maps $\hat{l}_n : \mathsf{Alt}^\bullet V \to V$.

Remark 4.9.10 (Curvature). The above definition can be generalized by including a nullary bracket l_0. Such L_∞-algebras are often said to be **curved**. The reason for this is that the coherence condition for l_0 says that

$$l_1 \circ l_1 = l_2(l_0, -)\,. \tag{4.76}$$

This terminology stems from the situation where l_1 is identified with the *exterior covariant derivative* on an *associated vector bundle* (see Formula 35.4.25).

Example 4.9.11 (Lie algebra). It can easily be checked that the L_∞-algebra with V concentrated in degree 1 is equivalent to the structure of an ordinary Lie algebra. Similarly, one obtains the notion of a Lie n-algebra by truncating an L_∞-algebra at degree n.

Property 4.9.12. 2-term L_∞-algebras or, equivalently, semistrict Lie 2-algebras are in correspondence with isomorphism classes of tuples $(\mathfrak{g}, V, \rho, l_3)$ where $\mathfrak{g}$ is a Lie algebra, (V, ρ) is Lie algebra representation of $\mathfrak{g}$ and l_3 is a V-valued *Lie algebra 3-cocycle* (see Section 32.6).

Proof (Sketch of construction). Using the representation ρ, one can extend the Lie bracket from $\mathfrak{g}$ to the complex $0 \to V \to \mathfrak{g} \to 0$ through the formulas $[g, v] := \rho(g)v$ and $[v, g] := -[g, v]$. The cocycle condition for l_3 gives rise to the Jacobiator. □

Example 4.9.13. If one chooses a finite-dimensional Lie algebra $\mathfrak{g}$ with the trivial representation on $\mathbb{R}$ (or, more generally, the underlying field of $\mathfrak{g}$), one obtains

$$H^3(\mathfrak{g}; \mathbb{R}) \cong \mathbb{R}\,. \tag{4.77}$$

The different classes can be represented by scalar multiples of the Killing cocycle (see Example 32.6.11). For every such scalar $\lambda \in \mathbb{R}$, one denotes the resulting Lie 2-algebra by $\mathfrak{g}_\lambda$.

Lie algebras and L_∞-algebras can also be dually characterized in terms of their *Chevalley–Eilenberg algebra* (see Definition 32.6.1).

Alternative Definition 4.9.14 (Lie algebra). Consider a finite-dimensional Lie algebra $\mathfrak{g}$. The transpose/dual of the Lie bracket $[\cdot,\cdot] : \mathfrak{g} \wedge \mathfrak{g} \to \mathfrak{g}$ is a morphism $\delta : \mathfrak{g}^* \to \mathfrak{g}^* \wedge \mathfrak{g}^*$:

$$\delta\omega(g, h) := \omega([g, h])\,. \tag{4.78}$$

In fact, it is not hard to see that this is exactly the Chevalley–Eilenberg differential of $\mathsf{CE}(\mathfrak{g})$ (see Definition 32.6.1). Conversely, given a semifree dgca $(\Lambda^\bullet V^*, \mathsf{d})$, for some finite-dimensional vector space V, one obtains a finite-dimensional Lie algebra by restricting the differential to V^* and taking the transpose. In fact, the nilpotency condition $\mathsf{d}^2 = 0$ is equivalent to the Jacobi identity.

More generally, by passing to graded vector spaces of finite type concentrated in positive degree, one can characterize L_∞-algebras as semifree DGCAs.

Alternative Definition 4.9.15 (L_∞-algebra). The (graded) Leibniz rule implies that the differential δ is completely defined by its restriction to the generators $V^* \leq \Lambda^\bullet V^*$. The differential can be decomposed as follows:

$$\delta t^a := -\sum_{k=1}^{+\infty} \frac{1}{k!}[t_{a_1}, \dots, t_{a_k}]_k^a \, t^{a_1} \wedge \cdots \wedge t^{a_k}\,, \tag{4.79}$$

where the basis t^a of V^* is dual to the basis t_a of V. Because δ is of degree 1, the coefficients $[\cdots]_k^a$ define a multilinear operator $[\cdots]_k : \Lambda^k V \to V$ of degree $n-1$. Some sources rephrase these brackets as morphism from the symmetric algebra $\mathsf{Sym}^\bullet V$, in which case their degree is just -1, cf. décalage (Remark 23.5.9).

The nilpotency condition $\delta^2 = 0$ implies a list of (quadratic) relations on the brackets $[\cdots]_k$ (with $\mathsf{d} := [\cdot]_1$):

$$\begin{aligned}
\mathsf{d}^2 &= 0 \\
\mathsf{d}[\cdot,\cdot]_2 &= [\mathsf{d}\cdot,\cdot]_2 + [\cdot,\mathsf{d}\cdot]_2 \\
[[v_1,v_2],v_3]_2 + \text{cyc. perm.} &= \mathsf{d}[v_1,v_2,v_3]_3 - [\mathsf{d}v_1,v_2,v_3]_3 - [v_1,\mathsf{d}v_2,v_3]_3 - [v_1,v_2,\mathsf{d}v_3]_3 \\
&\vdots
\end{aligned}$$

These relations can be interpreted as follows:

- d is a differential.
- d acts as a derivation with respect to the binary bracket.
- The Jacobi identity holds up to a chain homotopy (given by the ternary bracket).
- The higher relations are similar to the chain homotopy for the Jacobi identity.

When written out it in full detail, it can be checked that this is exactly the definition of an L_∞-algebra.

Definition 4.9.16 (Maurer–Cartan element). An element a of an L_∞-algebra V that satisfies the equation

$$\sum_{k=0}^{+\infty} \frac{1}{k!}[a, \dots, a]_k = 0\,. \tag{4.80}$$

For dg-Lie algebras, this reduces to the ordinary *Maurer–Cartan equation* (see Theorem 35.3.24):

$$\mathrm{d}a + \frac{1}{2}[a, a] = 0\,. \tag{4.81}$$

This is no coincidence since the complex $\Omega^\bullet(M) \otimes \mathfrak{g}$ of Lie algebra-valued differential forms on a smooth manifold M carries a canonical dg-Lie algebra structure.

4.10 Monoidal n-categories

Definition 4.10.1 (Monoidal n-category). In general, one can define a monoidal n-category as a one-object $(n+1)$-category, similar to how monoidal categories give one-object bicategories by delooping (Property 14.1.7). For the explicit definitions of monoidal bi- and tricategories, see the papers by Cheng and Gurski (2007) and Hoffnung (2011), respectively.

If one would put multiple compatible monoidal products on an n-category, by a version of the Eckmann–Hilton argument 3.5.1, all of these structures will be equivalent to a 'commutative' monoidal structure. By increasing the number of compatible structures, the 'commutativity' can be increased. This gives rise to the following definition which is stated in different terms (based on the *delooping hypothesis*).

Definition 4.10.2 (k-tuply monoidal n-categories). A pointed $(n+k)$-category (strict or weak) in which all parallel j-arrows for $j < k$ are equivalent. These categories form an $(n+k+1)$-category **kMonnCat**.

Example 4.10.3. For small values of k and n, the resulting structures coincide with some well-known constructions:

- $n = 0$:
  - $k = 0$: pointed set,
  - $k = 1$: monoid, and
  - $k \geq 2$: Abelian monoid.
- $n = 1$:
  - $k = 0$: 'pointed' category[9],
  - $k = 1$: monoidal category,
  - $k = 2$: braided monoidal category, and

[9] As in category with a specified element not as in category with a zero object (Definition 3.4.14).

– $k \geq 3$: symmetric monoidal category.

The stabilization occurring for higher values of k is the content of the following hypothesis[10] by Baez and Dolan (1995).

Theorem 4.10.4 (Stabilization hypothesis). *For values $k \geq n+2$, the structure of a k-tuply monoidal n-category becomes maximally symmetric. Formally, this means that the inclusion* $\mathbf{kMonnCat} \hookrightarrow (\mathbf{n}+\mathbf{2})\mathbf{MonnCat}$ *becomes an equivalence.*

4.10.1 Relation with group cohomology

See Section 5.6.1 for more information on group cohomology.

Consider a finite group G. As a first step, construct the group algebra $\mathbb{C}[G]$. As a monoid, one can consider this object as a G-graded monoidal 0-category. The ordinary multiplication $g * h = gh$ can be twisted to obtain a monoid $\mathbb{C}[G]^\omega$ with multiplication

$$g * h := e^{i\omega(g,h)} gh\,. \tag{4.82}$$

If associativity is still required to hold on the nose, one is led to the property that ω is in fact a group 2-cocycle. The equivalence classes of such twisted group algebras are then in correspondence with the second cohomology class $H^2(G; \mathrm{U}(1))$.

Before really going to higher category theory, one should first reflect on the different structures in the previous paragraph. Since the monoid is regarded as a monoidal category (call it M for convenience), one has a bifunctor $\mu : M \otimes M \to M$ (given by the twisted multiplication) that differs from the ordinary group multiplication by a phase. This phase can be viewed categorically as a natural isomorphism between the 'tensor products' in $\mathbb{C}[G]$ and M. At the same time, all the higher coherence conditions[11] (associativity, ...) are required to hold identically.

Now, drop the restriction on the product and take this to be a more general monoidal product bifunctor. To this end, replace the monoid $\mathbb{C}[G]$ by the G-graded monoidal category $\mathbf{Vect}_G$ and relax the associativity constraint up to a natural isomorphism α. When restricted to the simple objects of $\mathbf{Vect}_G$ this is given by a phase factor $e^{i\omega(g,h,k)}$. The pentagon condition for monoidal categories then implies that the function ω is a group 3-cocycle. In analogy with the case of monoids above, the equivalence classes of (twisted) monoidal structures on $\mathbf{Vect}_G$ is in correspondence with the third cohomology group $H^3(G; \mathrm{U}(1))$.

To go yet another step higher, move up a level in the chain of coherence conditions and relax the associativity constraint even more (for simplicity the one-object n-category

[10] For certain definitions of higher categories this has been proven in full generality.
[11] These can be parametrized by the *Stasheff polytopes/associahedra*.

point of view is adopted here). Instead of a natural isomorphism it only has to be an adjoint equivalence and at the same time the pentagon condition is replaced by an invertible modification. The coherence condition of this **pentagonator** then implies a classification of (twisted) monoidal bicategories, equivalent to $\mathbf{2Vect}_G^\omega$, by the fourth group cohomology $H^4(G; \mathsf{U}(1))$.

In a completely analogous way one can define more and more general structures. e.g. for monoidal tricategories one can translate the K_6-*associahedron* into an equation for an invertible perturbation which by the G-graded structure is equivalent to a group 5-cocycle.

Remark 4.10.5. This section is strongly related to the twisting procedure in n-dimensional *Dijkgraaf–Witten theories*.

Chapter 5

Homological algebra

References for this chapter are Hilton and Stammbach (1997); Weibel (1994).

Chapter Contents

5.1 Chain complexes **151**
5.2 Exact sequences **154**
5.3 Resolutions **155**
5.4 Derived functors **156**
5.4.1 Modules 158
5.5 Spectral sequences **160**
5.5.1 Exact couples 160
5.5.2 Filtered complexes 162
5.5.3 Convergence 163
5.6 Examples **164**
5.6.1 Group cohomology 164
5.6.2 Hochschild and cyclic homology ♣ 166

5.1 Chain complexes

Definition 5.1.1 (Chain complex). Let $\mathbf{A}$ be an additive category (often an Abelian category) and consider a collection $\{C_k\}_{k\in\mathbb{Z}}$ of objects and a collection $\{\partial_k : C_k \to C_{k-1}\}_{k\in\mathbb{Z}}$ of morphisms in $\mathbf{A}$ such that for all $k \in \mathbb{Z}$:

$$\partial_k \circ \partial_{k+1} = 0. \tag{5.1}$$

This structure is called a chain complex[1] and the morphisms ∂_k are called the **boundary operators** or **differentials**. Elements of $\mathsf{im}(\partial_k)$ are called **boundaries** or **exact elements** and elements of $\mathsf{ker}(\partial_k)$ are called **cycles** or **closed elements**. The chain complex $\{(C_k, \partial_k)\}_{k\in\mathbb{Z}}$ is often denoted by $(C_\bullet, \partial_\bullet)$ or simply by $C_\bullet$ if the choice of boundary operators is clear.

Morphisms between chain complexes are called **chain maps** and they are defined as collections of morphisms $\{f_k : C_k \to D_k\}_{k\in\mathbb{Z}}$ such that for all $k \in \mathbb{Z}$ the following equation holds:

$$\partial'_k \circ f_k = f_{k-1} \circ \partial_k \,, \tag{5.2}$$

where ∂_k, ∂'_k are the boundary operators of $C_\bullet$ and $D_\bullet$, respectively. Given an additive category $\mathbf{A}$, one writes $\mathbf{Ch}(\mathbf{A})$ for the category of chain complexes and chain maps in $\mathbf{A}$.

Remark 5.1.2 (Reversal). Given a chain (resp. cochain) complex C one can easily construct a cochain (resp. chain) complex $\widetilde{C}$ by setting $\widetilde{C}_k := C_{-k}$.

Definition 5.1.3 (Chain homology). Given a chain complex $C_\bullet$, one can define a sequence of **homology groups** $H_n(C_\bullet)$. Since $\partial^2 = 0$, the kernel $\mathsf{ker}(\partial_k)$ is a subgroup of the image $\mathsf{im}(\partial_{k+1})$ and it is even a normal subgroup. This way one can define the quotient group:

$$H_k(C_\bullet) := \frac{\mathsf{ker}(\partial_k)}{\mathsf{im}(\partial_{k+1})} \,. \tag{5.3}$$

The kernel in this definition, the group of k**-cycles**, is denoted by $Z_k(C_\bullet)$. The image, the group of $(k+1)$**-boundaries**, is denoted by $B_k(C_\bullet)$. The homology groups themselves also form a chain complex $H_\bullet(C_\bullet)$, but with trivial differentials.

Definition 5.1.4 (Quasi-isomorphism). A chain map for which the induced morphisms on homology are isomorphisms.

Definition 5.1.5 (Chain homotopy). Two chain maps $f, g : C_\bullet \to D_\bullet$ are said to be chain-homotopic if there exists a chain map $s : C_\bullet \to D_\bullet$ such that the following equation is satisfied:

$$f - g = s \circ \partial_C + \partial_D \circ s \,. \tag{5.4}$$

A chain map that is homotopy-equivalent to the zero map is said to be **null-homotopic**. If there exist two chain maps $f : C_\bullet \leftrightarrows D_\bullet : g$ such that both $f \circ g$ and $g \circ f$ are chain-homotopic to the identity, $C_\bullet$ and $D_\bullet$ are said to be **(chain-)homotopy equivalent**.

Property 5.1.6. Chain-homotopic maps induce coinciding maps in homology. In particular, every (chain-)homotopy equivalence is a quasi-isomorphism.

[1] A **cochain complex** is constructed similarly with an ascending order $\partial_k : C_k \to C_{k+1}$.

Corollary 5.1.7 (Vanishing homology). If a null-homotopic chain map $f : C_\bullet \to C_\bullet$ exists, then $H_\bullet(C_\bullet)$ vanishes.

Definition 5.1.8 (Differential modulo differential). Consider a chain complex $(C_\bullet, \partial_\bullet)$ together with a chain endomorphism d. This endomorphism is said to be a differential modulo ∂ if it satisfies the following conditions:

1. $\mathsf{d}\partial + \partial\mathsf{d} = 0$, and
2. $\mathsf{d}^2 = [\partial, D]$ for some other chain endomorphism D, i.e. d^2 is ∂-exact in $\mathsf{End}(C_\bullet)$.

The first condition states that d descends to a chain endomorphism on $H_\bullet(C_\bullet)$. The second condition states that d is actually a differential on $H_\bullet(C_\bullet)$. The resulting homology theory is denoted by $H_\bullet(\mathsf{d} \mid H_\bullet(C_\bullet))$.

The following definitions use the language of Chapter 4.

Definition 5.1.9 (Differential graded algebra). A differential graded algebra (often called a **dg-algebra**, **DGA** or **dga**) is a (co)chain complex that carries the structure of an algebra where the differential acts as a derivation. Equivalently, it is a graded algebra equipped with a nilpotent derivation of degree ± 1.

Definition 5.1.10 (Connective DGA). A DGA $A_\bullet$ with vanishing (co)homology in negative degree, i.e. $H_{<0}(A_\bullet) = 0$. For every connective DGA, one can find a quasi-isomorphic DGA concentrated in nonnegative degree.

Definition 5.1.11 (Semifree DGA). A DGA for which the underlying graded algebra is isomorphic (as a graded algebra) to the tensor algebra over a graded vector space.

Definition 5.1.12 (Differential graded-commutative algebra). A graded-commutative differential graded algebra. This is often abbreviated as **DGCA** or **dgca**.

Definition 5.1.13 (Semifree DGCA). A DGCA for which the underlying graded-commutative algebra is isomorphic (as a graded-commutative algebra) to the exterior algebra over a graded vector space.

Definition 5.1.14 (Differential ideal). A subset $I \subseteq (A, \mathsf{d})$ of a differential (graded) algebra satisfying:

1. I is an ideal of the underlying associative algebra of $A_\bullet$.
2. I is closed under differentials, i.e. $\mathsf{d}I \subseteq I$.

Definition 5.1.15 (Minimal model). Let $(C_\bullet, \partial_\bullet)$ be a (cohomological) DGCA of finite type. A model for $C_\bullet$ is a quasi-isomorphism $\rho : (A_\bullet, \mathsf{d}_\bullet) \to (C_\bullet, \partial_\bullet)$ from a semifree DGCA $(A_\bullet, \mathsf{d}_\bullet)$. This model is said to be **minimal** if $A_\bullet$ is freely generated in degrees ≥ 2 and satisfies $\mathsf{d}A \subseteq \Lambda^{\geq 2}A$.

Remark 5.1.16 (Model structure on DGCAs ♣). By Example 7.1.76, the (minimal) mod-

els of DGCAs are (minimal) Sullivan algebras. From a model theory point of view, the (minimal) Sullivan algebras are the cofibrant objects and the (minimal) models are the cofibrant replacements.

5.2 Exact sequences

Definition 5.2.1 (Exact sequence). Let $\mathbf{A}$ be an additive category and consider a sequence of objects and morphisms in $\mathbf{A}$:

$$C_0 \xrightarrow{\Phi_1} C_1 \xrightarrow{\Phi_2} \cdots \xrightarrow{\Phi_n} C_n \,. \tag{5.5}$$

This sequence is said to be exact if for every $k \in \mathbb{N}$:

$$\mathsf{im}(\Phi_k) = \mathsf{ker}(\Phi_{k+1}) \,. \tag{5.6}$$

In particular this means that $\Phi_{k+1} \circ \Phi_k = 0$ for all $k \in \mathbb{N}$, which in turn implies that exact sequences are a special type of chain complexes (Definition 5.1.1).

Definition 5.2.2 (Short exact sequence). An exact sequence with exactly three nonzero terms:

$$0 \longrightarrow C_0 \xrightarrow{\Phi_1} C_1 \xrightarrow{\Phi_2} C_3 \longrightarrow 0. \tag{5.7}$$

Usually, all other exact sequences are said to be **long**.

Property 5.2.3 (Morphisms in exact sequences). By looking at some simple examples, one can derive some important constraints for certain exact sequences. Consider the sequence

$$0 \longrightarrow C \xrightarrow{\Phi} D \,. \tag{5.8}$$

This sequence can only be exact if Φ is an injective morphism (monomorphism). This follows from the fact that the only element in the image of the map $0 \to C$ is 0 because the map is a morphism. It follows that the kernel of Φ is trivial and, hence, that Φ is injective.

Analogously, the sequence

$$C \xrightarrow{\Psi} D \longrightarrow 0$$

is exact if and only if Ψ is a surjective morphism (epimorphism). This follows from the fact that the kernel of the map $D \to 0$ is all of C, which implies that Ψ is surjective (by exactness).

Combining these two cases shows that

$$0 \longrightarrow C \xrightarrow{\Sigma} D \longrightarrow 0$$

is exact if and only if Σ is a **bimorphism** (if $\mathbf{A}$ is Abelian, Σ is even an isomorphism by Property 3.4.63).

5.3 Resolutions

Consider some Abelian category $\mathbf{A}$ and let $\mathbf{Ch}(\mathbf{A})$ denote its category of chain complexes.

Definition 5.3.1 (Acyclic complex). A chain complex $C_\bullet \in \mathbf{Ch}(\mathbf{A})$ is said to be acyclic if the sequence

$$\cdots \longrightarrow C_{k+1} \longrightarrow C_k \longrightarrow C_{k-1} \longrightarrow \cdots$$

is exact or, equivalently, if the homology complex $H_\bullet(C_\bullet)$ vanishes.

Remark. Some references, especially the older ones, use a slightly different definition of acyclicity. In their definition, the sequence is exact except in degree 0, i.e. $H_0(C_\bullet) \neq 0$.

Definition 5.3.2 (Resolution). Consider an object X in $\mathbf{A}$. A resolution of X is given by an acyclic chain complex in $\mathbf{A}$ of the form

$$\cdots \longrightarrow C_1 \longrightarrow C_0 \xrightarrow{\varepsilon} X \longrightarrow 0\,. \tag{5.9}$$

This also implies that X is the zeroth homology group of the chain complex $C_\bullet := \{C_k\}_{k\geq 0}$. The morphism $\varepsilon : C_0 \to X$ is often called the **augmentation map** and the complex $C_\bullet \to X \to 0$ is called the **augmentation** of $C_\bullet$.

In practice it is often convenient to restrict to a specific type of resolution. For example, by considering chain complexes with only injective or projective objects (Figs. 3.2a and 3.2b), one obtains injective or projective resolutions. If every object in A admits a projective (resp. injective) resolution, $\mathbf{A}$ is said to **have enough projectives** (resp. **injectives**).

Theorem 5.3.3 (Homological perturbation). *Consider a resolution $(C_\bullet, \partial_\bullet)$ and denote the grading in $C_\bullet$ by r, i.e. $r(x) = k$ whenever $x \in C_k$. Furthermore, consider a differential d modulo ∂ of degree 0 and denote the associated grading by deg. There exists a differential s satisfying the following properties:*

- $\mathsf{deg}(s) - r(s) = 1$, *and*
- $s = \delta + \mathsf{d} + \sum_{i=1}^{+\infty} s_{(i)}$, *where* $r(s_{(i)}) = i$ *and* $\mathsf{deg}(s_{(i)}) = i + 1$.

Moreover, any differential that satisfies these properties has the same homology as d*:*

$$H_\bullet(s) \cong H_\bullet(\mathsf{d} \mid H_\bullet(C_\bullet)) \equiv H_\bullet(\mathsf{d} \mid H_0(C_\bullet)). \tag{5.10}$$

Remark 5.3.4. Note that the cohomology of s is entirely determined by δ and d, the higher-degree terms $s_{(i)}$ are only required to ensure the nilpotency of s.

5.4 Derived functors

Given an additive functor (Definition 4.3.5), one can define its **prolongation** to the category of chain complexes.

Definition 5.4.1 (Prolongation). Let $F : \mathbf{A} \to \mathbf{A}'$ be an additive functor. The prolongation of F is a functor $\overline{F} : \mathbf{Ch}(\mathbf{A}) \to \mathbf{Ch}(\mathbf{A}')$ obtained by applying F to every object in a chain complex and to every diagram in the definition of a chain map. As is common, by abuse of notation the prolongation will also often be denoted by F.

To understand and unify the various long exact sequences in (co)homology and to formulate general statements about these theories, one can introduce the concept of derived functors.

Definition 5.4.2 (Left derived functor). Let $\mathbf{A}$ be an Abelian category with enough projectives and consider a right exact functor $F : \mathbf{A} \to \mathbf{A}'$. The left derived functors L_iF are defined in the following way. Pick an object X in A and construct a projective resolution $P_\bullet \xrightarrow{\varepsilon} X \to 0$. Then, apply the prolongation to this resolution and construct the homology of the resulting chain complex:

$$L_iF(X) := H_i(FP_\bullet)\,. \tag{5.11}$$

In particular, $L_0F(X) = F(X)$.

Right derived functors of left exact functors can be constructed dually by choosing an injective resolution, applying the prolongation and taking the cohomology of the resulting cochain complex. In the remainder of this section, all statements will be given for right exact functors and left derived functors.

Remark 5.4.3 (Contravariant functors). The above construction was given for covariant functors. For contravariant functors one defines the derived functors as those of the opposite functor. This is equivalent to starting with an injective (resp. projective) resolution for the calculation of left (resp. right) derived functors since injective objects are projective in the opposite category and similarly homology becomes cohomology in the opposite category.

Property 5.4.4 (Exact functors). If F is exact, the above construction immediately implies that the derived functors L_i vanish for $i \geq 1$.

Property 5.4.5 (Projective objects). Consider a right exact functor F together with its left derived functors L_iF. If an object P is projective, then $L_iF(P) = 0$ for all $i \geq 1$. This can easily be shown by noting that every projective object P admits a projective resolution of the form

$$\cdots \longrightarrow 0 \longrightarrow 0 \longrightarrow P \longrightarrow P \longrightarrow 0\,. \tag{5.12}$$

Now, of course, one could wonder why the resolutions used in the construction of derived functors are required to be projective or injective. This seems to be a very strong requirement. The reason is that, when using the above definitions, the result is independent of the resolution used in the sense that the derived functors are naturally isomorphic. However, in certain situations one might want to work with a more general resolution. For example, in the next section, when considering the tensor product, it would be useful if one could just work with *flat* modules.

Definition 5.4.6 (Acyclic resolution). Consider a right exact functor F together with its left derived functors L_iF. An object X is said to be F**-acyclic** if $L_iF(X) = 0$ for all $i \geq 1$. A resolution of an object is said to be F-acyclic if all objects in the resolution are F-acyclic.

Property 5.4.7 (Derived functors for acyclic resolutions). Derived functors of a right exact (resp. left exact) functor F constructed using an F-acyclic resolution are isomorphic to those obtained using a projective (resp. injective) resolution.

One of the motivating properties of derived functors are the long exact sequences in (co)homology. All of these are a result of the following property:

Property 5.4.8 (Long exact sequence). Let $F : \mathbf{A} \to \mathbf{A}'$ be a right exact functor (the left exact case proceeds in a similar way). Consider a short exact sequence in $\mathbf{A}$:

$$0 \longrightarrow A \longrightarrow B \longrightarrow C \longrightarrow 0\,. \tag{5.13}$$

Now, choose projective resolutions for A and C. By the *horseshoe lemma*, one obtains a projective resolution for B that fits in a short exact sequence of chain complexes:

$$0 \longrightarrow A_\bullet \longrightarrow B_\bullet \longrightarrow C_\bullet \longrightarrow 0\,. \tag{5.14}$$

Since F is additive and the above sequence is exact, the induced complex

$$0 \longrightarrow FA_\bullet \longrightarrow FB_\bullet \longrightarrow FC_\bullet \longrightarrow 0 \tag{5.15}$$

is also exact and so the *zig-zag lemma* is applicable. This theorem gives the following long exact sequence in homology:

$$\cdots \longrightarrow H_i(FB_\bullet) \longrightarrow H_i(FC_\bullet) \longrightarrow H_{i-1}(FA_\bullet) \longrightarrow H_{i-1}(FB_\bullet) \longrightarrow \cdots\,. \tag{5.16}$$

These homology groups are by definition the same as the left derived functors ($L_i = H_i \circ F$) and, accordingly, a long exact sequence relating the different derived functors is obtained.

Corollary 5.4.9. The above long exact sequence of derived functors shows that the first derived functor gives the obstruction to F being exact. Since exact functors have vanishing derived functors, one obtains the following result:

$$L_1F = 0 \implies L_iF = 0 \qquad \forall i \geq 1\,, \tag{5.17}$$

and, more generally:

$$L_i F = 0 \implies L_j F = 0 \qquad \forall j \geq i\,. \tag{5.18}$$

5.4.1 Modules

Consider the tensor and hom-bifunctors $- \otimes -$ and $\mathsf{Hom}(-,-)$ in the category $\mathbf{Mod}_R$ of modules over some ring. The tensor functor is right exact in both arguments, while the hom-functor is left exact in both arguments and, hence, one can construct the associated left and right derived functors. For simplicity, everything will be constructed with respect to the first argument of these bifunctors. A proof that the derived functors are *balanced*, i.e. that one can use a projective resolution for either argument and obtain isomorphic results, can be found in the references cited at the beginning of the chapter.

Definition 5.4.10 (Tor-functor). Consider a ring R and an R-module A. The Tor-functors $\mathsf{Tor}_n^R(-,A)$ are defined as the left derived functors of the tensor functor $- \otimes_R A$.

Definition 5.4.11 (Ext-functor). Consider a ring R and an R-module A. The Ext-functors $\mathsf{Ext}_R^n(-,A)$ are defined as the right derived functors of the hom-functor $\mathsf{Hom}_R(-,A)$.

Definition 5.4.12 (Flat module). An R-module M such that the induced tensor functor

$$- \otimes_R M : R\mathbf{Mod} \to R\mathbf{Mod} \tag{5.19}$$

is exact. By Property 5.4.4, this implies that a flat module is $\otimes$-acyclic (Definition 5.4.6), which, in turn, implies by Property 5.4.7 that these modules can be used to construct a good resolution for calculating Tor-functors. A ring morphism $f : R \to S$ is said to be **flat** if it turns S into a flat R-module.

Definition 5.4.13 (Bar resolution). Let A be an associative algebra. The bar complex (or **standard complex**) of A is the sequence

$$\cdots \to A^{\otimes 3} \to A^{\otimes 2} \to A \to 0\,, \tag{5.20}$$

with $B_k(A) := A^{\otimes(k+2)} \cong A^e \otimes A^{\otimes n}$, where A^e is the enveloping algebra (Definition 2.4.59) and the isomorphism is given by:

$$a_0 \otimes \cdots \otimes a_{k+1} \mapsto (a_0 \otimes a_{k+1}) \otimes (a_1 \otimes \cdots \otimes a_k)\,. \tag{5.21}$$

The differentials are defined as follows:

$$\mathsf{d}_k(a_0 \otimes \cdots \otimes a_{k+1}) := \sum_{i=0}^{k} a_0 \otimes \cdots \otimes a_i a_{i+1} \otimes \cdots \otimes a_{k+1}\,. \tag{5.22}$$

If A is free, so is the bar complex and, hence, it gives a (free) resolution of A.

Definition 5.4.14 (Koszul complex). Consider a commutative ring R together with a free rank-r module M over R. For every morphism $s : M \to R$ one constructs the Koszul complex $K(s)$ as follows:

$$0 \longrightarrow \Lambda^r M \longrightarrow \Lambda^{r-1} M \longrightarrow \cdots \longrightarrow M \xrightarrow{s} R \longrightarrow 0\,, \tag{5.23}$$

where the exterior powers $\Lambda^k M$ are defined as in Section 23.4.4, i.e. they are the free modules spanned by totally antisymmetric k-tuples in M. The differentials are defined as

$$\mathsf{d}_k(m_1 \wedge \ldots \wedge m_k) := \sum_{i=1}^{k} (-1)^{k+1} s(m_i)\, m_1 \wedge \ldots \wedge \widehat{m_i} \wedge \ldots \wedge m_k\,, \tag{5.24}$$

where the caret $\hat{\cdot}$ means that this element is omitted. It is clear that $\mathsf{d}_1 = s$. The homology of this complex is called the **Koszul homology** of s.

Example 5.4.15. Every finite sequence $(x_1, \ldots, x_n)$ in R, interpreted as a choice of basis for R^n, defines a morphism $s : R^n \to R$ by

$$s : R^n \to R : (r_1, \ldots, r_n) \mapsto r_1 x_1 + \cdots + r_n x_n\,. \tag{5.25}$$

The associated Koszul complex is denoted by $K(x_1, \ldots, x_n)$.

Property 5.4.16 (Koszul resolution). Let R be a commutative ring. If $(x_1, \ldots, x_n)$ is a **regular sequence** on R, i.e. for every $i \leq n$ the element x_i is a nonzero divisor of $R/(x_1, \ldots, x_{i-1})$, the Koszul homology of $K(x_1, \ldots, x_n)$ satisfies:

$$H_{i\geq 1}(K(x_1, \ldots, x_n)) = 0\,, \tag{5.26}$$

i.e. $K(x_1, \ldots, x_n)$ is a resolution of $R/(x_1, \ldots, x_n)$, called the Koszul resolution. By the very construction of the Koszul complex, it is even a free resolution.

Property 5.4.17 (Koszul–Tate resolution). Consider a commutative ring R with an ideal I. For any element $x \in I$, one can construct a differential on the polynomial algebra $R[t]$ on a formal generator t by setting $\partial t := x$. Because of this definition, the homology class of x in $R[t]$ vanishes. This procedure is said to 'kill' the homology of x.

In a similar way one can also kill the higher homology of I. If[2] $I \equiv (x_1, \ldots, x_k)$, one can consider the Koszul complex $(X^0, \mathsf{d}^0) := (K(x_1, \ldots, x_k), \mathsf{d})$. Its homology is exactly the quotient A/I. However, since the sequence is not necessarily regular, the higher homology groups need not vanish. To this end, choose a generating set $(x'_1, \ldots, x'_l)$ of $H_1(X^0, \mathsf{d}^0)$. Now, consider the Koszul complex $(X^1, \mathsf{d}^1) := (K(x_1, \ldots, x_k, x'_1, \ldots, x'_l), \mathsf{d}')$ induced by the morphism

$$s' : R^{k+l} \to R : (r_1, \ldots r_{k+l}) \mapsto r_1 x_1 + \cdots + r_k x_k + r_{k+1} x'_1 + \cdots + r_{k+l} x'_l\,, \tag{5.27}$$

[2] If R is Noetherian, this is always possible.

where the generators x_i are of degree 1 and the generators x_i' are degree 2 (in the definition of the Koszul complex one thus needs to replace the Grassmann algebra by a graded-commutative algebra (Example 23.5.8)). It should be clear that $H_0(X^1, \mathsf{d}^1) \cong R/I$ and $H_1(X^1, \mathsf{d}^1) = 0$. The direct limit of this construction is called the Koszul–Tate resolution of (R, I).

5.5 Spectral sequences

In this section the homological convention is adopted, i.e. differentials lower the degree.

Definition 5.5.1 (Spectral sequence). Consider a collection $\{(E_n, \mathsf{d}_n)\}_{n\in\mathbb{N}}$ of differential objects. This collection is called a spectral sequence if it satisfies

$$E_{n+1} \cong H(E_n, \mathsf{d}_n) \tag{5.28}$$

for every $n \in \mathbb{N}$. A morphism of spectral sequences is a collection of morphisms $(\varphi_n)_{n\in\mathbb{N}}$ satisfying

1. $\varphi_n \circ \mathsf{d}_n = \mathsf{d}_n' \circ \varphi_n$, and
2. $\varphi_{n+1} = H(\varphi_n)$.

The objects E_i are often called the **pages** or **terms** of the spectral sequence.

5.5.1 Exact couples

Definition 5.5.2 (Exact couple). A tuple $(A, B, \alpha, \beta, \gamma)$ that fits in a commutative diagram of the form 5.1a. A morphism of exact couples is a pair of morphisms $(f, g) : (A, B) \to (A', B')$ that fit in a commutative diagram of the form 5.1b.

$$\begin{tikzcd}
A \arrow[r, "\alpha"] & A \arrow[dl, "\beta"] \\
B \arrow[u, "\gamma"] &
\end{tikzcd}$$

(a) Exact couple.

$$\begin{tikzcd}
A' \arrow[rr, "\alpha'"] & & A' \\
& A \arrow[ul, "f"'] \arrow[r, "\alpha"] & A \arrow[ur, "f"] \arrow[d, "\beta"] \\
& B \arrow[u, "\gamma"] \arrow[d, "g"] & \\
& B \arrow[uul, "\gamma'"] \arrow[uur, "\beta'"'] &
\end{tikzcd}$$

(b) Morphism of exact couples.

Figure 5.1: Category of exact couples.

From any exact couple $(A, B, \alpha, \beta, \gamma)$ one can construct a spectral sequence using the following prescription:

$$E_0 := B \tag{5.29}$$

$$\mathsf{d}_0 := \beta \circ \gamma \tag{5.30}$$

$$\vdots$$

$$E_n := \frac{\gamma^{-1}(\alpha^n(A))}{\beta(\alpha^{-n}(0))} \tag{5.31}$$

$$\mathsf{d}_n := \beta \circ \alpha^{-n} \circ \gamma \tag{5.32}$$

It is not hard to see that $E_{n+1} = H(E_n, \mathsf{d}_n)$, so this construction gives a functor from the category of exact couples to the category of spectral sequences. The higher exact couples $(\alpha^n D, E_n, \ldots)$ are sometimes called **derived couples**.

One can also define the term E_∞ using the following limit procedure. For every $n \in \mathbb{N}$, take the elements in E_n that are closed under d_n and call these $E_{n,n+1}$. Since there exists a canonical surjection $E_{n,n+1} \to E_{n+1}$, one can look at all the elements in $E_{n,n+1}$ for which the image in E_{n+1} is closed under d_{n+1}. Call the set of these elements $E_{n,n+2}$. The elements that remain after taking the limit of this operation form the set $E_{n,\infty}$. Now, take the direct limit of the $E_{n,\infty}$ to obtain E_∞. This is equivalent to

$$E_\infty := \frac{\bigcap_i Z(E_i)}{\bigcup_i B(E_i)}, \tag{5.33}$$

i.e. E_∞ contains the equivalence classes of elements that are cycles for all d_n but boundaries for none. If E_∞ is the associated graded object of some filtered object G, one says that the spectral sequence **converges** to G.

Now, consider a differential object (C, d) together with a filtration $\{F_n C\}_{n\in\mathbb{N}}$. Definition 1.6.9 of a filtration immediately gives a short exact sequence for every $n \in \mathbb{N}$:

$$0 \longrightarrow F_{n-1}C \longrightarrow F_n C \longrightarrow F_n C / F_{n-1} C \longrightarrow 0\,. \tag{5.34}$$

This short sequence in turn gives rise to a long exact sequence in homology, which can be expressed as an exact triangle, and this triangle further leads to an exact couple:

$$\begin{tikzcd} D \arrow[rr, "\alpha"] & & D \arrow[dl, "\beta"] \\ & E \arrow[ul, "\gamma"] & \end{tikzcd} \tag{5.35}$$

where $D_n = H(F_nC)$ and $E_n = H(F_nC/F_{n-1}C)$. From a more abstract, yet more useful, point of view one can consider the object E as a functor from the category of filtered (differential) objects to the category of graded objects. As such it is constructed from the composition of the homology functor H and the 'associated graded object'-functor

$$\mathsf{Gr} : C \mapsto \left\{G_nC := F_nC/F_{n-1}C\right\}_{n\in\mathbb{N}}. \tag{5.36}$$

On the other hand, one could of course also construct the composition $\mathsf{Gr} \circ H$ that first maps a differential object to its homology object and then builds the graded object associated to the filtration

$$F_nH(C) := \mathsf{im}\big(H(F_nC) \to H(C)\big).$$

Some straightforward questions can arise at this point: *"How are the functors $H \circ \mathsf{Gr}$ and $\mathsf{Gr} \circ H$ related?"*, *"Do they coincide?"*, ... The latter question is easily answered: *"No, they do not."* However, they can be related and this exactly happens through a spectral sequence that says how the homology of the graded object associated to C can be related to the homology of C itself.

5.5.2 Filtered complexes

For the remainder of this section only graded differential objects will be considered, i.e. $(C_\bullet, \mathsf{d}_\bullet) = (\{C_p\}_{p\in\mathbb{Z}}, \mathsf{d})$ such that $\mathsf{d}C_p \subseteq C_p$. In this case the exact couple consist of $D_{p,q} := H_q(F_pC_\bullet)$ and $E_{p,q} := H_q(G_pC_\bullet)$ and, consequently, the objects are bigraded. The filtration is also required to be compatible with the differential, i.e. $\mathsf{d}F_iC_j \subseteq F_iC_{j-1}$.

Remark. In contrast to most of the literature, the 'complementary convention', i.e. the convention where $p + q$ denotes the total degree and hence $E_{p,q} = H_{p+q}(G_pC_\bullet)$, is not adopted.

Before introducing an expression for a general page E_r, the terms of degree zero and one are considered to get some intuition. The differential on E_0 is given by

$$\mathsf{d}_0 : \frac{F_pC_q}{F_{p-1}C_q} \to \frac{F_pC_{q-1}}{F_{p-1}C_{q-1}} \tag{5.37}$$

and is induced by the differential d on $C_\bullet$. The kernel of this map is clearly given by all elements $x \in F_pC_q$ such that $\mathsf{d}x = 0 \mod F_{p-1}C_{q-1}$ (with the additional remark that one also has to take the quotient by $F_{p-1}C_q$). As a result one finds that the homology $E_1 = H(E_0, \mathsf{d}_0)$ is given by

$$E_1^{p,q} := \frac{\{x \in F_pC_q \mid \mathsf{d}x \in F_{p-1}C_{q-1}\}}{F_{p-1}C_q + \mathsf{d}F_pC_{q+1}}. \tag{5.38}$$

The first term in the denominator was already explained above. The second term comes from the $\mathsf{im}(\mathsf{d}_0)$-part in the definition of $H(E_0, \mathsf{d}_0)$. One might suspect that some data

is missing since the relevant map $\mathrm{d}_0^{p,q+1}$ goes from $\frac{F_pC_{q+1}}{F_{p-1}C_{q+1}}$ to $\frac{F_pC_q}{F_{p-1}C_q}$. However, the image of $F_{p-1}C_{q+1}$ is a subspace of $F_{p-1}C_q$ and this is already included in the first term, so one might as well work with all of F_pC_{q+1}.

For arbitrary $r > 0$, one defines the page E_r as follows:

$$E_r^{p,q} := \frac{\{x \in F_pC_q \mid \mathrm{d}x \in F_{p-r}C_{q-1}\}}{F_{p-1}C_q + \mathrm{d}F_{p+r-1}C_{q+1}} . \tag{5.39}$$

To relate this more to the usual notions of (co)homology, one can rephrase this in terms of (co)chains, (co)cycles and (co)boundaries. Consider again a filtered complex $F_\bullet C_\bullet$. The following definitions are used:

- The elements of G_pC_q are called the (p,q)-**chains** (in filtering degree p).
- The elements of

  $$Z_{p,q}^r := \{c \in G_pC_q \mid \mathrm{d}c = 0 \bmod F_{p-r}C_{q-1}\}$$

  are called r-**almost** (p,q)-**cycles**.
- The elements of

  $$B_{p,q}^r := \mathrm{d}F_{p+r-1}C_{q+1}$$

  are called r-**almost** (p,q)-**boundaries**.

It is then easy to see that the page E^r satisfies

$$E_{p,q}^r = Z_{p,q}^r / B_{p,q}^r , \tag{5.40}$$

i.e. the homology is given by the quotient of the cycles by the boundaries. All these objects fit in a nice sequence of inclusions:

$$B_{p,q}^0 \hookrightarrow \cdots \hookrightarrow B_{p,q}^\infty \hookrightarrow Z_{p,q}^\infty \hookrightarrow \cdots \hookrightarrow Z_{p,q}^0 . \tag{5.41}$$

5.5.3 Convergence

Definition 5.5.3 (Limit term). Consider a spectral sequence $\{E_{p,q}^r\}$. If there exists a for every two integers $p, q \in \mathbb{Z}$ an integer $r(p,q) \in \mathbb{N}$ such that for all $r \geq r(p,q)$:

$$E_{p,q}^r \cong E_{p,q}^{r(p,q)} , \tag{5.42}$$

the object $E^\infty := \{E_{p,q}^{r(p,q)}\}$ is called the **limit term** and the sequence is said to **abut** to E^∞.

Example 5.5.4 (Collapsing sequence). If there exists an integer $r \in \mathbb{N}$ such that for all $s \geq r : d_s = 0$, the sequence is said to **collapse** at r and E^r is a limit term. A common example is where the nonvanishing elements of a term are concentrated in a single row or column.

Definition 5.5.5 (Convergence). A spectral sequence $E^r_{p,q}$ is said to converge to a graded object $H_\bullet$ with filtering $F_\bullet H_\bullet$, denoted by

$$E^r_{p,q} \Rightarrow H_\bullet\,,$$

if

$$E^\infty_{p,q} \cong G_p H_q \qquad \forall p,q \in \mathbb{Z}\,. \tag{5.43}$$

Definition 5.5.6 (Bounded sequence). A spectral sequence is said to be bounded if for all numbers $n, r \in \mathbb{Z}$, there only exists a finite number of nonvanishing elements of the form $E^r_{k,n-k}$. A common example are the **first quadrant spectral sequences** where the only nonvanishing elements have $p, q \geq 0$.

Property 5.5.7. Every bounded spectral sequence abuts.

Property 5.5.8 (Filtered complex). If the spectral sequence of a filtered complex $F_\bullet C_\bullet$, it converges to the chain homology of the complex:

$$E^r_{p,q} \Rightarrow H_\bullet(C). \tag{5.44}$$

5.6 Examples

5.6.1 Group cohomology

In this section, an important application of derived functors is studied. In fact, historically, this was one of the motivating examples. In different areas of mathematics and physics, the concept of group cohomology pops up. Some examples are the obstruction to group extensions, the classification of projective representations and the application of these concepts to the study of *symmetry-protected topological order* in condensed matter physics (see Chapter 65). Although a very concrete description exists (which will be given towards the end of this section), the more abstract definition is given first.

Every G-module (Example 2.3.7) can be regarded as a module over the group ring $\mathbb{Z}[G]$, i.e. there exists an equivalence of categories between **Ab** and $\mathbb{Z}[G]$**Mod**. Assuming the axiom of choice, every module category over a ring has enough projectives and, hence, it makes sense to define group (co)homology using derived functors in $\mathbb{Z}[G]$**Mod**. An explicit construction of a resolution that is not just $\mathbb{Z}[G]$-projective but even $\mathbb{Z}[G]$-free will be given.

The homology and cohomology of a finite group G with coefficients in a G-module A is defined using the Ext- and Tor-functors defined above:

$$\begin{aligned} H^\bullet(G;A) &:= \mathsf{Ext}^\bullet_{\mathbb{Z}[G]}(\mathbb{Z},A)\,, \\ H_\bullet(G;A) &:= \mathsf{Tor}^{\mathbb{Z}[G]}_\bullet(A,\mathbb{Z})\,, \end{aligned} \tag{5.45}$$

where $\mathbb{Z}$ carries the trivial G-module structure. To explicitly calculate the (co)homology groups, one has to find an acyclic resolution of $\mathbb{Z}$.

Construction 5.6.1 (Normalized bar resolution). Let P'_k be a free rank-k G-module. The boundary maps are defined as follows:

$$\partial_k(g_1, \dots, g_k) = g_1(g_2, \dots, g_k) + \sum_{i=1}^{k} (-1)^i (g_1, \dots, g_i g_{i+1}, \dots, g_k) + (-1)^k (g_1, \dots, g_{k-1}) . \tag{5.46}$$

To obtain the normalized bar[3] resolution (in inhomogeneous form), one has to quotient out the submodule of P'_n generated by tuples $(g_1, \dots, g_n)$ where one of the g_i's is the identity. It can be shown that the resulting quotient modules P_n form a $\mathbb{Z}[G]$-free resolution of $\mathbb{Z}$.

To explicitly calculate the cohomology groups $H^k(G; A) = H^k(\mathsf{Hom}_{\mathbb{Z}[G]}(P_\bullet, A))$, it is often easier to work with a more explicit description of the involved hom-sets. Since P_k is a free $\mathbb{Z}[G]$-module on G^k, it is isomorphic (as a module) to $\mathbb{Z}[G^{k+1}]$. This can be seen as follows. The generating set consists of all k-tuples of elements in G:

$$S = \{(g_1, \dots, g_k) \mid \forall i \leq k : g_i \in G\} .$$

Since the module is free over $\mathbb{Z}[G]$, one can write every element as a formal linear combination of elements of the form

$$g_0(g_1, \dots, g_k) .$$

One can now construct a morphism φ between this module and $\mathbb{Z}[G^{k+1}]$, which carries the diagonal G-action, in the following way. On the generating set S, define φ as follows:

$$\varphi(g_1, \dots, g_k) := (e, g_1, g_1 g_2, \dots, g_1 g_2 \cdots g_k) . \tag{5.47}$$

It is not hard to show that this morphism is in fact an isomorphism (of G-modules) and that

$$H^k(G; A) = H^k(\mathsf{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}[G^{k+1}], A)) . \tag{5.48}$$

By a little more algebra, it can also be shown that this hom-set is isomorphic to the (set-theoretic) mapping space $\mathsf{Map}(G^k, A)$. This space can be given an Abelian group structure induced by the group structure on A. Combining these facts, one gets the following construction for the cohomology of groups.

[3] One of the possible explanations for this name is that the formal generating elements are often written as $[g_1|g_2|\dots|g_k]$.

Definition 5.6.2 (Group cohomology). Consider a group G together with a G-module A. (For simplicity only discrete, e.g. finite, groups and Abelian coefficients will be considered.) For every $k \in \mathbb{N}$, define the kth **chain group** as

$$C^k(G;A) := \{\text{all set-theoretic functions from } G^k \text{ to } A\}. \tag{5.49}$$

The coboundary map $\mathsf{d}^k : C^k(G;A) \to C^{k+1}(G;A)$ is defined as follows:

$$\begin{aligned}\mathsf{d}^k f(g_1,\dots,g_k,g_{k+1}) := g_1 \cdot f(g_2,\dots,g_k,g_{k+1}) + (-1)^{k+1} f(g_1,\dots,g_k) \\ + \sum_{i=1}^{k} (-1)^{i+1} f(g_1,\dots,g_i g_{i+1},\dots,g_k,g_{k+1}).\end{aligned} \tag{5.50}$$

The cohomology groups are defined as the following quotient groups:

$$H^k(G;A) := \frac{\ker(\mathsf{d}^k)}{\operatorname{im}(\mathsf{d}^k)}. \tag{5.51}$$

Property 5.6.3 (Finiteness). Let G be a finite group and let A be a G-module such that the underlying group is finitely generated. Since in this case the hom-groups are themselves finitely generated, the cohomology groups $H^k(G;A)$ for $k \geq 1$ are also finitely generated. Furthermore, they are annihilated by the order of G, so in particular they are all torsion. It follows that all cohomology groups are finite.

Property 5.6.4 (Bockstein homomorphism). Let G be a group and consider a short exact sequence of $\mathbb{Z}[G]$-modules

$$0 \longrightarrow A \longrightarrow B \longrightarrow C \longrightarrow 0.$$

This exact sequence induces a long exact sequence in group cohomology. The connecting homomorphism

$$H^\bullet(G;C) \to H^{\bullet+1}(G;A) \tag{5.52}$$

is called the Bockstein homomorphism.

@@ COMPLETE (ADD e.g. classification of extensions) @@

5.6.2 Hochschild and cyclic homology ♣

Definition 5.6.5 (Hochschild homology groups). Let A be an associative algebra and consider an A-bimodule M. The Hochschild chain complex is obtained by tensoring the bar complex of A by M:

$$HC_k(A;M) := M \otimes A^{\otimes k} \cong M \otimes_{A^e} B_n(A), \tag{5.53}$$

where $B_\bullet(A)$ is the bar complex (Definition 5.4.13). The differential d on the bar complex $B_\bullet(A)$ can be extended to $M \otimes_{A^e} B_\bullet(A)$ as $\mathbb{1}_M \otimes \mathsf{d}$ and the above isomorphism then

induces a differential on $M \otimes A^k$. The boundary operator $\mathsf{d}_k : HC_k(A;M) \to HC_{k-1}(A;M)$ is then given by:

$$\mathsf{d}_k : m \otimes a_1 \otimes \cdots \otimes a_k \mapsto ma_1 \otimes \cdots \otimes a_k + \sum_{i=1}^{k-1} (-1)^i m \otimes a_1 \otimes \cdots \otimes a_i a_{i+1} \otimes \cdots \otimes a_k + (-1)^k a_k m \otimes \cdots \otimes a_{k-1} . \tag{5.54}$$

The Hochschild homology of A with coefficients in M is then given by the homology of this chain complex:

$$HH_k(A;M) := \mathsf{ker}(\mathsf{d}_k)/\,\mathsf{im}(\mathsf{d}_{k+1}) . \tag{5.55}$$

This complex can be turned into a graded-commutative algebra by equipping it with the *shuffle product*.

Remark 5.6.6 (Cohomology). The Hochschild cochain complex is obtained by applying the Hom-functor to the bar complex:

$$HC^k(A;M) := \mathsf{Hom}_{A_e}\big(B_k(A);M\big) , \tag{5.56}$$

where $\mathsf{Hom}_{A_e}\big(B_k(A);M\big) \cong \mathsf{Hom}(A^{\otimes k};M)$ by sending $f \in \mathsf{Hom}(A^{\otimes k};M)$ to

$$\overline{f} : a_0 \otimes \cdots \otimes a_{k+1} \mapsto a_0 f(a_1 \otimes \cdots \otimes a_k) a_{k+1} . \tag{5.57}$$

The cohomology of this cochain complex is called the Hochschild cohomology $HH^\bullet(A;M)$ of A with coefficients in M.

Definition 5.6.7 (Cyclic homology). The cyclic chain complex $CC_\bullet(A)$ is the subcomplex of the Hochschild complex $HC_\bullet(A;A)$ by quotienting out the coinvariants of the permutation operator

$$\lambda : HC_k(A) \to HC_k(A) : a_1 \otimes \cdots \otimes a_k \mapsto (-1)^k a_k \otimes a_0 \otimes \cdots \otimes a_{k-1} . \tag{5.58}$$

This gives the **Connes complex**

$$CH_\bullet(A) := HH_\bullet(A;A)/\langle 1-\lambda \rangle . \tag{5.59}$$

It can be shown that the Hochschild boundary operator commutes with the permutation operator and, hence, descends to the cyclic complex. The resulting homology is called cyclic homology $CH_\bullet(A)$. As usual, by dualizing, one obtains cyclic cohomology.

Theorem 5.6.8 (Hochschild–Kostant–Rosenberg). *Let $A = C^\infty(M)$ for M a smooth manifold (see Definition 31.1.4).*

$$HH_\bullet(A) \cong \Omega^\bullet(M) . \tag{5.60}$$

Example 5.6.9 (Smooth algebras). Consider the case of $A = C^\infty(M)$, where M is a smooth manifold.

$$CH_k(A) \cong \Omega^k_{\mathrm{dR}}(M)/\mathsf{d}\Omega^{k-1}_{\mathrm{dR}}(M) \oplus \bigoplus_{\substack{i=1\\ k-2i\geq 0}} H^{k-2i}_{\mathrm{dR}}(M) . \tag{5.61}$$

For this reason, cyclic cohomology is used in *noncommutative geometry* to generalize *de Rham cohomology*.

Part II

Topology

Table of Contents

6 General Topology **172**

6.1 Topological spaces . . . 172

6.1.1 Neighbourhoods . . . 174

6.1.2 Separation axioms . . . 175

6.1.3 Convergence . . . 176

6.2 Morphisms . . . 176

6.2.1 Continuity . . . 176

6.2.2 Homeomorphisms . . . 178

6.3 Associated constructions . . . 179

6.4 Connected spaces . . . 181

6.5 Compact spaces . . . 182

6.5.1 Compactness . . . 182

6.5.2 Compactifications . . . 184

6.6 Uniform spaces . . . 184

6.7 Bornological spaces . . . 185

6.8 Locales ♣ . . . 185

6.8.1 Frames . . . 185

6.8.2 Stone spaces . . . 186

7 Algebraic Topology **188**

7.1 Homotopy theory . . . 189

7.1.1 Homotopy . . . 189

7.1.2 Homotopy groups . . . 190

7.1.3 CW complexes . . . 195

7.1.4 Fibrations . . . 196

7.1.5 Rational homotopy theory ♣ . . . 198

7.1.6 Equivariant homotopy theory ♣ . . . 200

7.2 Simplicial homology . . . 201

7.2.1 Simplices . . . 201

7.2.2 Simplicial homology . . . 202

7.2.3 Relative homology . . . 205

7.2.4 Examples . . . 206
7.3 Singular homology . . . 207
7.4 Local coefficients . . . 210
7.5 Axiomatic approach ♣ . . . 210
7.6 Equivariant cohomology ♣ . . . 212
7.6.1 Borel construction . . . 213
7.6.2 Classifying topoi and stacks . . . 213

8 Sheaf Theory **215**
8.1 Presheaves . . . 215
8.2 Sheaves . . . 216
8.3 Equivariance . . . 219
8.4 Abelian sheaf cohomology . . . 220
8.4.1 Derived cohomology . . . 220
8.4.2 Čech cohomology . . . 221
8.5 Non-Abelian sheaf cohomology . . . 222
8.5.1 Čech cohomology . . . 222
8.6 Ringed spaces . . . 223

9 Metric Spaces **225**
9.1 Definition . . . 225
9.2 Topology . . . 226
9.3 Examples . . . 228
9.4 Completeness . . . 229
9.4.1 Convex metric spaces . . . 229
9.5 Compactness . . . 229

10 Algebraic Geometry **231**
10.1 Varieties . . . 231
10.1.1 Topology . . . 234
10.1.2 Varieties . . . 236
10.1.3 Projective varieties . . . 236
10.2 Schemes . . . 239
10.2.1 Morphisms . . . 242
10.2.2 Tangent space . . . 244
10.3 Sheaf theory ♣ . . . 245
10.4 Algebraic groups & Galois theory . . . 248
10.5 Tropical geometry . . . 248

11 Number Theory **250**
11.1 Adic numbers . . . 250

11.1.1 Finite fields 250
11.1.2 Rational numbers 251
11.1.3 Number and global fields 252
11.2 Algebraic geometry ♣ 254
11.2.1 Modular forms 254
11.2.2 Algebraic functions 256
11.3 L-functions 256

Chapter 6

General Topology

Chapter Contents

6.1 Topological spaces **172**
6.1.1 Neighbourhoods 174
6.1.2 Separation axioms 175
6.1.3 Convergence 176
6.2 Morphisms **176**
6.2.1 Continuity 176
6.2.2 Homeomorphisms 178
6.3 Associated constructions **179**
6.4 Connected spaces **181**
6.5 Compact spaces **182**
6.5.1 Compactness 182
6.5.2 Compactifications 184
6.6 Uniform spaces **184**
6.7 Bornological spaces **185**
6.8 Locales ♣ **185**
6.8.1 Frames 185
6.8.2 Stone spaces 186

6.1 Topological spaces

Definition 6.1.1 (Topology). Let X be a set and consider a collection of subsets $\tau \subseteq 2^X$. The set τ is a topology on X if it satisfies the following axioms:

1. $\emptyset \in \tau$ and $X \in \tau$,
2. $\forall \mathcal{F} \subseteq \tau : \bigcup_{V \in \mathcal{F}} V \in \tau$, and
3. $\forall\, U, V \in \tau : U \cap V \in \tau$.

The elements of τ are called **open sets** and the couple (X, τ) is called a **topological space**. **Closed sets** are defined as the sets that have an open complement. Because complements are uniquely defined, one could just as well define a topology in terms of closed subsets.

Property 6.1.2 (Category of opens ♣). Consider a topological space (X, τ) and let $U \subseteq V \in \tau$. The topology τ together with the collection of inclusion maps $U \hookrightarrow V$ forms a poset and, by extension, a small category $\mathbf{Open}(X)$.

Definition 6.1.3 (Pointed topological space). A topological space X with a designated **base point** $x_0 \in X$. Equivalently, a pointed object (Definition 3.4.17) in the category **Top** of topological spaces.

Example 6.1.4 (Relative topology[1]). Any subset Y of a topological space (X, τ) can be turned into a topological space by equipping it with the following topology:

$$\tau_{\text{rel}} := \{U_i \cap Y \mid U_i \in \tau\}\,. \tag{6.1}$$

Example 6.1.5 (Discrete topology). The topology in which every subset is open (and thus also closed).

Example 6.1.6 (Indiscrete topology). The topology in which only the empty set and the space itself are open.

Definition 6.1.7 (Interior). The interior Y° of a subset Y of a topological space X is defined as the union of all open subsets of Y. Elements of the interior are called **interior points** of Y.

Definition 6.1.8 (Closure). The closure $\overline{Y}$ of a subset Y of a topological space X is defined as the intersection of all closed sets containing Y.

Definition 6.1.9 (Boundary). The boundary ∂Y of a subset Y of a topological space X is defined as $\overline{Y} \setminus Y^\circ$.

Definition 6.1.10 (Borel algebra). The σ-algebra 1.6.12 generated by all open subsets of a topological space. The elements of this σ-algebra are called Borel sets.

Property 6.1.11. The closed subsets of a topological space also generate the Borel σ-algebra.

Definition 6.1.12 (Topological group). A group equipped with a topology such that both the multiplication and inversion morphisms are continuous.

[1] Sometimes called the **subspace** or **trace topology**.

Definition 6.1.13 (G_δ-set). A countable intersection of open sets.

Definition 6.1.14 (F_σ-set). A countable union of closed sets.

6.1.1 Neighbourhoods

Definition 6.1.15 (Neighbourhood). A set $N \subseteq X$ is a neighbourhood of a point $x \in X$ if there exists an open set U such that $x \in U \subseteq N$.

Although the following two notions are often treated as synonyms in the literature, they can be given a separate meaning.

Definition 6.1.16 (Limit point). Let Y be a subset of X. A point $x \in X$ is a limit point of Y if every neighbourhood of x contains at least one point of Y different from x.

Definition 6.1.17 (Adherent point). Let Y be a subset of X. A point $x \in X$ is an adherent point of Y if every neighbourhood of x contains at least one point of Y. A point x is an adherent point of Y if and only if it is an element of the closure $\overline{Y}$.

Definition 6.1.18 (Accumulation point[2]). Let $x \in X$ be a limit point of Y. It is called an accumulation point of Y if every open neighbourhood of x contains infinitely many points of Y.

Definition 6.1.19 (Basis). A collection $\mathcal{B} \subseteq \tau$ of open subsets of a topological space (X, τ) is called a basis for τ if every $U \in \tau$ can be written as

$$U = \bigcup_{V \in \mathcal{F}} V, \tag{6.2}$$

for some $\mathcal{F} \subseteq \mathcal{B}$.

Definition 6.1.20 (Local basis). A collection $\mathcal{B}_x$ of open neighbourhoods of a point $x \in X$ is a local basis of x if every neighbourhood of x contains at least one element in $\mathcal{B}_x$.

Definition 6.1.21 (First-countable space). A topological space (X, τ) for which, for every point $x \in X$, there exists a countable local basis.

Property 6.1.22 (Decreasing basis). Let $x \in X$. If there exists a countable local basis for x, there also exists a countable decreasing local basis for x.

Definition 6.1.23 (Second-countable space). A topological space (X, τ) for which there exists a countable (global) basis.

Property 6.1.24 (Closure). Let X be a topological space. The closure of a subset $Y \subseteq X$ is given by

$$\overline{Y} = \{x \in X \mid \exists \text{ a net } (x_\alpha)_{\alpha \in I} \text{ in } X : x_\alpha \longrightarrow x\}. \tag{6.3}$$

[2]Sometimes called a **cluster point**.

This implies that the topology on X is completely determined by the convergence of nets (Definition 1.7.16).

Definition 6.1.25 (Fréchet–Urysohn space). A topological space for which the closure of every subset is equal to its sequential closure, i.e. the subset obtained as in Property 6.1.24, but with nets replaced by sequences.

Fréchet–Urysohn spaces form an important subclass of **sequential spaces**, i.e. topological spaces where the topology is uniquely determined by the convergence of sequences: a subset of a sequential space is closed if and only if every convergent sequence converges to a point in the set.

The following property is of great practical importance.

Property 6.1.26. Every first-countable space is Fréchet–Urysohn and, therefore, only convergent sequences have to be considered in these spaces.

Definition 6.1.27 (Germ). Let X be a topological space and let Y be a set. Consider two functions $f, g : X \to Y$. If there exists a neighbourhood N of a point $x \in X$ such that

$$f(u) = g(u) \qquad \forall u \in N\,, \tag{6.4}$$

then this property defines an equivalence relation denoted by $f \sim_x g$ and the equivalence classes are called germs at x.

6.1.2 Separation axioms

Definition 6.1.28 (Irreducible space). A topological space that is not the union of two proper closed subsets or, equivalently, if the intersection of two nonempty open subsets is again nonempty.

Definition 6.1.29 (T_0-space). A topological space such that for every two distinct points at least one of them has a neighbourhood not containing the other, i.e. they are **topologically distinguishable**. T_0-spaces are also said to carry a **Kolmogorov topology**.

Definition 6.1.30 (T_1-space). A topological space such that for every two distinct points x, y there exists neighbourhood N, N' of x and y respectively such that $y \notin N$ and $x \notin N'$, i.e. they can topologically **separated**. T_1-spaces are also said to carry a **Fréchet topology** (not to be confused with *Fréchet spaces* from functional analysis).

Definition 6.1.31 (Hausdorff space). A topological space X is called a Hausdorff space or T_2-space if it satisfies the following condition:

$$\forall x, y \in X : \exists \text{ neighbourhoods } N \ni x, N' \ni y : N \cap N' = \emptyset\,. \tag{6.5}$$

The points are said to be **separated by neighbourhoods**. It can be shown that this

definition is equivalent to requiring that the diagonal Δ_X is closed in the product space $X \times X$.

Property 6.1.32 (Closed points). Every singleton and, by extension, every finite subset of a Hausdorff space is closed.

Definition 6.1.33 (Urysohn space). A topological space is called an Urysohn space or $T_{2^1/2}$-space if every two distinct points can be separated by closed neighbourhoods.

Definition 6.1.34 (Regular space). A topological space such that for every closed subset V and every point $x \notin V$ there exist disjoint open subsets U, U' such that $x \in U$ and $V \subset U'$.

Definition 6.1.35 (T_3-space). A space that is both regular and T_0.

Definition 6.1.36 (Normal space). A topological space such that every two closed subsets have disjoint neighbourhoods.

Definition 6.1.37 (T_4-space). A space that is both normal and T_1.

Property 6.1.38 (Nesting of axioms). A space satisfying the separation axiom T_k also satisfies all separation axioms $T_{i \leq k}$.

6.1.3 Convergence

Definition 6.1.39 (Convergence). A sequence $(x_n)_{n \in \mathbb{N}}$ in X is said to (**sequentially**) converge to a point $x \in X$ if

$$\forall \text{ neighbourhoods } U \text{ of } x : \exists N \in \mathbb{N}_0 : \forall n > N : x_n \in U\,. \tag{6.6}$$

Property 6.1.40 (Uniqueness). The limit of a convergent sequence in a Hausdorff space is unique.

Property 6.1.41 (Subsequences). Every subsequence of a convergent sequence converges as well. Moreover, in a Hausdorff space, the limits coincide.

6.2 Morphisms

6.2.1 Continuity

Definition 6.2.1 (Continuous function). A function between topological spaces for which the preimage of every open set is also open.

Notation 6.2.2. The set of all continuous functions between two topological spaces X, Y is often denoted by $C(X, Y)$.

Definition 6.2.3 (Initial topology). Consider a collection of functions $\{f_i : X \to Y_i\}_{i \in I}$ between topological spaces. The initial (or **limit**) topology on X, with respect to this family, is the coarsest topology on X for which all maps f_i are continuous.

Definition 6.2.4 (Final topology). Consider a collection of functions $\{f_i : Y_i \to X\}_{i\in I}$ between topological spaces. The final (or **inductive**[3]) topology on X, with respect to this family, is the finest topology on X for which all maps f_i are continuous.

Property 6.2.5 (Sequential continuity). Consider a function $f : X \to Y$ of topological spaces, where X is first-countable. The following statements are equivalent:

- f is continuous.
- The sequence $(f(x_n))_{n\in\mathbb{N}}$ converges to $f(a) \in Y$ whenever the sequence $(x_n)_{n\in\mathbb{N}}$ converges to $a \in X$.

Remark 6.2.6. If the space Y in the previous theorem is Hausdorff, the limit is irrelevant since it is unique by Property 6.1.40 above.

Remark 6.2.7. If the space X is not first-countable, one has to consider the convergence of nets (Definition 1.7.16).

Theorem 6.2.8 (Urysohn's lemma). *A topological space X is normal if and only if every two closed disjoint subsets $A, B \subset X$ can be separated by a continuous function $f : X \to [0,1]$, i.e. $\forall a \in A, b \in B$ there exists a continuous function $f : X \to [0,1]$ such that*

$$f(a) = 0 \qquad \text{and} \qquad f(b) = 1\,. \tag{6.7}$$

The following, seemingly unrelated, theorem is actually equivalent to Urysohn's lemma.

Theorem 6.2.9 (Tietze extension theorem). *Consider a continuous function $f : V \to \mathbb{R}$, where V is a closed subset of a normal space X. There exists a continuous function $F : X \to \mathbb{R}$ such that $\forall x \in V : F(x) = f(x)$. Furthermore, if the function f is bounded, then F can be chosen to be bounded by the same number.*

Definition 6.2.10 (Limit superior and inferior). Consider a continuous function $f : X \to \mathbb{R}$. The limit inferior and superior of f at $x_0 \in (X, \tau)$ are defined as follows:

$$\liminf_{x\to x_0} f(x) := \sup\big\{\inf\{f(x) \mid x \in U\backslash\{x_0\}\} \mid U \in \tau, x_0 \in U, U\backslash\{x_0\} \neq \emptyset\}\big\}\,, \tag{6.8}$$

$$\limsup_{x\to x_0} f(x) := \inf\big\{\sup\{f(x) \mid x \in U\backslash\{x_0\}\} \mid U \in \tau, x_0 \in U, U\backslash\{x_0\} \neq \emptyset\}\big\}\,. \tag{6.9}$$

For functions on a subspace $Y \subset X$ one should look at the values $\{f(x) \mid x \in Y \cap U\backslash\{x_0\}\}$. For the definition to make sense, x_0 should be a limit point of Y.

Definition 6.2.11 (Semicontinuity). Consider a function $f : X \to \overline{\mathbb{R}}$ on a topological space. It is said to be **upper semicontinuous** at a point $x_0 \in X$ if for all $y \in \mathbb{R}$ with $y > f(x_0)$ there exists a neighbourhood of x_0 such that $f(x) < y$ for all $x \in U$. Analogously, the function is said to be **lower semicontinuous** at a point $x_0 \in X$ if for all $y \in \mathbb{R}$ with $y < f(x_0)$ there exists a neighbourhood of x_0 such that $f(x) > y$ for all $x \in U$.

[3] The reason for this name is that the direct (or inductive) limit (Definition 1.8.2) of topological spaces is usually endowed with this topology.

These definitions can be expressed more succintly as follows:

- Upper semicontinuous: $\limsup_{x \to x_0} f(x) \leq f(x_0)$ for all $x_0 \in X$, and
- Lower semicontinuous: $\liminf_{x \to x_0} f(x) \geq f(x_0)$ for all $x_0 \in X$.

6.2.2 Homeomorphisms

Definition 6.2.12 (Homeomorphism). A bijection f such that both f and f^{-1} are continuous.

Definition 6.2.13 (Embedding). A continuous function that is a homeomorphism onto its image.

Definition 6.2.14 (Local homeomorphism). A continuous function $f : X \to Y$ such that every point $x \in X$ has an open neighbourhood U for which $f|_U$ is an embedding. Local homeomorphisms are also called **étale morphisms**.

Definition 6.2.15 (Covering space). Consider two topological spaces C, X and a continuous surjection $p : C \to X$. C is said to be a covering space of X (and p is called a **covering map**) if, for all points $x \in X$, there exists an open neighbourhood U of x such that $p^{-1}(U)$ can be written as a disjoint union of open sets in C, where every set is mapped homeomorphically onto U. The neighbourhoods U are sometimes said to be **evenly covered**.

Notation 6.2.16. Because the covering map $p : C \to M$ is surjective, the space M can be left implicit. Therefore, covering spaces are often just denoted by the couple (C, p).

Definition 6.2.17 (Covering transformation). Consider two covering spaces (C, p) and (C', p'). A continuous function $f : C \to C'$ is called a covering transformation if $p' \circ f = p$.

Definition 6.2.18 (Deck transformation). Let $p : C \to X$ be a covering map. The automorphism group of (C, p) in the category of covering spaces (over X) is given by all homeomorphisms φ satisfying $p \circ \varphi = p$. These automorphisms are called deck transformations.

Definition 6.2.19 (Étale space). Let X be a topological space. A topological space Y is called an étale space over X if there exists a surjective local homeomorphism $\pi : Y \to X$. The preimage $\pi^{-1}(x)$ of a point $x \in X$ is called the **stalk** of π over x.

Example 6.2.20. Every covering space is an étale space.

Definition 6.2.21 (Pseudogroup ♣). Let (X, τ) be a topological space. A pseudogroup is a collection $\mathcal{G}$ of homeomorphisms $\phi : U \to V$ between open subsets of X such that:

1. $\mathbb{1}_U \in \mathcal{G}$ for all $U \in \tau$.
2. If $\phi \in \mathcal{G}$, then $\phi^{-1} \in \mathcal{G}$.

3. If $V \subset U \in \tau$, then $\phi|_V \in \mathcal{G}$.
4. If $U = \bigcup_{i \in I} U_i$ and $\phi|_{U_i} : U_i \to V$ is an element of $\mathcal{G}$ for all $i \in I$, then $\phi \in \mathcal{G}$.
5. If $\phi : U \to V$ and $\psi : U' \to V'$ are elements of $\mathcal{G}$ and $V \cap U' \neq \emptyset$, then $\psi \circ \phi|_{\phi^{-1}(V \cap U')} \in \mathcal{G}$.

6.3 Associated constructions

Construction 6.3.1 (Product topology). First, consider the case with only a finite number of spaces $\{(X_i, \tau_i)\}_{i \in I}$. The Cartesian product $X := \prod_{i \in I} X_i$ can be turned into a topological space by equipping it with the topology generated by the following basis:

$$\mathcal{B} := \left\{ \prod_{i \in I} U_i \middle| U_i \in \tau_i \right\}. \tag{6.10}$$

In the general case the topology can be defined using the canonical projections $\pi_i : X \to X_i$. The general product topology, called the **Tychonoff topology**, is the initial topology with respect to the projections π_i.

Construction 6.3.2 (Disjoint union). Let $\{(X_i, \tau_i)\}_{i \in I}$ be a family of topological spaces and consider the disjoint union

$$X := \bigsqcup_{i \in I} X_i \tag{6.11}$$

together with the canonical inclusion maps $\phi_i : X_i \to X : x_i \mapsto (i, x_i)$. The set X can be turned into a topological space by equipping it with the following topology:

$$\tau_X := \left\{ U \subseteq X \mid \forall i \in I : \phi_i^{-1}(U) \in \tau_i \right\}. \tag{6.12}$$

Construction 6.3.3 (Quotient space). Consider a topological space (X, τ) and a subset $Y \subseteq X$. The quotient X/Y is defined as the set $X \backslash Y \sqcup \{*\}$ where the point $*$ can be regarded as the result of identifying all points in Y. This canonically turns the quotient space into a pointed space.

Let π be the canonical projection $X \to X/Y$. The quotient space can be turned into a topological space by equipping it with the following topology:

$$\tau_Q := \left\{ U \subseteq X/Y \mid \pi^{-1}(U) \in \tau \right\}. \tag{6.13}$$

Remark 6.3.4 (Degenerate quotient). For the degenerate case $Y = \emptyset$ one can also apply the above definition. However, this has the awkward effect that it adjoins a new point to the space X instead of a collapsing it:

$$X/\emptyset = X \sqcup *. \tag{6.14}$$

Construction 6.3.5 (Wedge sum). Consider two pointed spaces (X, x_0) and (Y, y_0). The wedge sum $X \vee Y$ is defined as the quotient of the disjoint union $X \sqcup Y$ obtained by identifying the basepoints $x_0 \sim y_0$.

Definition 6.3.6 (Smash product). Consider two pointed topological spaces (X, x_0) and (Y, y_0). The smash product $X \wedge Y$ is defined as the quotient

$$X \wedge Y := (X \times Y)/(X \vee Y), \tag{6.15}$$

where $X \vee Y$ sits inside the product as the union of $X \times \{y_0\}$ and $\{x_0\} \times Y$.

Construction 6.3.7 (Suspension). Let X be a topological space. The suspension of X is defined as the following quotient space:

$$SX := (X \times [0,1])/\big\{(x,0) \sim (y,0) \text{ and } (x,1) \sim (y,1) \mid x, y \in X\big\}. \tag{6.16}$$

By the remark about degenerate quotients the suspension of the empty set is in fact not empty, but equal to the two-point space S^0.

An often more interesting construction is the **reduced suspension** ΣX. This is obtained by taking the ordinary suspension SX of a pointed space (X, x_0) and identifying all copies of x_0:

$$\Sigma X := SX/(x_0 \times [0,1]). \tag{6.17}$$

An equivalent definition of the reduced suspension can be given in terms of the smash product:

$$\Sigma X \cong X \wedge S^1. \tag{6.18}$$

Example 6.3.8 (Spheres). Up to homeomorphism, the spheres are related by (reduced) suspensions:

$$SS^n \cong S^{n+1} \cong \Sigma S^n. \tag{6.19}$$

If one identifies the empty set with the (-1)-sphere, this relation can be continued to the case $n = -1$.

Construction 6.3.9 (Attaching space). Let X, Y be two topological spaces and consider a subspace $A \subseteq X$. For every continuous function $f : A \to Y$, called the **attaching map**, one can construct the attaching space or **adjunction space** $X \cup_f Y$ in the following way:

$$X \cup_f Y := (X \sqcup Y)/\{A \sim f(A)\}. \tag{6.20}$$

In categorical terms, it is the pushout (Definition 3.4.70) in **Top** of the inclusion $\iota : A \hookrightarrow X$ along $f : A \to Y$.

Construction 6.3.10 (Join). Let $\{X_i\}_{i\leq n}$ be a finite collection of topological spaces. The join $X := X_1 \circ \cdots \circ X_n$ is defined as follows. Every point of X is defined by the following data:

1. An element of the standard n-simplex (Definition 7.2.1), i.e. an n-tuple of non-negative numbers $\{t_i\}_{i\leq n}$ satisfying $\sum_i t_i = 1$.
2. For each index i such that $t_i \neq 0$, a point $x_i \in X_i$.

This point in X is denoted by $t_1x_1 \oplus \cdots \oplus t_nx_n$.

In the case of two spaces, there exists a more intuitive construction. Let X, Y be two topological spaces. The join $X \circ Y$ is equal to the quotient space $(X \times Y \times [0,1])/\sim$, where the equivalence relation is defined as follows:

- For all $x \in X$ and $y, y' \in Y$: $(x, y, 0) \sim (x, y', 0)$.
- For all $x, x' \in X$ and $y \in Y$: $(x, y, 1) \sim (x', y, 1)$.

This can be interpreted as collapsing one end of the cylinder $(X \times Y) \times [0,1]$ to X and the other end to Y.

Property 6.3.11 (Monoidal structure ♣). The join induces a monoidal structure on the category **Top**, where the tensor unit is given by the empty space $\emptyset$.

6.4 Connected spaces

Definition 6.4.1 (Connected space). A topological space that cannot be written as the disjoint union of two nonempty open sets. Equivalently, a space is said to be connected if the only clopen sets are the empty set and the space itself.

Property 6.4.2 (Locally constant). Let X be a connected space and let f be a function on X. If f is **locally constant**, i.e. for every $x \in X$ there exists a neighbourhood U on which f is constant, then f is constant on all of X.

Theorem 6.4.3 (Intermediate value theorem). *Let X be a connected space and let $f : X \to \mathbb{R}$ be a continuous function. If $a, b \in f(X)$, then $\forall c \in]a, b[: c \in f(X)$.*

Definition 6.4.4 (Path-connected space[4]). Let X be a topological space. If for every two points $x, y \in X$ there exists a continuous function $\varphi : [0,1] \to X$ (i.e. a **path**) such that $\varphi(0) = x$ and $\varphi(1) = y$, the space is said to be path-connected.

Property 6.4.5 (Path-connected implies connected). Every path-connected space is connected. The converse does not hold. A connected and locally path-connected space is path-connected.

[4] A similar notion is that of **arcwise-connectedness** where the function φ is required to be a homeomorphism.

Remark 6.4.6 (Connected components). (Path-)connectedness defines an equivalence relation on the space X. The equivalence classes are closed in X and form a cover of X. The set of path components of X is often denoted by $\pi_0(X)$.

6.5 Compact spaces

6.5.1 Compactness

Definition 6.5.1 (Sequentially compact space). A topological space in which every sequence has a convergent subsequence (the sequence itself does not have to be convergent).

Property 6.5.2 (First-countable spaces). A first-countable space is sequentially compact if and only if every countable open cover has a finite subcover.

Definition 6.5.3 (Lindelöf space). A space for which every open cover has a countable subcover.

Property 6.5.4. Every second-countable space is a Lindelöf space.

Definition 6.5.5 (Finite intersection property). A collection $\mathcal{F} \subseteq P(X)$ of subsets has the finite intersection property (FIP) if

$$\bigcap_{V \in \mathcal{F}'} V \neq \emptyset \tag{6.21}$$

for all finite $\mathcal{F}' \subset \mathcal{F}$.

Definition 6.5.6 (Locally finite cover). An open cover of a topological space such that every point has a neighbourhood that intersects only finitely many sets in the cover.

Definition 6.5.7 (Paracompact space). A topological space for which every open cover has a locally finite open refinement.

Property 6.5.8. Every paracompact Hausdorff space is normal.

Definition 6.5.9 (Partition of unity). A collection $\{f_i : X \to [0,1]\}_{i \in I}$ of continuous functions such that for every $x \in X$ the following conditions hold:

1. **Locally finite**: For every neighbourhood U of x, the set $\{f_i \mid \mathsf{supp} f_i \cap U \neq \emptyset\}$ is finite.
2. **Normalization**: $\sum_{i \in I} f_i(x) = 1$.

Consider an open cover $\{V_i\}_{i \in I}$ of X. If there exists a partition of unity, also indexed by I, such that $\mathrm{supp}(\varphi_i) \subseteq U_i$, this partition of unity is said to be **subordinate** to the given cover.

Property 6.5.10 (Hausdorff spaces). A paracompact space is Hausdorff if and only if it admits a partition of unity subordinate to any open cover.

Definition 6.5.11 (Numerable open cover). An open cover of a topological space is said to be numerable if the space admits a partition of unity subordinate to the given cover.

Definition 6.5.12 (Compact space). A topological space for which every open cover has a finite subcover.

Theorem 6.5.13 (Heine–Borel[5]). *If a topological space X is sequentially compact and second-countable, every open cover has a finite subcover and, therefore, X is compact.*

Corollary 6.5.14 (Real numbers). A subset of $\mathbb{R}^n$ is compact if and only if it is closed and bounded.

Definition 6.5.15 (Proper function). A continuous function for which the inverse of every compact set is again compact.

Theorem 6.5.16 (Tychonoff's theorem). *Any product of compact topological spaces is compact under the (Tychonoff) product topology (Construction 6.3.1).*

Definition 6.5.17 (Relatively compact space). A topological space for which its closure is compact.

Definition 6.5.18 (Locally compact space). A topological space in which every point has a compact neighbourhood.

Theorem 6.5.19 (Dini). *Let (X, τ) be a compact space and let $(f_n)_{n\in\mathbb{N}}$ be a monotone sequence of continuous functions $f_n : X \to \mathbb{R}$. If $(f_n)_{n\in\mathbb{N}}$ converges pointwise to a continuous function, the convergence is uniform.*

Definition 6.5.20 (ω-bounded space). A topological space in which the closure of every countable subset is compact.

Definition 6.5.21 (Compact-open topology). Consider the mapping space $C(X, Y)$ between two topological spaces. This space is often endowed with a topology generated by the subbasis of subsets of the form

$$U^K := \{f : X \to Y \mid K \text{ compact}, U \text{ open and } f(K) \subseteq U\}. \tag{6.22}$$

Property 6.5.22 (Internal hom). Consider two topological spaces X, Y with X locally compact and equip the mapping space $C(X, Y)$ with the compact-open topology. The following relation is satisfied for all topological spaces Z:

$$C(Z \times X, Y) \cong C\big(Z, C(X, Y)\big), \tag{6.23}$$

i.e. the mapping space $C(X, Y)$ is an internal hom (Definition 4.1.18) in the category **Top** and, because the topological product is the product in **Top**, $C(X, Y)$ is even an exponential object (Definition 4.1.22). For this reason the mapping spaces $C(X, Y)$ are also sometimes denoted by Y^X.

[5] Also called the **Borel–Lebesgue** theorem.

Definition 6.5.23 (Proper action). Consider a topological group G together with an action $\triangleright\colon G \times X \to X$ on a topological space X. The action is said to be proper if $G \times X \to X \times X : (g, x) \mapsto (x, g \triangleright x)$ is proper (Definition 6.5.15), i.e. the subset of elements $g \in G$ such that

$$g \triangleright K \cap K' \neq \emptyset \tag{6.24}$$

for compacta $K, K' \subseteq X$ is itself compact.

6.5.2 Compactifications

Definition 6.5.24 (Dense). A subset $V \subseteq X$ is said to be dense in a topological space X if $\overline{V} = X$.

Definition 6.5.25 (Generic point). A point whose closure is the whole space.

Definition 6.5.26 (Separable space). A topological space that contains a countable, dense subset.

Property 6.5.27. Every second-countable space is separable.

Definition 6.5.28 (Compactification). A compact topological space (X', τ') is a compactification of a topological space (X, τ) if X is a dense subspace of X'.

Example 6.5.29. Standard examples of compactifications are the extended real line $\mathbb{R} \cup \{-\infty, +\infty\}$ and the extended complex plane $\mathbb{C} \cup \{\infty\}$ for the real line and the complex plane, respectively.

Remark. It is important to note that compactifications are not necessarily unique.

Definition 6.5.30 (One-point compactification). Let X be a Hausdorff space. A one-point compactification or **Alexandrov compactification** is a compactification $\widehat{X}$ such that $\widehat{X} \backslash X$ is a singleton.

Example 6.5.31 (Real line). The classic example of a (one-point) compactification is that of the real line. By adjoining the points $\pm\infty$ and identifying them, the circle S^1 is obtained. In general one can obtain the n-dimensional sphere S^n as the one-point compactification of $\mathbb{R}^n$. This can be regarded as an inverse *stereographic projection*.

6.6 Uniform spaces

Definition 6.6.1 (Uniform structure). A uniform structure on a set X consists of a collection $\mathfrak{U}$ of subsets $U \subseteq X \times X$ that satisfy the following properties:

1. **Monotonicity**: If $U \in \mathfrak{U}$ and $U \subset V$, then $V \in \mathfrak{U}$.
2. **Intersections**: If $U, V \in \mathfrak{U}$, then $U \cap V \in \mathfrak{U}$.

3. **Diagonals**: If $U \in \mathfrak{U}$, then $\Delta_X \subset U$.
4. **Halving**: If $U \in \mathfrak{U}$, there exists $V \in \mathfrak{U}$ such that $V \circ V \subseteq U$.
5. **Transposition**: If $U \in \mathfrak{U}$, then $U^t \in \mathfrak{U}$.

U^t denotes the converse (Definition 1.4.7) of U and $\circ$ is the relational composition (Definition 1.4.8) of V and V. The elements of the uniformity $\mathfrak{U}$ are called **entourages**. If $(x, y) \in U$ for some entourage $U \in \mathfrak{U}$, then x and y are said to be U**-close**.

Remark 6.6.2. The first three conditions imply that a uniform structure is in particular a filter (Definition 1.6.8).

@@ COMPLETE (see Bourbaki) @@

6.7 Bornological spaces

Definition 6.7.1 (Bornology). A bornology on a set X is a collection $\mathfrak{b} \subset P(X)$ such that:

1. **Singletons**: $\forall x \in X : \{x\} \in \mathfrak{b}$.
2. **Monotonicity**: $A \in \mathfrak{b}, B \subset A \implies B \in \mathfrak{b}$.
3. **Union**: $A, B \in \mathfrak{b} \implies A \cup B \in \mathfrak{b}$.

The sets in $\mathfrak{b}$ are said to be **bounded** and the pair $(X, \mathfrak{b})$ is called a **bornological space**.

@@ COMPLETE (see e.g. Rejzner) @@

6.8 Locales ♣

6.8.1 Frames

Property 6.8.1 (Opens form a frame). Consider the poset **Open**(X) of opens of a topological space X. This set is closed under finite intersections (limits) and arbitrary unions (colimits). Furthermore, arbitrary unions distribute over finite intersections:

$$V \cap \left(\bigcup_{i \in I} U_i \right) = \bigcup_{i \in I} (V \cap U_i) \,. \tag{6.25}$$

This implies that the poset **Open**(X) is a frame (Definition 1.7.40).

Definition 6.8.2 (Locale). The previous property can be used to generalize the notion of topological spaces to include 'pointless spaces'. Let **Frame** denote the category of

frames together with frame homomorphisms. The category of locales is defined as its opposite category:

$$\mathbf{Loc} := \mathbf{Frame}^{\mathrm{op}} . \tag{6.26}$$

Construction 6.8.3 (From locale to topological space). There exists an adjunction

$$\begin{tikzcd} \mathbf{Loc} \arrow[r, shift right, "\mathrm{Point}"'] \arrow[r, phantom, "\perp"] & \mathbf{Top} \arrow[l, shift right, "\iota"'] \end{tikzcd} , \tag{6.27}$$

where the right adjoint is defined as follows. Let L be a locale. For a topological space, the points are given by continuous functions $* \to X$ and, hence, by frame morphisms $\mathbf{Open}(X) \to 1 \equiv \Omega_{\mathrm{Frame}} = \{0, 1\}$. Generalizing this to locales, one defines the set of points of L as the Ω_{Loc}-elements:

$$\mathrm{Point}(L) := \mathbf{Loc}(1, L) . \tag{6.28}$$

This set can be given a topology by declaring for every $U \in L$ the set $\{p \in \mathrm{Point}(L) \mid p^{-1}(U) = 1\}$ to be open.

Definition 6.8.4 (Sober space). A topological space X for which the map $X \to \mathrm{Point}(X)$ is a homeomorphism, i.e. the points of X are precisely determined by its frame of opens. Equivalently, a topological space such that every irreducible closed subset is the closure of a unique point, i.e. has a unique generic point (Definition 6.5.25). Important examples are Hausdorff spaces.

6.8.2 Stone spaces

Definition 6.8.5 (Stone space). The inverse limit (Definition 1.8.4) of an inverse system of finite, discrete spaces. Equivalently, it is a totally disconnected, compact Hausdorff space (Sections 6.4 and 6.5).

Property 6.8.6 (Stone duality). By *Stone duality*, these spaces are equivalent to pro-objects (Definition 3.4.55) in the category of finite sets, i.e. they are equivalent to **profinite sets**. Moreover, the opposite category of Stone spaces is equivalent to the category of Boolean algebras (Definition 1.7.45):

$$\mathbf{Bool} \cong \mathbf{Pro}(\mathbf{FinSet})^{\mathrm{op}} \cong \mathbf{Stone}^{\mathrm{op}} . \tag{6.29}$$

The isomorphism between the category of Boolean algebras and Stone spaces is the content of **Stone's representation theorem**. Every Boolean algebra B is isomorphic to the algebra of clopen subsets of the Stone space $S(B)$, consisting of the ultrafilters on B (Definition 1.7.18). A topology on $S(B)$ is generated by the sets $\{x \in S(B) \mid b \in x\}$, where b ranges over all elements of B. The space $S(B)$ might also be called the **spectrum** of B.

Definition 6.8.7 (Profinite group). A Stonean topological group or, equivalently, a group object in the category of profinite sets.

Definition 6.8.8 (Profinite completion). Consider a group G. Its profinite completion $\widehat{G}$ is constructed as the inverse limit of the system consisting of the quotient groups G/N, where the N are normal subgroups of finite index (Definition 2.2.15).

Although, by Stone duality, pro-objects in the category of finite sets can be identified with their limit (in the category of topological spaces), this is not the case for infinite sets.

Definition 6.8.9 (Progroup). A pro-object (Definition 3.4.55) in the category of groups. If all morphisms in the underlying system are surjective, the progroup itself is called **surjective**.[6]

Remark 6.8.10. Due to the strong relation between locales and topological spaces, there is also a relation between topological groups and localic groups. Because the functor of points $Point : \mathbf{Loc} \to \mathbf{Top}$ is a right adjoint, it preserves limits and, therefore, group objects. This induces a functor $\mathbf{LocGrp} \to \mathbf{TopGrp}$. The converse is, however, not true in general (although it is, for example, true when restricting to locally compact groups).

Property 6.8.11. Surjective progroups are equivalent to **localic groups**, i.e. group objects in $\mathbf{Loc}$, that are inverse limits of discrete groups (these are also called **prodiscrete**).

Localic groups have a strong connection to *topos theory* (see Chapter 12).

Property 6.8.12 (Galois topos). Every localic group G gives rise to a classifying topos $\mathbf{B}G$. Although the functor $\mathbf{LocGrp} \to \mathbf{Topos}$ is not an embedding, it becomes one when restricting to the subcategory of prodiscrete localic groups. The topoi arising in this way are also called **Galois topoi**.

[6]Surjective progroups are exactly those satisfying the *Mittag-Leffler condition*.

Chapter 7

Algebraic Topology

References for this chapter are Massey (1991); Merry (2018).

Chapter Contents

7.1 Homotopy theory **189**
7.1.1 Homotopy 189
7.1.2 Homotopy groups 190
7.1.3 CW complexes 195
7.1.4 Fibrations 196
7.1.5 Rational homotopy theory ♣ 198
7.1.6 Equivariant homotopy theory ♣ 200
7.2 Simplicial homology **201**
7.2.1 Simplices 201
7.2.2 Simplicial homology 202
7.2.3 Relative homology 205
7.2.4 Examples 206
7.3 Singular homology **207**
7.4 Local coefficients **210**
7.5 Axiomatic approach ♣ **210**
7.6 Equivariant cohomology ♣ **212**
7.6.1 Borel construction 213
7.6.2 Classifying topoi and stacks 213

7.1 Homotopy theory

7.1.1 Homotopy

Definition 7.1.1 (Retraction). Let X be a topological space and let $A \subseteq X$ be a subspace. A continuous function $f : X \to A$ is called a retraction (and A is called a **retract** of X) if[1]

$$f|_A = \mathbb{1}_A \,. \tag{7.1}$$

Definition 7.1.2 (Homotopy). Two continuous functions $f, g : X \to Y$ are said to be homotopic if there exists a continuous function $H : X \times [0,1] \to Y$ such that $f(x) = H(x,0)$ and $g(x) = H(x,1)$. This relation induces an equivalence relation on $C(X,Y)$ for which the quotient space is denoted by $[X,Y]$. A homotopy H such that $H(\cdot,t)$ is a homeomorphism for all $t \in [0,1]$ is called an **isotopy**.

Definition 7.1.3 (Deformation retraction). Let X be a topological space and let $A \subseteq X$ be a subspace. A is called a deformation retract if there exists a homotopy between the identity function on X and a retraction $f : X \to A$.

Definition 7.1.4 (Homotopy type). Two topological spaces X and Y are said to be **homotopy equivalent** or to be of the same homotopy type[2] if there exist continuous functions $f : X \to Y$ and $g : Y \to X$ such that $f \circ g$ is homotopic to $\mathbb{1}_Y$ and $g \circ f$ is homotopic to $\mathbb{1}_X$. The maps f, g are called **homotopy equivalences**.

Property 7.1.5 (Homeomorphisms). Every homeomorphism is a homotopy equivalence.

Definition 7.1.6 (Null-homotopic). A continuous function is said to be null-homotopic if it is homotopic to a constant function.

Definition 7.1.7 (Contractible space). A topological space for which the identity map is null-homotopic or, equivalently, a space that is homotopy-equivalent to a point.

Definition 7.1.8 (Good cover). Let X be a topological space with an open cover $\mathcal{U} \equiv \{U_i\}_{i \in I}$. The cover $\mathcal{U}$ is called a good cover (or **nice cover**) if every nonempty finite intersection $U_{i_1} \cap \ldots \cap U_{i_k}$ is contractible.

Property 7.1.9 (Path space). Consider a topological space X. By Property 6.5.22, functions $Y \to X^{[0,1]}$ to the path space represent functions $Y \times [0,1] \to X$, i.e. the path space represents homotopies to X.

Definition 7.1.10 (Mapping class group). Consider a topological space X with some additional structure denoted by $\mathfrak{X}$ (often $\mathfrak{X}$ will be trivial). The mapping class group

[1] It is a retraction of the inclusion map $A \hookrightarrow X$ in the sense of Definition 3.4.1.

[2] The associated equivalence classes are sometimes called **strong homotopy types** to distinguish them from the homotopy types associated to the *weak equivalences* introduced further on.

of X is defined as the quotient

$$\mathsf{MCG}(X, \mathfrak{X}) := \mathsf{Aut}(X, \mathfrak{X}) \,/\, \mathsf{Aut}_0(X, \mathfrak{X}), \tag{7.2}$$

i.e. as the set of path components of its automorphism group (path are, in this case, given by isotopies).

7.1.2 Homotopy groups

Definition 7.1.11 (Loop space). The set of all **loops** in a pointed topological space, i.e. all continuous functions $\gamma : (S^1, t_0) \to (X, x_0)$ for which $\gamma(t_0) = x_0$. This space is denoted by $\Omega_{x_0} X$ (if the base point is clear, the subscript is often omitted). It can be turned into a topological space by equipping it with the compact-open topology (Definition 6.5.21).

When one drops the requirement of based loops, i.e. when one considers the space of all continuous functions $S^1 \to X$, the resulting space is called the **free loop space** on X. This space is denoted by LX.

Definition 7.1.12 (Loop group). In the case of topological groups, one can define a group structure on the (free) loop space. With the compact-open topology, it becomes a topological group.

Remark 7.1.13 (H-structure ♣). Loop spaces can be equipped with a multiplication corresponding to the concatenation of loops[3]. However, this operation is not strictly associative and, hence, it does not endow the loop space with a group structure. Instead, it turns the loop space into an *H-group*, which is in particular an A_∞-space (Definition 3.7.7), i.e. a group up to homotopy.

Definition 7.1.14 (Fundamental group). The fundamental group $\pi_1(X, x_0)$ is defined as the loop space modulo homotopies, i.e. $\pi_1(X, x_0) := \pi_0(\Omega_{x_0} X)$, where π_0 denotes the set of path components (Remark 6.4.6). As the name implies, the fundamental group can be given the structure of a multiplicative group where the operation is inherited from that of the loop space.

Remark 7.1.15. In general, as the notation implies, the fundamental group depends on the base point x_0. However, when the space X is path connected, the fundamental groups belonging to different base points are isomorphic. It follows, that one can speak of 'the' fundamental group in the case of path-connected spaces.

Property 7.1.16 (Groups). The fundamental group of a topological group is Abelian. This follows from an Eckmann–Hilton argument (Property 2.1.4).

[3] It should be noted that the rate at which the concatenated loops are traversed is doubled because the parameter t should remain an element of $S^1 \cong [0, 1]/_{0\sim 1}$.

Definition 7.1.17 (Fundamental groupoid ♣). Let X be a topological space. The fundamental (or **Poincaré**) groupoid $\mathbf{\Pi}_1(X)$ is the groupoid consisting of the following data:

- **Objects**: X, and
- **Morphisms**: Endpoint-preserving homotopy classes of continuous functions $f : [0,1] \to X$.

The fundamental group $\pi_1(X, x)$ can be recovered as the automorphism group of $x \in \mathsf{ob}(\mathbf{\Pi}_1(X))$.

This construction can be generalized to an ∞-*groupoid* (see Definition 14.4.1), where the class of k-morphisms is given by k-dimensional 'paths'. More explicitly, it is the *Kan complex* (see Definition 13.4.3) given by

$$\mathsf{Sing}(X) : [k] \mapsto \mathbf{Top}(\Delta^k, X)\,, \tag{7.3}$$

the *singular simplicial complex* of X (to be defined in Section 7.3).

Definition 7.1.18 (Simply-connected space). A topological space that is path connected and whose fundamental group is trivial.

Definition 7.1.19 (Semilocally simply connected). A topological space in which every point admits a neighbourhood in which every loop can be contracted to a point.

Definition 7.1.20 (Universal covering space). A simply-connected covering space (Definition 6.2.15).

Universal Property 7.1.21. Let X be a topological space and let $\widetilde{X}$ be its the universal covering space. Every other covering space of X is also covered by $\widetilde{X}$.

Property 7.1.22 (Automorphisms). Consider a topological space X and let $\widetilde{X}$ be its universal covering space. The group of deck transformations $\mathsf{Aut}(\widetilde{X})$ (Definition 6.2.18) is isomorphic to the fundamental group $\pi_1(X)$. Hence, one obtains

$$X \cong \widetilde{X}/\pi_1(X)\,. \tag{7.4}$$

Property 7.1.23 (Existence). If the space is connected, locally path-connected and semilocally simply-connected, a universal cover exists.

An explicit construction for a path-connected space is the following (when multiple path-components exist, one should take the disjoint union of the individual covers):

$$\widetilde{X} = \left\{[\gamma] \mid \gamma : [0,1] \to X, \gamma(0) = x_0\right\}, \tag{7.5}$$

where x_0 is the base point of X and $[\gamma]$ denotes the homotopy class of γ.

Corollary 7.1.24. This property also implies that, for connected, locally path-connected and semilocally simply-connected spaces, the category of coverings is equivalent to the category of continuous representations of the fundamental group.

The definition of fundamental groups can be generalized to arbitrary dimensions. (Note that, in the following definition, the interval $[0,1]$ is replaced by the sphere S^1. This is nonrestrictive as one can construct S^n by identifying the boundary of $[0,1]^n$ with the basepoint x_0.)

Definition 7.1.25 (Homotopy group). The homotopy group $\pi_n(X, x_0)$ is defined as the set of homotopy classes of continuous functions $f : S^n \to X$ based at $x_0 \in X$. The set $\pi_0(X, x_0)$ can be seen to be the set of path-connected components of X (Remark 6.4.6). This explains why the notation $\pi_0(X)$ was introduced before.

Property 7.1.26. For $n \geq 1$, the sets $\pi_n(X, x_0)$ are groups. For $n \geq 2$, the homotopy groups $\pi_n(X, x_0)$ are Abelian. This follows from an Eckmann–Hilton argument (Property 2.1.4).

Remark 7.1.27 (Relative homotopy groups). As for *(co)homology* (see Section 7.2), one can also define relative homotopy groups given a subset inclusion $A \subset X$. Two continuous functions $f, g : X \to Y$ are said to be homotopic relative to A if there exists a continous function $H : [0,1] \times X \to Y$ such that

1. $H(0, x) = f(x)$ and $H(1, x) = g(x)$, and
2. $H(t, a) = f(a) = g(a)$ for all $a \in A$.

Property 7.1.28 (Path-connectedness). If X is path connected, the homotopy groups for different basepoints are isomorphic.

Property 7.1.29 (Homeomorphisms). Homeomorphic spaces have isomorphic homotopy groups.

Formula 7.1.30 (Products). Let (X, x_0) and (Y, y_0) be pointed topological spaces with homotopy groups $\pi_n(X, x_0)$ and $\pi_n(Y, y_0)$. The homotopy groups of their product is given by

$$\pi_n(X \times Y, (x_0, y_0)) = \pi_n(X, x_0) \times \pi_n(Y, y_0) . \tag{7.6}$$

Property 7.1.31 (Whitehead bracket ♣). Consider a simply-connected topological space X. The complex $\bigoplus_{n=1} \pi_n(X)$ obtains the structure of a graded Lie algebra when equipped with the *Whitehead bracket*.

Definition 7.1.32 (Weak homotopy equivalence). A continuous function that induces isomorphisms on all homotopy groups. Two spaces connected via a weak homotopy equivalence are said to have the same **homotopy type**.

Definition 7.1.33 (n-connected space). A topological space is said to be n-connected if it is path connected and if its first $n \in \mathbb{N}$ homotopy groups are trivial. A continuous function is said to be n-connected if its induced maps on homotopy groups are isomorphisms in degrees $k < n$ and surjective in degree $n \in \mathbb{N}$.

Definition 7.1.34 (Homotopy n-type). A topological space X for which the homotopy

groups $\pi_i(X)$ vanish for $i > n$.

Definition 7.1.35 (Eilenberg–MacLane space). Let G be a group (regarded as a discrete topological space) and choose a positive integer $n \in \mathbb{N}_0$. An Eilenberg–MacLane space $K(G, n)$ is a topological space with the following property:

$$\pi_i\big(K(G,n)\big) = \begin{cases} G & \text{if } i = n\,, \\ 0 & \text{if } i \neq n\,. \end{cases} \tag{7.7}$$

It follows from Property 7.1.26 above that, for $n > 1$, the group G has to be Abelian.

Property 7.1.36 (Uniqueness). For every group G and integer $n \in \mathbb{N}_0$, the space $K(G, n)$ is unique up to weak homotopy equivalences.

Property 7.1.37 (Loop spaces). For all groups G and integers $n \geq 2$, the loop space $\Omega K(G, n)$ is homotopy equivalent to $K(G, n-1)$.

Definition 7.1.38 (Postnikov tower[4]). Consider a path-connected topological space X. A Postnikov tower of X is an inverse system of topological spaces (X_i, ϕ_i) with the following properties:

1. for all $i \leq n$, there exists an isomorphism $\pi_i(X) \cong \pi_i(X^n)$, and
2. for all $n \in \mathbb{N}$, the space X^n is a homotopy n-type.

In some cases, the morphisms $\phi : X_i \to X_{i-1}$ in the inverse system are required to be *fibrations* (see Section 7.1.4). This also implies that the fibres are Eilenberg–MacLane spaces.

The (categorically) dual notion is called the **Whitehead tower** of X. This consists of a sequence of topological spaces

$$\cdots \longrightarrow X_2 \longrightarrow X_1 \longrightarrow X \tag{7.8}$$

such that:

1. for all $i > n$, the induced maps $\pi_i(X_n) \to \pi_i(X)$ are isomorphisms, and
2. for all $n \in \mathbb{N}$, the space X_n is n-connected.

Again, one can add the requirement that the maps $\phi_i : X_i \to X_{i-1}$ are *fibrations*.

The following conjecture is due to *Baez*. Proofs are available depending on which model is used for the definition of ∞-*groupoids*.

Theorem 7.1.39 (Homotopy hypothesis ♣). **Top** *and* ∞**Grpd** *are equivalent as* $(\infty, 1)$-*categories. In particular, this means that n-groupoids are equivalent to homotopy n-types.*

[4]Often called the **Moore–Postnikov tower** or **Postnikov system** (especially in category theory).

Remark 7.1.40. The statement of the theorem is sometimes used as a consistency condition for the definition of ∞-*categories* and, in some cases, it is even used as the very definition of higher groupoids. It should also be noted that the relation is very important in *homotopy type theory* as introduced in Chapter 15.

Property 7.1.41 (Homotopy category ♣). The homotopy category **hTop** has as objects the topological spaces and as morphisms the homotopy classes of continuous functions. It is immediately clear that there exists a functor $F : \mathbf{Top} \to \mathbf{hTop}$ that acts as the identity on spaces and maps continuous functions to their homotopy classes.

However, the above definition is often too restrictive. *Quillen* gave a more general construction. The homotopy category (in the sense of Quillen) is obtained as the localization of **Top** with respect to the collection of weak homotopy equivalences. (See Section 13.3 for more information.)

Recall the reduced suspension functor Σ from Construction 6.3.7. The functors Σ and Ω are related in the following way.

Property 7.1.42 (Eckmann–Hilton duality). The reduced suspension functor Σ and the loop space functor Ω form an adjunction in the category of pointed topological spaces:

$$\mathsf{Map}_*(\Sigma X, Y) \cong \mathsf{Map}_*(X, \Omega Y)\,. \tag{7.9}$$

This also passes down to an equivalence in the associated homotopy category:

$$[\Sigma X, Y] \cong [X, \Omega Y]\,. \tag{7.10}$$

Corollary 7.1.43. By choosing $X = S^n$ and using the result from Example 6.3.8, one obtains the following important result about homotopy groups:

$$\pi_{n+1}(Y) \cong \pi_n(\Omega Y)\,. \tag{7.11}$$

Theorem 7.1.44 (Freudenthal suspension theorem). *The suspension morphism in homotopy*

$$\pi_{n+k}(S^n) \to \pi_{n+k+1}(S^{n+1}) \tag{7.12}$$

is an isomorphism for all $k \leq n-2$.

This section will be closed by giving some results that can be proven using the content of this section.

Theorem 7.1.45 (Brouwer fixed point theorem). *Every continuous function from a convex subset of* $\mathbb{R}^n$ *to itself has a fixed point.*

This theorem can be extended in the context of *Banach spaces* (see Chapter 25). There, it is called the **Schauder fixed point theorem**.

Theorem 7.1.46 (Borsuk–Ulam). *For every continuous function $f : S^n \to S^n$, there exists a point $x \in S^n$ such that $f(x) = f(-x)$.*

This theorem is often interpreted as follows in the two-dimensional setting: '*There exists a pair of antipodal points on the Earth where the temperature and air pressure are the same*'.

An application of the Borsuk–Ulam theorem is the following.

Theorem 7.1.47 (Ham-sandwich theorem). *For any n connected subsets in $\mathbb{R}^n$ with non-vanishing volume, there exists a hyperplane that cuts all of these objects exactly in half (with respect to their volume).*

Remark 7.1.48. The theorem is even stronger in that one can replace the volume by any other *measure* (see Chapter 18).

The Borsuk–Ulam theorem is equivalent to the following purely point-set theoretic theorem.

Theorem 7.1.49 (Lusternik–Schnirelmann). *For every closed cover $\{V_i\}_{1 \le i \le n+1}$ of S^n, there exists an index $i \in I$ such that V_i contains a pair of antipodal points, i.e. $V_i \cap (-V_i) \neq \emptyset$.*

7.1.3 CW complexes

Definition 7.1.50 (n-cell). An open n-cell is a subset of a topological space homeomorphic to an n-dimensional open ball. A closed n-cell is the image of an n-dimensional closed ball under an attaching map (Construction 6.3.9).

Definition 7.1.51 (CW complex). A CW complex is a Hausdorff space X together with a partition of X in open cells satisfying following conditions:

1. A subset of X is closed if and only if it intersects the closure of each cell in a closed set.
2. For each open n-cell C in the partition, there exists an attaching map $f : \overline{B}_n \to X$ such that:
   - $f|_{B_n}$ is homeomorphic to C, and
   - $f(\partial \overline{B}_n)$ is covered by a finite number of open cells in the partition, each having dimension smaller than n.

Definition 7.1.52 (Regular CW complex). A CW complex such that, for every open cell C, the attaching map f is a homeomorphism onto the closure $\overline{C}$.

Definition 7.1.53 (Finite type). A CW complex is said to be of finite type if there are only a finite number of cells in each degree.

Construction 7.1.54. Every CW complex can be constructed inductively (up to isomorphism). First, choose a discrete space X_0. This space forms the collection of 0-cells.

Then, one adds 1-cells C_1 using appropriate attaching maps $f : \partial\overline{B}_1 \to X_0$. This way, a 1-dimensional CW complex X_1 is obtained. Inductively, one obtains a sequence of nested n-dimensional CW complexes $X_0 \subset X_1 \subset \cdots \subset X_n$. The spaces X_i are also called ***i*-skeletons**.

Remark 7.1.55. Infinite-dimensional CW complexes can be obtained by taking the direct limit of the sequence above (Definition 1.8.2).

Theorem 7.1.56 (Whitehead). *A continuous function between CW complexes is a homotopy equivalence if and only if it is a weak homotopy equivalence.*

Theorem 7.1.57 (CW approximation theorem). *For every topological space X, there exists a CW complex Y together with a weak homotopy equivalence $f : X \to Y$.*

Property 7.1.58 (Suspensions). The suspension and reduced suspension of a CW complex are weakly homotopy equivalent.

Because the unit $X \to \Omega\Sigma X$ of the Eckmann–Hilton adjunction is $(2n+1)$-connected, the Freudenthal suspension theorem (Theorem 7.1.44) can be generalized to CW complexes.

Theorem 7.1.59 (Freudenthal suspension theorem). *If X is n-connected, the suspension morphism*

$$\pi_k(X) \to \pi_{k+1}(\Sigma X) \tag{7.13}$$

is an isomorphism for all $k \leq 2n$.

7.1.4 Fibrations

Definition 7.1.60 (Homotopy lifting property). Consider a continuous function $\pi : E \to B$. This function is said to have the homotopy lifting property with respect to a topological space X if for every homotopy $f : X \times [0,1] \to B$ and lift $\tilde{f}_0 : X \to E$ of $f_0 := f|_{X\times\{0\}}$ there exists a homotopy $\tilde{f} : X \times [0,1]$ lifting f such that Diagram 7.1 commutes.

$$\begin{tikzcd}
X \arrow[r, "\tilde{f}_0"] \arrow[d, hook, "X \times \{0\}"'] & E \arrow[d, "\pi"] \\
X \times [0,1] \arrow[r, "f"'] \arrow[ur, dashed, "\tilde{f}"] & B
\end{tikzcd}$$

Figure 7.1: Homotopy lifting property.

Definition 7.1.61 (Fibration). A continuous function satisfying the homotopy lifting property with respect to every topological space is called a **Hurewicz fibration**. If the

homotopy lifting property only holds with respect to CW complexes, it is called a **Serre fibration**.

Property 7.1.62 (Model fibre). Consider a fibration $\pi : E \to B$ with B path-connected. All fibres, i.e. sets $\pi^{-1}(\{b\})$ with $b \in B$, are homotopy equivalent. Therefore, a fibration is often denoted by the diagram $F \hookrightarrow E \to B$.

Example 7.1.63 (Hopf fibration). The Hopf fibration is given by

$$S^1 \hookrightarrow S^3 \to S^2 \,. \tag{7.14}$$

Adam's theorem states that this fibration can be generalized to other dimensions as $S^n \hookrightarrow S^{2n+1} \to S^{2n}$ only for $n \in \{0, 1, 3, 7\}$. (It is not a coincidence that these dimensions correspond to the dimensions of Euclidean spaces where one can consistently define a cross product or the dimensions of the real division algebras as classified by the Hurwitz theorem 22.7.5.)

Example 7.1.64. For all $n \in \mathbb{N}$, the following sequence forms a fibration:

$$\mathsf{SO}(n) \hookrightarrow \mathsf{SO}(n+1) \to S^n \,. \tag{7.15}$$

Definition 7.1.65 (Homotopy extension property). Consider a continuous function $\iota : A \to X$. This function is said to have the homotopy extension property with respect to a topological space Y if for every homotopy $f : A \times [0,1] \to Y$ and extension $\tilde{f}_0 : X \to Y$ of $f_0 := f|_{A \times \{0\}}$ there exists a homotopy $\tilde{f} : X \times [0,1] \to Y$ extending f such that Diagram 7.2 commutes, where Property 6.5.22 is used to represent homotopies in terms of the path space of Y.

$$\begin{tikzcd}
Y & & X \arrow[ll, "\tilde{f}_0"'] \arrow[dll, dashed, "\tilde{f}"] \\
Y^{[0,1]} \arrow[u, two heads, "\pi_0"] & & A \arrow[ll, "f"] \arrow[u, "\iota"']
\end{tikzcd}$$

Figure 7.2: Homotopy extension property.

Definition 7.1.66 (Cofibration). A continuous function $\iota : A \to X$ satifying the homotopy extension property with respect to every topological space Y, i.e. every extension along ι induces an extension of homotopies or, equivalently, the extensions $C(A, Y) \to C(X, Y)$ pass down to extensions $[A, Y] \to [X, Y]$.

Construction 7.1.67 (Mapping path space). Consider a continuous function $f : X \to Y$. The mapping path space P_f is defined as the following pullback:

$$P_f := \left\{ (x, p) \in X \times Y^{[0,1]} \mid f(x) = p(0) \right\} . \tag{7.16}$$

The projection

$$\pi : P_f \to Y : (x, p) \mapsto p(1) \tag{7.17}$$

is a fibration. The homotopy type of the fibres of this fibration is called the **homotopy fibre** or **mapping fibre** of f.

Construction 7.1.68 (Mapping cylinder). Let $f : X \to Y$ be a continuous function. The mapping cylinder M_f or $\mathsf{Cyl}(f)$ is defined as follows:

$$M_f := \left(X \times [0,1] \bigsqcup Y\right) / \sim_f, \tag{7.18}$$

where the equivalence relation $\sim_f$ is generated by the relations $(0, x) \sim f(x)$. So the mapping cylinder of $f : X \to Y$ is the attaching space $(X \times [0,1]) \cup_f Y$. From this definition, it follows that the 'top' of the cylinder is homeomorphic to X and the 'base' is homeomorphic to $f(X) \subseteq Y$.

By also quotienting out the relation $(1, x) \sim (1, x')$, i.e. by collapsing the top of the cylinder to a point, one obtains the so-called **mapping cone** C_f or $\mathsf{Cone}(f)$. The canonical map $\iota : Y \to C_f$ is a cofibration for all continuous functions $f : X \to Y$. Furthermore, if the image $\iota(Y)$ is closed, the coprojection $C_f \to Y/f(X)$ is a homotopy equivalence.

Remark 7.1.69 (Pushouts ♣). Just like the attaching space (Construction 6.3.9) could be obtained as a pushout in **Top**, one can also characterize the mapping cone as a pushout. The cone $\mathsf{Cone}(X)$ can be obtained as the pushout of the span $\{*\} \leftarrow X \xrightarrow{\iota_0} X \times [0,1]$. The mapping cone is then obtained as the pushout of the span $\mathsf{Cone}(X) \leftarrow X \times [0,1] \to M_f$. By the *pasting law* for pushouts, this implies that the mapping cone of f can be obtained as the pushout

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d] \arrow[dr, phantom, "\text{po}"] & Y \arrow[d] \\ \mathsf{Cone}(X) \arrow[r] & \mathsf{Cone}(f) \end{tikzcd}$$

7.1.5 Rational homotopy theory ♣

For the theory on differential graded algebras, see Chapter 4.

Definition 7.1.70 (Rational space). A simply-connected topological space X for which the homotopy groups $\pi_n(X)$ are rational vector spaces.

Definition 7.1.71 (Rational homotopy equivalence). A continuous function $f : X \to Y$ for which the induced maps on rational homotopy groups

$$\pi_n(f) \otimes \mathbb{Q} : \pi_n(X) \otimes \mathbb{Q} \to \pi_n(Y) \otimes \mathbb{Q} \tag{7.19}$$

are isomorphisms for all $n \in \mathbb{N}$. An equivalent requirement is that the induced map on rational (co)homology is an isomorphism (see the section on singular homology below).

Definition 7.1.72 (Rational homotopy category). Consider the category **Top** of topological spaces. The rational homotopy category is obtained as the *localization* (see Definition 13.2.5) of **Top** with respect to the collection of rational homotopy equivalences.

Definition 7.1.73 (Polynomial differential forms). Consider the standard n-simplex Δ^n (Definition 7.2.1). On this topological space, one can define differential forms analogous to those from Section 34.4. Let $\{t_i\}_{0\leq i\leq n}$ be $n+1$ generators in degree 0 (the barycentric coordinates). Together with $n+1$ associated generators $\mathrm{d}t_i$ in degree 1, one can construct the free GCA over $\mathbb{Q}$. To preserve the geometric structure of Δ^n, one has to quotient out the following relations:

$$\sum_{i=0}^{n} t_i = 1 \qquad \text{and} \qquad \sum_{i=0}^{n} \mathrm{d}t_i = 0\,. \tag{7.20}$$

The resulting graded algebra is denoted by $\Omega^\bullet_{\text{poly}}(\Delta^n)$. In degree 0, this complex can be identified with the space of polynomial functions on Δ^n. The ordinary differential forms can be obtained by taking the tensor product of $\Omega^\bullet_{\text{poly}}$ with the space of smooth functions on Δ^n. Under this isomorphism, the generators $\mathrm{d}t_i$ are identified with the de Rham differentials of the barycentric coordinates.

Morphisms $f : [m] \to [n]$ in the *simplex category* $\mathbf{\Delta}$ (Definition 13.1.1) induce morphisms of simplices and these in turn induce morphisms $F : \Omega^\bullet_{\text{poly}}(\Delta^n) \to \Omega^\bullet_{\text{poly}}(\Delta^m)$ defined by the following action on generators:

$$F(t_i) := \sum_{f(j)=i} t_j\,. \tag{7.21}$$

It can be seen that this turns the above construction into a functor $\Omega^\bullet_{\text{poly}} : \mathbf{\Delta}^{op} \to \mathbf{dgcAlg}$. By passing to the opposite functor and taking a left Kan extension along the Yoneda embedding $\mathbf{\Delta} \hookrightarrow \mathbf{sSet}$, one obtains a functor $\Omega^\bullet_{\text{poly}} : \mathbf{sSet} \to \mathbf{dgcAlg}^{\mathbf{op}}$. Composition with the singular set functor $\mathsf{Sing} : \mathbf{Top} \to \mathbf{sSet}$ gives the **piecewise-polynomial differential forms** functor $\Omega^\bullet_{\text{pwpoly}}$.

Definition 7.1.74 (Relative Sullivan algebra). An inclusion of DGCAs of the form

$$(A, \mathrm{d}) \hookrightarrow (A \otimes \Lambda^\bullet V, \mathrm{d}')\,,$$

where V is a $\mathbb{N}_0$-graded vector space such that:

- there is a well-ordered set I indexing a linear basis $\{e_i\}_{i\in I}$ of V.
- for all $k \in I$ and all e_k, one has that

$$\mathrm{d}' e_k \in A \otimes \Lambda^\bullet V_{<k}\,, \tag{7.22}$$

  where $V_{<k} := \mathsf{span}\{e_i\}_{i<k}$.

If, in addition, the implication

$$i < j \implies \deg(e_i) \leq \deg(e_j) \tag{7.23}$$

holds for all $i, j \in I$, the relative Sullivan algebra is said to be **minimal**. If the Sullivan algebra is defined relative to the tensor unit $(k, 0)$, where k is the underlying field, it is just called a **Sullivan algebra**.

Remark 7.1.75. The minimality condition admits an (equivalent) reformulation:

$$\mathrm{d}V \subseteq A_{\geq 1} \otimes \Lambda^\bullet V + A \otimes \Lambda^{\geq 2} V\,. \tag{7.24}$$

Example 7.1.76. The free DGCA $\Lambda^\bullet V$ on a graded vector space V of finite type with $V_0 = V_1 = 0$ is a Sullivan algebra.

Definition 7.1.77 (Sullivan model). Let X be a simply-connected topological space. A (minimal) Sullivan model for X is a (minimal) Sullivan algebra equipped with a quasi-isomorphism to the DGA of piecewise-polynomial differential forms on X.

Property 7.1.78. Minimal Sullivan models are unique up to isomorphism.

Property 7.1.79. If $(\Lambda^\bullet V^*, \mathrm{d})$ is a minimal Sullivan model of a rational space X, then

$$\pi_\bullet(X) \cong V. \tag{7.25}$$

There is also another approach due to *Quillen*. Instead of working with differential graded algebras, *Quillen* used differential graded Lie algebras, i.e. *strict L_∞-algebras* (see Section 4.9.2).

Construction 7.1.80 (Sketch, Quillen). To every simply-connected topological space X, one can associate a differential graded Lie algebra $L_\bullet(X)$ such that the homology of this complex is isomorphic to the (shifted) rational homotopy complex $\pi_{\bullet+1}(X) \otimes \mathbb{Q}$ of X.

7.1.6 Equivariant homotopy theory ♣

In this section, topological spaces equipped with an action of a topological group G will be considered (the group will often be a compact Lie group). Some notions will be defined very similarly to those in the previous section, but some others will look very differently.

Notation 7.1.81 (Fixed-point space). Consider a topological G-space X. The set of all its fixed points is denoted by X^G.

Definition 7.1.82 (Equivariant homotopy equivalence). A continuous function that is both an equivariant function and a homotopy equivalence in the ordinary sense.

Definition 7.1.83 (Weak equivariant homotopy equivalence). An equivariant continuous function that restricts to an ordinary weak equivalence between the fixed-point subspaces for all closed subgroups $H \subset G$.

Theorem 7.1.84 (Whitehead[5]). *A continuous function between G-CW complexes is a weak equivariant homotopy equivalence if and only if it is a equivariant homotopy equivalence.*

Definition 7.1.85 (Orbit category). Let G be a topological group. The orbit category $\mathbf{Orb}_G$ is the category defined by the following data:

1. the coset spaces G/H with $H \subset G$ a closed subgroup as objects, and
2. the equivariant homomorphisms as morphisms.

Theorem 7.1.86 (Elmendorf). *The map $(X, H) \mapsto X^H$ sending a G-space to its fixed-point spaces can be interpreted as a functor $X : \mathbf{Orb}_G^{op} \to \mathbf{Top}$ or, equivalently, as a $\mathbf{Top}$-valued presheaf on the orbit category. By taking this a step further, one obtains a functor sending every topological space X to such a presheaf.[6] To study the homotopy theory on these categories, one needs a choice of weak equivalences. Similar to the model structure on $\mathbf{sSet}$, one can choose the weak equivalences on $\mathbf{Psh}(\mathbf{Orb}_G)$ to be the levelwise weak equivalences. This gives an $(\infty, 1)$-equivalence*

$$\mathbf{Ho}(G\mathbf{Top}) \cong \mathbf{Psh}(\mathbf{Orb}_G)\,. \tag{7.26}$$

7.2 Simplicial homology

7.2.1 Simplices

Definition 7.2.1 (Simplex). A k-simplex $\sigma^k \equiv [t_0, \dots, t_k]$ is defined as the following set:

$$\sigma^k := \left\{ \sum_{i=0}^{k} \lambda_i t_i \,\middle|\, \sum_{i=0}^{k} \lambda_i = 1 \text{ and } \lambda_i \geq 0 \right\}. \tag{7.27}$$

where the **vertices** $t_i \in \mathbb{R}^n$ are **affinely independent**, i.e. the vectors $t_i - t_0$ are linearly independent. Equivalently, a simplicial k-simplex is the convex hull of the $k+1$ vertices $\{t_0, \dots, t_k\}$.

Remark 7.2.2 (Barycentric coordinates). The numbers λ_i from the previous definition are called barycentric coordinates. This terminology stems from the fact that the point $\sum_{i=0}^{k} \lambda_i t_i$ represents the *barycenter* of a gravitational system consisting of masses λ_i placed at the points t_i.

Example 7.2.3 (Standard simplex).

$$\Delta^k := \left\{ (x_0, \dots, x_k) \in \mathbb{R}^{k+1} \,\middle|\, \sum_i x_i = 1 \text{ and } x_i \geq 0 \right\}. \tag{7.28}$$

[5] This equivariant version of Theorem 7.1.56 is due to *Bredon*.

[6] This is the restriction of the Yoneda embedding to the subcategory $\mathbf{Orb}_G$ of $G\mathbf{Top}$.

This simplex is also called the **probability simplex** or, sometimes, the **unit simplex**.

Notation 7.2.4 (Face). Consider a k-simplex $[v_0, \dots, v_k]$. The **face** opposite to the vertex v_i is the $(k-1)$-simplex $[v_0, \dots, \hat{v}_i, \dots, v_k]$ obtained by removing the vertex v_i.

Definition 7.2.5 (Simplicial complex). A simplicial complex $\mathcal{K}$ is a set of simplices satisfying the following conditions:

- If σ is a simplex in $\mathcal{K}$, so are its faces.
- If $\sigma_1, \sigma_2 \in \mathcal{K}$, either $\sigma_1 \cap \sigma_2 = \emptyset$ or $\sigma_1 \cap \sigma_2$ is a face of both σ_1 and σ_2.

A simplicial k-complex is a simplicial complex where every simplex has dimension at most $k \in \mathbb{N}$.

Definition 7.2.6 (Path-connected complex). A simplicial complex in which every two vertices are connected by an edge.

Definition 7.2.7 (Polyhedron). Consider a simplicial complex. The polyhedron associated with it is the topological space constructed by equipping the complex with the Euclidean subspace topology.

Definition 7.2.8 (Triangulable spaces). Let X be a topological space and let $\mathcal{K}$ be a polyhedron. If there exists a homeomorphism $\varphi : \mathcal{K} \to X$, then X is said to be triangulable and $\mathcal{K}$ is called a **triangulation** of X.

Property 7.2.9 (Fundamental group). Let $\mathcal{K}$ be a path-connected polyhedron with basepoint a_0 and consider a contractible one-dimensional subpolyhedron $\mathcal{C} \subset \mathcal{K}$ containing all vertices of $\mathcal{K}$. Let G be the free group (Definition 2.2.52) with generators g_{ij} corresponding to the ordered 1-simplices $[v_i, v_j] \in \mathcal{K}$. On this group, one can define the following relations:

- $g_{ij} = e$ if $[v_i, v_j] \in \mathcal{C}$.
- $g_{ij} g_{jk} = g_{ik}$ for every ordered 2-simplex $[v_i, v_j, v_k] \in \mathcal{K} \backslash \mathcal{C}$.

The quotient group of G by these relations is isomorphic to the fundamental group $\pi_1(\mathcal{K}, a_0)$.

Corollary 7.2.10. Property 7.1.29, which states that homeomorphic spaces have the same homotopy groups, implies that the fundamental group of a triangulable space can be computed by looking at any of its triangulations.

7.2.2 Simplicial homology

In this section, a chain complex (Definition 5.1.1) and its associated homology theory are constructed for triangulable spaces.

Definition 7.2.11 (Chain group). Let $\mathcal{K}$ be a simplicial n-complex. The kth chain group

$C_k(\mathcal{K})$ is defined as the free Abelian group generated by the k-simplices in $\mathcal{K}$:

$$C_k(\mathcal{K}) := \left\{ \sum_i a_i \sigma_i \,\middle|\, \sigma_i \text{ is a } k\text{-simplex in } \mathcal{K}, a_i \in \mathbb{Z} \right\}. \tag{7.29}$$

For $k > n$ and $k < 0$, $C_k(\mathcal{K})$ is defined to be $\{0\}$.

Definition 7.2.12 (Boundary operator). The boundary operator $\partial_k : C_k(\mathcal{K}) \to C_{k-1}(\mathcal{K})$ is the group morphism defined by the following properties:

- **Linearity**:

$$\partial_k \Big(\sum_i a_i \sigma_i \Big) = \sum_i a_i \partial_k \sigma_i \,, \tag{7.30}$$

- **Action on generators**

$$\begin{aligned} \partial_k [v_0, \dots, v_k] &= \sum_{i=0}^{k} (-1)^i [v_0, \dots, \hat{v}_i, \dots, v_k] \,, \\ \partial_0 [v] &= 0 \,. \end{aligned} \tag{7.31}$$

Property 7.2.13 (Chain condition). The boundary operators satisfy the following relation:

$$\partial_k \circ \partial_{k+1} = 0 \,. \tag{7.32}$$

This property turns the system (C_k, ∂_k) into a chain complex (Definition 5.1.1).

Definition 7.2.14 (Cycle group). The kth cycle group $Z_k(\mathcal{K})$ is defined as the set of k-chains σ_k such that $\partial_k \sigma_k = 0$. These chains are called **cycles**.

Definition 7.2.15 (Boundary group). The kth boundary group $B_k(\mathcal{K})$ is defined as the set of k-chains σ_k for which there exists a $(k+1)$-chain b such that $\partial_{k+1} b = \sigma_k$. These chains are called **boundaries**.

Definition 7.2.16 (Homology group). From Property 7.2.13, it follows that $B_k(\mathcal{K}) \subset Z_k(\mathcal{K})$ is a subgroup (it is a normal subgroup because of the Abelian character). One can thus define the kth homology group $H_k(\mathcal{K})$ as the following quotient group:

$$H_k(\mathcal{K}) := Z_k(\mathcal{K}) / B_k(\mathcal{K}) \,. \tag{7.33}$$

Theorem 2.2.60 says that $H_k(\mathcal{K})$ can be written as $G_k \oplus T_k$. Both of these groups say something about $\mathcal{K}$. The rank of G_k, denoted by $R_k(\mathcal{K})$, is equal to the number of $(k+1)$-dimensional holes in $\mathcal{K}$. The torsion subgroup T_k says how the space $\mathcal{K}$ is 'twisted'.

Definition 7.2.17 (Betti numbers). The ranks $R_k(\mathcal{K})$ are called the Betti numbers of $\mathcal{K}$.

Definition 7.2.18 (Euler characteristic). The Euler characteristic of a triangulable space X is defined as follows:

$$\chi(X) := \sum_i (-1)^i R_i(X) \,. \tag{7.34}$$

This formula is sometimes called the **Poincaré** or **Euler–Poincaré formula**.

Property 7.2.19 (Isomorphisms). If two topological spaces have the same homotopy type, in particular when they are homeomorphic, they have isomorphic homology groups.

Corollary 7.2.20. As was the case for the fundamental group, it follows from the definition of a triangulation that one can study the homology groups of a given triangulable space by looking at any of its triangulations.

Although the homology invariants do not depend on the choice of triangulation, a remark about the existence of nonequivalent triangulations should be given.

Remark 7.2.21 (Hauptvermutung). Before the construction of a counterexample, it was believed (hence the terms *Hauptvermutung* in German or *main conjecture* in English) that every two triangulations of a topological space allowed a common refinement and, hence, where equivalent for many constructions. However, it was shown that this conjecture is generally false, e.g. for topological manifolds in dimensions 5 and higher, there exist an infinite number of nonequivalent triangulations. In dimensions up to 4, it was proven by Radó and Moise that the Hauptvermutung holds for all topological manifolds (see also Theorem 31.1.7).

Construction 7.2.22. The definition of homology groups can be generalized by letting the (formal) linear combinations used in the definition of the chain group be of the following form:

$$c = \sum_i g_i \sigma_i^k \,, \tag{7.35}$$

where $g_i \in G$ for some Abelian group G. The kth homology group of X with coefficients in G is denoted by $H_k(X; G)$.

Property 7.2.23 (Vanishing torsion). When G is a field, such as $\mathbb{Q}$ or $\mathbb{R}$, the torsion subgroups T_k vanish. The relation between integral homology and homology with coefficients in a group is given by the *universal coefficient theorem*.

Formula 7.2.24 (Künneth formula). Let X, Y be two simplicial complexes. The homology groups of the Cartesian product $X \times Y$ with coefficients in a field $\mathfrak{K}$ is given by

$$H_k(X \times Y; \mathfrak{K}) = \bigoplus_{k=i+j} H_i(X; \mathfrak{K}) \otimes H_j(Y; \mathfrak{K}) \,. \tag{7.36}$$

When the requirement of $\mathfrak{K}$ being a field is relaxed to it merely being a group, the torsion subgroups have to be taken into account. See the literature for this general form.

7.2.3 Relative homology

In this section, the homology of a simplicial complex $\mathcal{K}$ 'modulo' a subcomplex $\mathcal{L}$ is considered.

Definition 7.2.25 (Relative chain group). The k-chain group of $\mathcal{K}$ modulo $\mathcal{L}$ is defined as the following quotient group:

$$C_k(\mathcal{K}, \mathcal{L}) := C_k(\mathcal{K})/C_k(\mathcal{L}) \,. \tag{7.37}$$

Equivalence classes will be denoted by square brackets $[c]$, where $c \in C_\bullet(\mathcal{K})$.

Definition 7.2.26 (Relative boundary operator). The relative boundary operator $\overline{\partial}_k$ is defined as follows:

$$\overline{\partial}_k[c_k] := [\partial_k c_k] \,, \tag{7.38}$$

where ∂_k on the right-hand side denotes the boundary operator in ordinary homology. This is a well-defined operation since $\partial C_\bullet(\mathcal{L}) \subseteq C_\bullet(\mathcal{L})$.

Definition 7.2.27 (Relative homology groups). The relative cycle and relative boundary groups are defined analogous to their ordinary counterparts. The relative homology groups are defined as follows:

$$H_k(\mathcal{K}, \mathcal{L}) := \frac{\ker(\overline{\partial}_k)}{\operatorname{im}(\overline{\partial}_{k+1})} \,. \tag{7.39}$$

Elements $h_k \in H_k(\mathcal{K}, \mathcal{L})$ can be represented as $h_k = z_k + C_k(\mathcal{L})$ such that $\partial_k z_k \in C_{k-1}(\mathcal{L})$.

Property 7.2.28 (Homotopy invariance). Consider two (triangulable) topological spaces X, Y with subspaces $A \subset X, B \subset Y$. If a continuous function $f : X \to Y$ is a homotopy equivalence such that the restriction to A gives a homotopy equivalence $f|_A : A \to B$, the relative homology complexes $H_\bullet(X, A)$ and $H_\bullet(Y, B)$ are isomorphic.

Property 7.2.29 (Long sequence for a pair). The relative homology groups fit in the following (long) exact sequence:

$$\cdots \longrightarrow H_k(\mathcal{L}) \xrightarrow{i_*} H_k(\mathcal{K}) \xrightarrow{j_*} H_k(\mathcal{K}, \mathcal{L}) \xrightarrow{\partial_k} H_{k-1}(\mathcal{L}) \longrightarrow \cdots \,, \tag{7.40}$$

where i_* and j_* are the homology morphisms induced by the inclusions $i : \mathcal{L} \to \mathcal{K}$ and $j : \mathcal{K} \to (\mathcal{K}, \mathcal{L})$.

Theorem 7.2.30 (Excision theorem). *Let U, V and X be simplicial complexes such that $U \subset V \subset X$. If the closure $\overline{U}$ is contained in the interior V°, then*

$$H_k(X, V) \cong H_k(X \backslash U, V \backslash U) \,. \tag{7.41}$$

Definition 7.2.31 (Reduced homology). Consider the chain complex $C_\bullet(X)$. Example 7.2.34 says that a point $\{*\}$ has homology $\mathbb{Z}$ concentrated in degree 0. However, it would be nice if $\{*\}$ had vanishing homology. To this end, one can augment (Definition 5.3.2) the chain complex $C_\bullet(X)$ of any topological space by the group $\mathbb{Z}$ to obtain the reduced homology complex $\widetilde{H}_\bullet(X)$:

$$H_k(X) \cong \begin{cases} \widetilde{H}_0(X) \oplus \mathbb{Z} & \text{if } k = 0\,, \\ \widetilde{H}_k(X) & \text{if } k \neq 0\,. \end{cases} \tag{7.42}$$

Property 7.2.32. Consider a pointed topological space (X, x_0). The reduced homology $\widetilde{H}_\bullet(X)$ is isomorphic to the relative homology $H_\bullet(X, \{x_0\})$.

Property 7.2.33 (Good pair). Consider a topological space X together with a subspace A. The pair (X, A) is called a **neighbourhood deformation retract (NDR) pair**[7] if there exists a neighbourhood $V \subset X$ of A such that V deformation retracts onto A. Equivalently, the inclusion $A \subset X$ is a closed cofibration (Definition 7.1.66).

Given an NDR pair (X, A), there exists an isomorphism

$$H_\bullet(X, A) \cong \widetilde{H}_\bullet(X/A)\,. \tag{7.43}$$

@@ STREAMLINE TERMINOLOGY AND NOTATION (topological space vs. triangulable space vs. simplicial complex, X vs K, ...) @@

7.2.4 Examples

Example 7.2.34. Let X be a contractible space.

$$H_k(X) = \begin{cases} \mathbb{Z} & \text{if } k = 0 \\ 0 & \text{otherwise} \end{cases} \tag{7.44}$$

Example 7.2.35. Let $\mathcal{K}$ be a path-connected simplicial complex.

$$H_0(\mathcal{K}) = \mathbb{Z} \tag{7.45}$$

Furthermore, every point $p \in \mathcal{K}$ determines a generator $\langle p \rangle \in H_0(\mathcal{K})$.

Example 7.2.36. The homology groups of the n-sphere S^n are given by

$$H_k(S^n) = \begin{cases} \mathbb{Z} & \text{if } k = 0 \text{ or } k = n\,, \\ 0 & \text{otherwise}\,. \end{cases} \tag{7.46}$$

Definition 7.2.37 (Homology sphere). An *n-dimensional manifold* (see Chapter 31) having the same homology groups as the n-sphere.

[7] Some authords call this a **good pair**, while others define an NDR pair more restrictively.

Definition 7.2.38 (Degree). The example above says that $H_n(S^n) = \mathbb{Z}$ for any $n \in \mathbb{N}$. Given a map $f : S^n \to S^n$, the induced map f_* on homology is an endomorphism of $\mathbb{Z}$ and, hence, is of the form $f(x) = \lambda x$ where $\lambda \in \mathbb{Z}$. The coefficient λ is called the degree of f, sometimes denoted by $\deg(f)$.

Property 7.2.39. Two maps $f : S^n \to S^n$ have the same degree if and only if they are homotopic.

Example 7.2.40. Consider the closed (or open) disks D_n.

$$H_k(D_n, \partial D_n) = \begin{cases} \mathbb{Z} & \text{if } k = n \\ 0 & \text{otherwise} \end{cases} \tag{7.47}$$

This holds for any n-skeleton X^n of a CW-complex.

7.3 Singular homology

Definition 7.3.1 (Singular simplex). Recall the k-simplex Δ^k from Example 7.2.3. A singular k-simplex in a topological space X is defined as a continuous function $\sigma : \Delta^k \to X$.

Remark. The name singular comes from the fact that the function σ need not be injective.

Definition 7.3.2 (Δ-complex). Let $\{\sigma_i : \Delta^n \to X\}_{i \in I}$ be a collection of singular simplices in X, where the dimension $n \in \mathbb{N}$ may depend on the index $i \in I$. This collection forms a Δ-complex on X if it satisfies the following conditions:

- The restrictions $\sigma_i|_{\mathring{\Delta}^n}$ are injective and every point in X lies in the image of exactly one such restriction.
- The restriction of a simplex σ_i to any one of the faces of Δ^n is equal to some other σ_j.
- A subset $Y \subseteq X$ is open if and only if all of its inverse images $\sigma_i^{-1}(Y)$ are open.

Similar to a CW-complex or *cellular complex*, these conditions imply that every Δ-complex can be constructed inductively from a (discrete) set of vertices by gluing and identifying edges.

Definition 7.3.3 (Singular chain group). The singular chain group $S_k(X)$ with coefficients in a group G is defined as the set of formal linear combinations $\sum_i g_i \sigma_i$, where the σ_i are singular k-simplices in X. The basis of this free group is, in most cases, infinite as there are, in general, many ways to map Δ^k to X.

Before continuing, one first need to introduce an important concept in the context of simplicial objects.

Definition 7.3.4 (Face map). The face maps are morphisms $\varepsilon_i^k : \Delta^{k-1} \to \Delta^k$ that map Δ^{k-1} onto the ith face of Δ^k. They are explicitly given by

$$\varepsilon_i^k(s_0, \dots, s_{k-1}) := (s_0, \dots, s_{i-1}, 0, s_i, \dots, s_{k-1}) \,. \tag{7.48}$$

Their defining property is the following relation:

$$\varepsilon_i^k \circ \varepsilon_j^{k-1} = \varepsilon_j^k \circ \varepsilon_{i-1}^{k-1} \,, \tag{7.49}$$

where $j \leq i$.

Remark. Some authors (for example the authors at *nLab*) call these maps *degeneracy maps* and call what in these notes are called *degeneracy maps*, face maps. This is a consequence of working in the dual picture (they work in the opposite of the *simplex category* $\mathbf{\Delta}$).

Definition 7.3.5 (Singular boundary operator). The singular boundary operator[8] ∂ is defined by its linear action on the singular chain groups $S_k(X;G)$. It follows, that it is uniquely defined by its action on the singular k-simplices σ^k.

The action of the boundary operator on the singular k-simplex σ^k is given by

$$\partial_k \sigma^k = \sum_{i=0}^{k} (-1)^i \sigma^k \circ \varepsilon_i^k \,, \tag{7.50}$$

where the ε_i^k are the face maps defined above. The singular boundary operators satisfy the same relation as the the simplicial boundary operators:

$$\partial_{k-1} \circ \partial_k = 0 \,. \tag{7.51}$$

It follows, that $S_k(X;G)$ is also a chain complex.

Definition 7.3.6 (Singular homology group). The singular homology groups of a topological space with coefficients in an Abelian group G are defined as follows:

$$H_k(X;G) := \frac{\ker(\partial_k)}{\operatorname{im}(\partial_{k+1})} \,. \tag{7.52}$$

Property 7.3.7 (Simplicial homology). For triangulable spaces, singular homology is isomorphic to simplicial homology. When a space is not triangulable, this property is not valid. The singular approach to homology is more general, but it is often more difficult to compute (even in the case of triangulable spaces).

Property 7.3.8 (Induced morphism). Consider a continuous function $f : X \to Y$ between topological spaces. This induces a morphism $f_* : S_k(X;G) \to S_k(Y;G)$ on the chain groups as follows:

$$f_*\Big(\sum_\sigma c_\sigma \sigma\Big) := \sum_\sigma c_\sigma f \circ \sigma \,. \tag{7.53}$$

[8] The same notation as for simplicial boundary operators is used for simplicity.

This map takes cycle (resp. boundary) groups to (subgroups of) cycle (resp. boundary) groups and, hence, induces a morphism of homology groups, called the **pushforward**:

$$f_* : H_k(X) \to H_k(Y) : \langle h \rangle \mapsto \langle f_*(h) \rangle \,. \tag{7.54}$$

Corollary 7.3.9. H_k is a functor **Top** $\to$ **Ab** that maps topological spaces to their homology groups and continuous functions f to their pushforwards f_*.

Theorem 7.3.10 (Hurewicz). *Let X be path connected and let $[\cdot]$ and $\langle\cdot\rangle$ denote the equivalence classes in the homotopy and homology groups, respectively. Because every path can be obtained as a singular one-chain, the map $\pi_1(X) \to H_1(X) : [\gamma] \mapsto \langle\gamma\rangle$ defines a group morphism. Furthermore, this map induces an isomorphism*

$$\pi_1(X)/[\pi_1(X), \pi_1(X)] \cong H_1(X) \,. \tag{7.55}$$

More generally, for every topological space X and every $k \in \mathbb{N}$, there exists a morphism $h_* : \pi_k(X) \to H_k(X)$. If Y is $(n-1)$-connected, then, for every $k \leq n$, this morphism is in fact an isomorphism.

Definition 7.3.11 (Singular cohomology). The singular cohomology groups of a topological space X with coefficients in an Abelian group G are defined as follows:[9]

$$H^k(X; G) := \mathsf{Hom}\big(H_k(X), G\big) \,, \tag{7.56}$$

where, on the right-hand side, the integral (singular) homology of X is used. A continuous function $f : X \to Y$ induces a **pullback** morphism $f^* : H^\bullet(Y; G) \to H^\bullet(X; G)$ on cohomology by duality:

$$f^* g(\sigma) := g(f_* \sigma) \,, \tag{7.57}$$

where f_* is the pushforward (Property 7.3.8) induced by f.

Property 7.3.12 (Representability). Let X be a CW complex. There exists an isomorphism

$$[X, K(G, n)] \to H^n(X; G) \tag{7.58}$$

between the homotopy classes of maps $X \to K(G, n)$ and the nth singular cohomology of X with coefficients in G. (This result can be widely generalized cf. Theorem 7.5.3 further below.)

[9] Actually this is the *Universal Coefficient Theorem* when defining cohomology through a suitable hom-complex.

Proof. Consider the Eilenberg–Maclane space $K(G,n)$ for G (Definition 7.1.35). By the Hurewicz theorem, there exists an isomorphism $H_n\big(K(G,n)\big) \cong \pi_n\big(K(G,n)\big) \cong G$. By definition of cohomology, $H^n\big(K(G,n);G\big) = \mathsf{Hom}\big(H_n(K(G,n)),G\big)$ and, thus, $H^n\big(K(G,n);G\big) \cong \mathsf{Hom}(G,G)$. In particular, there corresponds a cohomology class ψ to the identity mapping on G. The cohomology class in $H^n(X;G)$ associated to a homotopy class of functions $f : X \to K(G,n)$ is given by $f^*\psi$. □

7.4 Local coefficients

For some applications, one needs to generalize the notion of (co)homology. For example, when considering integration theory on *manifolds* (see Section 34.9.3), one needs to take into account the nonorientability.

Consider a topological space X that admits a universal cover $\widetilde{X}$. By Property 7.1.22, one then has

$$X \cong \widetilde{X}/\pi_1(X)\,. \tag{7.59}$$

Given a $\mathbb{Z}[\pi_1(X)]$-module A, one defines the A-twisted chain (singular) group of X as

$$C(X;A) := C(\widetilde{X}) \otimes_{\mathbb{Z}[\pi_1(X)]} A\,. \tag{7.60}$$

The automorphism group $\pi_1(X)$ of $\widetilde{X}$ acts on singular simplices by postcomposition.

When A is a trivial $\pi_1(X)$-module, the tensor product structure enforces that the orbit of every automorphism collapses and, hence, $C(\widetilde{X}) \otimes_{\mathbb{Z}[\pi_1(X)]} A \cong C(X) \otimes_{\mathbb{Z}} A$. It follows, that the ordinary singular cohomology with coefficients in A is recovered (as expected).

@@ COMPLETE @@

7.5 Axiomatic approach ♣

Definition 7.5.1 (Eilenberg–Steenrod axioms). All (co)homology theories have a set of properties in common. By treating these properties as axioms, one can construct relative (co)homology theories as a sequence of functors $H_k : \mathbf{Top} \times \mathbf{Top} \to \mathbf{Ab}$.[10] The axioms are as follows:

1. **Homotopy invariance**: If f, g are homotopic maps, their induced homology maps are the same:
$$f \cong g \implies \forall k \in \mathbb{N} : H_k(f) = H_k(g).$$

[10] Technically, one should replace $\mathbf{Top} \times \mathbf{Top}$ by the category consisting of subspace inclusions $A \hookrightarrow X$.

2. **Excision**: If $U \subset V \subset X$ and $\overline{U} \subset V^\circ$, then $H_k(X, V) \cong H_k(X \backslash U, V \backslash U)$.
3. **Additivity**: If $X = \bigsqcup_i X_i$, then $H_k(X) \cong \bigoplus_i H_k(X_i)$.
4. **Exactness**: Each pair (X, A), where $A \subset X$, induces a long exact sequence
$$\cdots \longrightarrow H_k(A) \xrightarrow{i_*} H_k(X) \xrightarrow{j_*} H_k(X, A) \xrightarrow{\partial_k} H_{k-1}(A) \longrightarrow \cdots, \tag{7.61}$$
where i_* and j_* are the pushforwards of the inclusions $i : A \to X$ and $j : X \to (X, A)$.
5. **Dimension**: If X is a singleton, then $H_k(X) = \{0\}$ for all $k \geq 1$. The group $H_0(X)$ is called the **coefficient group** and gives the coefficients used in the linear combinations of the chain group.

Remark 7.5.2. If the dimension axiom is removed from the set of axioms, the so-called *extraordinary* (*or generalized*) *homology theories* are obtained.

Although the following theorem sounds like more of a purely category-theoretic statement, its main application is the definition of cohomology theories.

Theorem 7.5.3 (Brown's representability theorem). *Consider a presheaf H on the homotopy category of pointed connected topological spaces* $\mathbf{Ho}(\mathbf{Top}_*^{con})$. *This functor is representable if and only if it satisfies the following conditions:*

- *It maps coproducts to products.*
- *It maps weak pushouts to weak pullbacks.* (*Weak pushouts in the homotopy category* $\mathbf{Ho}(\mathbf{Top})$ *come from homotopy pushouts in* $\mathbf{Top}$.)

Remark 7.5.4. If one constructs the homotopy category using the CW-model structure in $\mathbf{Top}$, the second condition can be restated as a *Mayer–Vietoris axiom*. Consider a CW complex U and two subcomplexes V_1, V_2. If $x \in H(V_1), y \in H(V_2)$ and $x = y$ on the intersection $V_1 \cap V_2$, there exists a $z \in H(U)$ that restricts to x (resp. y) on V_1 (resp. V_2).

Corollary 7.5.5 (Cohomology). Generalized cohomology theories are representable. Given cohomology functors $\{H^n\}_{n\in\mathbb{N}}$, there exists a sequence of pointed topological spaces $(X_n)_{n\in\mathbb{N}}$ such that
$$H^n(Y) := [Y, X_n] \tag{7.62}$$
for all $n \in \mathbb{N}$.

For reduced cohomology theories, there exist suspension isomorphisms
$$\widetilde{H}^n(Y) \cong \widetilde{H}^{n+1}(\Sigma Y)\,. \tag{7.63}$$
Under Brown's theorem, these induce isomorphisms $X_n \cong \Omega X_{n+1}$. This endows the sequence of spaces $(X_n)_{n\in\mathbb{N}}$ with the following structure.

Definition 7.5.6 (Spectrum[11]). A sequence of pointed topological spaces $(X_n)_{n\in\mathbb{N}}$ satisfying

$$X_n \cong \Omega X_{n+1}\,. \tag{7.64}$$

(The isomorphism can be a weak homotopy equivalence or homeomorphism depending on the context.) A weaker definition, that of **prespectra** or **sequential spectra**, is that of a sequence of pointed topological spaces $(X_n)_{n\in\mathbb{N}}$ with structure morphisms $\Sigma X_n \to X_{n+1}$, where Σ is the reduced suspension functor. By Eckman–Hilton duality, every Ω-spectrum is a prespectrum.

Every prespectrum gives rise to a spectrum by taking a direct limit over loop spaces:[12]

$$(LX)_n := \varinjlim_{k\in\mathbb{N}} \Omega^k X_{n+k}\,. \tag{7.65}$$

This gives a *fibrant replacement functor* in the *model structure* on spectra.

Property 7.5.7. Brown's representability theorem shows that there exists a bijection between isomorphism classes of (generalized) cohomology theories and homotopy-equivalence classes of spectra.

Property 7.5.8 (Eilenberg–MacLane spectrum). The Eilenberg–MacLane spectrum of an Abelian group A is the sequence of Eilenberg–MacLane spaces $\{K(A,n)\}_{n\in\mathbb{N}}$. By the *Dold–Kan theorem* (see Theorem 13.1.19), Abelian groups form a subcategory of Abelian chain complexes: $\mathbf{Ab} \hookrightarrow \mathbf{Ch}^+(\mathbf{Ab})$. The Eilenberg–MacLane assignment extends to a functor $\mathcal{D}(\mathbf{Ab}) \to \mathbf{Spectra}$ from the derived category of Abelian chain complexes to the category of (Ω-)spectra.

Property 7.5.9. There exists an (∞-)equivalence between $\mathbf{Sp}$ and $\infty\mathbf{Grpd}$. The left adjoint sends (a topological space representing) a given homotopy type to its suspension spectrum. The right adjoint sends a spectrum $(X_n)_{n\in\mathbb{N}}$ to (the homotopy type of) its infinite loop space

$$\Omega^\infty X := X_0 \cong \varinjlim_{k\in\mathbb{N}} \Omega^k X_k\,. \tag{7.66}$$

7.6 Equivariant cohomology ♣

In this section, the action of (topological) groups G on topological spaces will be considered. These spaces will be called topological G-spaces or just G-spaces. See, for example, Cartier (2001); Rozenblyum, Grantcharov, and Slipper (2023). The content of Chapter 8 is used.

[11] Sometimes called an Ω-spectrum to distinguish it from other, possibly more general, spectra.

[12] The notion, similar to the one in Definition 7.1.11, stems from the relation to loop spaces in the case of suspensions $X_n = \Sigma^n Y$ (Construction 6.3.7).

7.6.1 Borel construction

Definition 7.6.1 (Equivariant cohomology). Let X be a topological G-space for which the G-action is free. The equivariant cohomology of X is defined as

$$H^\bullet_G(X) := H^\bullet(X/G)\,, \tag{7.67}$$

where X/G is the orbit space with respect to the action of G on X.

If the action is not free, a more general construction needs to be used. Consider the *universal bundle* $\pi : EG \to BG$ from Definition 35.2.1. From this bundle, one can construct the associated bundle $EG \times_G X$. This is sometimes called the **Borel construction**. It gives a model for the homotopy quotient $X /\!\!/ G$, with $BG \cong * /\!\!/ G$. The G-equivariant cohomology of X is then defined as the singular cohomology of the Borel construction:

$$H^\bullet_G(X) := H^\bullet(EG \times_G X)\,. \tag{7.68}$$

@@ COMPLETE (E.G. LECTURES OF TU ON YOUTUBE) @@

7.6.2 Classifying topoi and stacks

The topological constructions from the previous section can reformulated in a more categorical/topos-theoretic way. As a first step, the Borel construction is rephrased.

Property 7.6.2. The geometric realization (Remark 13.1.12) of the action groupoid (Example 3.5.10) is isomorphic to the Borel construction:

$$|X /\!\!/ G| \cong EG \times_G X\,. \tag{7.69}$$

Definition 7.6.3 (Classifying topos). First, consider a discrete group G. The classifying topos $\mathcal{B}G$ is defined as the category of G-sets together with equivariant maps. Equivalently, consider the delooping groupoid $\mathbf{B}G$ (Example 3.5.11). $\mathcal{B}G$ is equivalent to the (pre)sheaf topos on the classifying space $BG \cong |\mathbf{B}G|$.

For general topological groups, the situation is more complicated. Instead of having the trivial (Grothendieck) topology on the category of G-sets, G-equivariant covers have to be considered, i.e. topological covers consisting of equivariant maps. The classifying topos $\mathcal{B}G$ then consist of sheaves on this site, i.e. the *G-equivariant sheaves* (see Definition 8.3.1).

Definition 7.6.4 (Quotient stack). Consider a topological group G and a topological space X. The quotient stack $[X/G]$ is the stack (Section 14.3) that assigns to any topological space Y the following groupoid:

1. **Objects**: Pairs $(\pi : P \to Y, f : P \to X)$, where π is a principal G-bundle and f is G-equivariant.

2. **Morphisms**: G-equivariant isomorphisms.

Equivalently, consider the action groupoid $X /\!\!/ G$ (Example 3.5.10). This groupoid induces a prestack that assigns to every topological space Y the action groupoid $\underline{X}(Y) /\!\!/ \underline{G}(Y)$. The sheafification of this prestack is equivalent to $[X/G]$.

Example 7.6.5. When X is trivial, the delooping groupoid $\mathbf{B}G$ is recovered, which is also called the **classifying stack** in this form. On the other hand, when G is trivial, the representable stack $\underline{X}$ is recovered.

Chapter 8

Sheaf Theory

A reference aimed towards the study of differential geometry is Brylinski (1993). For the concept of a sheaf in category theory, see Section 12.4.

Chapter Contents

8.1 Presheaves **215**

8.2 Sheaves **216**

8.3 Equivariance **219**

8.4 Abelian sheaf cohomology **220**

8.4.1 Derived cohomology 220

8.4.2 Čech cohomology 221

8.5 Non-Abelian sheaf cohomology **222**

8.5.1 Čech cohomology 222

8.6 Ringed spaces **223**

8.1 Presheaves

Definition 8.1.1 (Presheaf). Let (X, τ) be a topological space. A presheaf on X consists of a choice of algebraic structure $\mathcal{F}(U)$ for every open set $U \in \tau$ and a morphism $\Phi_V^U : \mathcal{F}(U) \to \mathcal{F}(V)$ for every two open sets $U, V \in \tau$ with $V \subseteq U$ such that the following conditions are satisfied:

1. $\Phi_U^U = \mathbb{1}_{\mathcal{F}(U)}$, and
2. $W \subseteq V \subseteq U \implies \Phi_W^U = \Phi_W^V \circ \Phi_V^U$.

The set $\mathcal{F}(U)$ is called the set of **sections** over U and the morphisms Φ_V^U are called the **restriction maps**.

Definition 8.1.2 (Morphism of presheaves). Let $\mathcal{F}, \mathcal{F}'$ be two presheaves on a topological space X. A morphism $\mathcal{F} \to \mathcal{F}'$ is a set of morphisms $\Psi_U : \mathcal{F}(U) \to \mathcal{F}'(U)$ that commute with the restriction maps Φ_V^U.

Alternative Definition 8.1.3 (Categorical definition). Using the language of category theory one can give a more concise definition. Let $\mathbf{C}$ be a category and let X be a topological space. A $\mathbf{C}$-valued presheaf on X is a contravariant functor $\mathcal{F} : \mathbf{Open}(X) \to \mathbf{C}$. A morphism of presheaves is a natural transformation between such functors. As such the category of presheaves on a topological space X is in fact the presheaf topos on $\mathbf{Open}(X)$.

Example 8.1.4 (Constant presheaf). Let S be any set. The constant presheaf on X with target S is defined by

$$\mathcal{F}(U) := S \tag{8.1}$$

for every open set $U \subseteq X$.

8.2 Sheaves

Definition 8.2.1 (Sheaf). Let (X, τ) be a topological space. A sheaf on X is a presheaf $\mathcal{F}$ satisfying the following conditions:

1. **Locality (or separation)**: Let $\{U_i\}_{i \in I} \subset \tau$ be an open cover of $U \subseteq X$ and consider sections $s, t \in \mathcal{F}(U)$. If $\forall i \in I : s|_{U_i} = t|_{U_i}$, then $s = t$. This is equivalent to saying that that $\mathcal{F}(U)$ injects into $\prod_i \mathcal{F}(U_i)$ for all open covers of U.

2. **Gluing**: Let $\{U_i\}_{i \in I} \subset \tau$ be an open cover of $U \subseteq X$ and let $\{s_i \in \mathcal{F}(U_i)\}_{i \in I}$ be a collection of local sections. If $\forall i, j \in I : s_i|_{U_i \cap U_j} = s_j|_{U_i \cap U_j}$, there exists a section $s \in \mathcal{F}(U)$ such that $\forall i \in I : s|_{U_i} = s_i$.

Remark 8.2.2 (Separated presheaves). If a global section exists by the gluing condition, it is automatically unique by the separation axiom. In some texts, these two conditions are combined in a single gluing condition that requires a unique global section. Presheaves satisfying only the first condition are said to be **separated**. (See also the footnote in Definition 12.4.11.)

Notation 8.2.3 (Category of sheaves). Similar to the case of presheaves, one can define a morphism of sheaves as a collection of morphisms that commute with the restriction maps. The sheaves and sheaf morphisms on a space X form a full subcategory of the category of presheaves, denoted by $\mathbf{Sh}(X)$.

Property 8.2.4 (Sheaf topos). The category of sheaves $\mathbf{Sh}(X)$ is in fact an *elementary topos*, called the sheaf topos on X. (See Property 12.4.30.)

Property 8.2.5. Let X be a topological space and let $\mathcal{F}$ be a presheaf on X. $\mathcal{F}$ is a sheaf

on X if for any open $U \subseteq X$ and every open cover $\{U_i\}_{i\in I}$ of U the following diagram is an equalizer diagram:

$$\mathcal{F}(U) \to \prod_{i\in I} \mathcal{F}(U_i) \rightrightarrows \prod_{i,j\in I} \mathcal{F}(U_i \cap U_j)\,. \tag{8.2}$$

The two morphisms on the right are induced by the restriction morphisms $\Phi^{U_i}_{U_i\cap U_j}$ and $\Phi^{U_j}_{U_i\cap U_j}$.

Definition 8.2.6 (Stalk). Consider a point $x \in X$ together with the set of all neighbourhoods of x. This set can be turned into a directed set (Definition 1.7.15) by equipping it with the (partial) order relation

$$U \geq V \implies U \subseteq V\,. \tag{8.3}$$

This turns the sheaf $\mathcal{F}$ on X into a directed system. The stalk at x is defined as the following direct limit (Definition 1.8.2):

$$\mathcal{F}_x := \varinjlim_{U\ni x} \mathcal{F}(U)\,. \tag{8.4}$$

The equivalence class of a section $s \in \mathcal{F}(U)$ in $\mathcal{F}_x$ is called the **germ** of s at x. Two sections belong to the same germ at x if there exists a neighbourhood of x on which they coincide.

Notation 8.2.7. Similar to the notation of the restriction morphisms, the morphism that maps every section to its germ at x is denoted by Φ^U_x.

Property 8.2.8. Two subsheaves of a sheaf $\mathcal{F}$ on X are equal if and only if their stalks are equal as subsets of $\mathcal{F}_x$ for all points $x \in X$. However, this does not imply that two sheaves with isomorphic stalks are equal (or even isomorphic)!

Example 8.2.9 (Global sections functor). Let X be a topological space. The global sections functor $\Gamma(X,-)$ is defined as the functor $\Gamma(X,-) : \mathbf{Sh}(X) \to \mathbf{Set} : \mathcal{F} \to \mathcal{F}(X)$.

Example 8.2.10 (Sheaf of sections). Consider a continuous function $f : X \to Y$. This function induces a sheaf on Y in the following way:

$$\mathcal{F}(U) := \{s : U \to X \mid f \circ s = \mathbb{1}_U\}\,. \tag{8.5}$$

It is the sheaf that assigns to every open set the local sections of f in the sense of Definition 3.4.1.

Construction 8.2.11 (Associated sheaf). Consider a presheaf $\mathcal{F}$ on a topological space X. From this presheaf one can construct a sheaf $\overline{\mathcal{F}}$, called the **sheafification** or associated sheaf of $\mathcal{F}$, in the following way. First, define $\overline{\mathcal{F}}$ as the presheaf

$$\overline{\mathcal{F}}(U) := \left\{ (s_x)_{x\in U} \in \prod_{x\in U} \mathcal{F}_x \,\middle|\, \forall x \in U : \exists \text{ open } V \ni x, t \in \mathcal{F}(V) : \forall v \in V : s_v = \Phi^V_v(t) \right\}. \tag{8.6}$$

Sections of this sheaf are said to be **continuous**. This statement can be made formal using the concept of an étalé space (see Construction 8.2.15 further on). The restriction maps ρ_V^U are defined as follows:

$$\rho_V^U : (s_x)_{x\in U} \mapsto (s_x)_{x\in V} \,. \tag{8.7}$$

It is easily proven that this presheaf is in fact a sheaf, the sheafification of $\mathcal{F}$. This construction also gives a canonical morphism $\varphi : \mathcal{F} \to \overline{\mathcal{F}}$ since the canonical injection

$$\varphi(s) : U \to \prod_{x\in U} \mathcal{F}_x : x \mapsto s_x = \Phi_x^U(s) \,, \tag{8.8}$$

where $s \in \mathcal{F}(U)$ and $x \in U$, takes image in $\overline{\mathcal{F}}(U)$.

Universal Property 8.2.12. Let $\mathcal{F}$ be a presheaf on X with associated sheaf $\overline{\mathcal{F}}$. Every sheaf morphism $\mathcal{F} \to \mathcal{G}$ factors uniquely through the canonical morphism $\mathcal{F} \to \overline{\mathcal{F}}$.

Property 8.2.13 (Stalks). Let $\mathcal{F}$ be a presheaf on X with associated sheaf $\overline{\mathcal{F}}$. The morphism $\varphi : \mathcal{F} \to \overline{\mathcal{F}}$ induces an isomorphism $\varphi_x : \mathcal{F}_x \to \overline{\mathcal{F}}_x$ for all $x \in X$.

Property 8.2.14. Let $\mathcal{F}$ be a sheaf on X with associated sheaf $\overline{\mathcal{F}}$ (obtained by regarding $\mathcal{F}$ as a presheaf). The morphism $\varphi : \mathcal{F} \to \overline{\mathcal{F}}$ is an isomorphism.

There exists another, more topological, construction of the associated sheaf.

Construction 8.2.15 (Étale spaces). Let $\mathcal{F}$ be a presheaf on X and consider the disjoint union

$$\mathcal{F}^* := \bigsqcup_{x\in X} \mathcal{F}_x \,. \tag{8.9}$$

Define for every local section $s \in \mathcal{F}(U)$ a function $\bar{s} : U \to \mathcal{F}^* : x \mapsto s_x \in \mathcal{F}_x$. The union $\mathcal{F}^*$ can be turned into an étale space (Definition 6.2.19) over X by equipping it with the topology with basis

$$\left\{\bar{s}(U) \mid U \text{ open in } X, s \in \mathcal{F}(U)\right\} . \tag{8.10}$$

The projection map π is given by $\pi : s_x \mapsto x$. The sheafification $\overline{\mathcal{F}}$ is isomorphic to the sheaf of sections of $\mathcal{F}^*$.

Property 8.2.16 (Paracompact spaces). Let X be a paracompact space (Definition 6.5.7) and consider a sheaf F of Abelian groups on X. For every closed subset $V \subset X$, there exists an isomorphism

$$\varinjlim_{U\supset K} F(U) \cong F(K) \,, \tag{8.11}$$

where $F(K)$ is defined as the set of sections of the restriction of the étale space F^* to K. By definition this means that every section over a closed subset can be extended to a local section over some open neighbourhood.

Definition 8.2.17 (Flabby sheaf). A sheaf F on a topological space X such that for every two open subsets $U \subseteq V \subseteq X$ the restriction morphism $F(V) \to F(U)$ is surjective.

Definition 8.2.18 (Soft sheaf). A sheaf on a topological space (often required to be paracompact Hausdorff) such that every section over a closed subset can be extended to a global section.

From the previous property and definition it is clear that (on a paracompact Hausdorff space) every flabby sheaf is soft.

The sheafification can even be constructed in a third way.

Construction 8.2.19. Consider the equalizer in Eq. (8.2). To every presheaf $\mathcal{F}$, one can assign a separated presheaf $\mathcal{F}^{\#}$ by defining $\mathcal{F}^{\#}(U)$ as the direct limit (Definition 1.8.2) of the equalizers over all open covers of U ordered by refinement. A sheaf is obtained by applying this construction a second time.

Example 8.2.20 (Constant sheaf). Consider the constant presheaf on X with target S (Example 8.1.4). The constant sheaf, denoted by $\underline{S}$ or $\flat S$, is defined as the associated sheaf of this presheaf. The stalks at every point $x \in X$ can be identified with S. The continuous sections $\underline{S}(U)$ are the locally constant functions $f : U \to S$.

8.3 Equivariance

In this section, actions of a topological group on a topological sheaf are considered (cf. Section 2.3).

Definition 8.3.1 (Equivariant sheaf). A G-equivariant sheaf on a topological G-space X is a sheaf $\mathcal{F}$ such that G acts equivariantly on the étale space $\overline{\mathcal{F}}$, i.e. it is a G-equivariant object in the category of sheaves on X.

Equivalently, consider the canonical action and projections maps $\rho, \pi : G \times X \to X$. A sheaf $\mathcal{F}$ is G-equivariant if

$$\rho^* \mathcal{F} \cong \pi^* \mathcal{F} \tag{8.12}$$

and, denoting this isomorphism by Φ, the following cocycle condition is satisfied:

$$\pi_{23}^* \Phi \circ (\mathbb{1}_G \times \rho)^* \Phi = (\mu \times \mathbb{1}_X)^* \Phi \,, \tag{8.13}$$

where $\pi_{23} : G \times G \times X \to G \times X$ is the projection on the last two factors.

Example 8.3.2 (G-sets). Consider the one-element set $*$ and let G be locally connected. Since étale spaces over $*$ are simply discrete topological spaces, the G-equivariant sheaves on $*$ correspond to G/G_0-sets (Definition 2.3.1).

Example 8.3.3 (Free actions). If G acts freely on X, G-equivariant sheaves are equivalent to sheaves on the quotient space X/G.

8.4 Abelian sheaf cohomology

In this section only Abelian sheaves will be considered, i.e. e. sheaves with values in **Ab** (unless stated otherwise).

Property 8.4.1. The global sections functor is only left exact.

Because of Chapter 5 one can now construct derived functors. These give rise to a new cohomology theory. Although it will appear a lot more abstract than the (co)homology theories from Chapter 7, these theories are in fact specific instances.

Definition 8.4.2 (Local system). A locally constant sheaf, of Abelian groups. The cohomology of a topological space with coefficients in a local system is then simply the sheaf cohomology of this sheaf.

8.4.1 Derived cohomology

Property 8.4.3 (Injective resolutions). Every Abelian sheaf admits an injective resolution or, equivalently, the category $\mathcal{AB}(X)$ of Abelian sheaves has enough injectives.

Because of Property 8.4.1, one can construct nontrivial (right) derived functors of $\Gamma(X, -)$.

Definition 8.4.4 (Sheaf cohomology group). Let $\mathcal{F}$ be a sheaf on X. Given an injective resolution I of $\mathcal{F}$ (as usual the result will be independent of the chosen resolution), the sheaf cohomology groups of $\mathcal{F}$ on X are defined as the cohomology groups of the complex

$$\cdots \longrightarrow \Gamma(X, I^i) \longrightarrow \Gamma(X, I^{i+1}) \longrightarrow \cdots . \tag{8.14}$$

The cohomology group $H^0(X; \mathcal{F})$ is equal to $\Gamma(X, \mathcal{F})$.

Definition 8.4.5 (Acyclic sheaf). A sheaf is said to be acyclic if its higher cohomology groups vanish (cf. Definition 5.4.6).

Example 8.4.6 (Soft sheaves). Soft sheaves (Definition 8.2.18) are acylic. Let $(M, \mathcal{O}_M)$ be a smooth manifold with its sheaf of smooth functions (resp. complex manifold with its sheaf of holomorphic functions). All sheaves of $\mathcal{O}_M$-modules are soft and, hence, acyclic.

The following theorem is a specific instance of Property 5.4.7.

Theorem 8.4.7 (de Rham & Weil). *There exists an isomorphism between the sheaf cohomology groups obtained using injective resolutions and the ones obtained using an acyclic resolution.*

Definition 8.4.8 (Image and kernel). Given a morphism of sheaves $\phi : \mathcal{F} \to \mathcal{G}$ on a space X one can define the kernel/image presheaves that assign to every open subset $U \subseteq X$ the image/kernel of ϕ_U.

The kernel presheaf is already a sheaf and will be denoted by $\ker(\phi)$. The sheafification of the image presheaf will be denoted by $\operatorname{im}(\phi)$. In a similar way one can also define cokernels or any other notion that makes sense in Abelian categories.

Definition 8.4.9 (Cohomology sheaves). Let $\mathcal{F}^\bullet$ be a cochain complex of sheaves on X. The cohomology sheaves $\underline{H}^i(X;\mathcal{F}^\bullet)$ are obtained by sheafifying the presheaves that assign to every open subset $U \subseteq X$ the quotient group $\ker(d^i_U)/\operatorname{im}(d^{i-1}_U)$.

8.4.2 Čech cohomology

Consider a chain complex $(A_\bullet, \partial_\bullet) \in \mathbf{Ch}(\mathcal{AB}(X))$ of Abelian sheaves on a topological space X (for simplicity assume that the complex is connective).

Definition 8.4.10 (Čech cohomology). For an open cover $\mathcal{U} = \{U_i \subseteq X\}_{i\in I}$, denote the intersection $U_{i_0} \cap \cdots \cap U_{i_k}$ by $U_{i_0\ldots i_k}$. The cochain groups are defined for all $p \in \mathbb{N}$ as:

$$C^p(\mathcal{U};A_\bullet) := \bigoplus_{\substack{p=k-n\\ i_0<\cdots<i_k}} A_n(U_{i_0\ldots i_k})\,. \tag{8.15}$$

Since the (pre)sheaf takes values in Abelian groups, one can define the subtraction of elements and, hence, the following definition of the differential makes sense:

$$\begin{aligned}(\mathrm{d}\omega)_{i_0\ldots i_{k+1}} &:= \left(\partial\omega + (-1)^k \sum_{i=0}^{k} (-1)^i A_\bullet(\iota_i)\omega\right)_{i_0\ldots i_{k+1}} \\ &= \partial\omega_{i_0\ldots i_{k+1}} + (-1)^k \sum_{j=0}^{k+1} (-1)^j\, \omega_{i_0\ldots i_{j-1} i_{j+1}\ldots i_{k+1}}\Big|_{U_{i_0\ldots i_{k+1}}}\,,\end{aligned} \tag{8.16}$$

where ι_i are the inclusion maps of the cover. The cohomology $\check{H}^\bullet(\mathcal{U};A_\bullet)$ of this complex is called the Čech cohomology of $\mathcal{U}$ with values in $A_\bullet$.

By taking the direct limit over the direct system of open covers (the partial ordering is given by refinement of covers) one can define the Čech cohomology $\check{H}^\bullet(X;A_\bullet)$ of X with values in $A_\bullet$.

Remark 8.4.11 (Hypercohomology). Two remarks should be made here. The above construction is essentially building a double complex and calculating the cohomology of the total complex. The definition Eq. (8.16) of the total differential might differ from others in the literature by a factor $(-1)^k$ as is often the case, since vertical and horizontal directions can be interchanged. Furthermore, sometimes the definition of Čech cohomology is only given with values in an Abelian group, such that the first term in Eq. (8.16) also vanishes. This case can be recovered from the above construction by considering chain complexes concentrated in a single degree. The more general case is sometimes called **hypercohomology**.

The following two properties characterize when Čech cohomology calculates the (derived) cohomology of sheaves:

Property 8.4.12. In degrees 0 and 1 one always has

$$\check{H}^{0,1}(X;\mathcal{F}) \cong H^{0,1}(X;\mathcal{F})\,. \tag{8.17}$$

For a paracompact Hausdorff space, the Čech cohomology and (derived) sheaf cohomology coincide in all degrees.

Property 8.4.13 (Leray). An open cover $\mathcal{U} = \{U_i \subset X\}_{i\in I}$ of a topological space is said to be **acyclic** with respect to a sheaf $\mathcal{F}$ if $\mathcal{F}$ is acyclic with respect to any finite subcover of $\mathcal{U}$:

$$H^{\bullet\geq 1}(U_{i_1} \cap \ldots \cap U_{i_k};\mathcal{F}) = 0 \tag{8.18}$$

for all $i_1, \ldots, i_k \in I$. If $\mathcal{U}$ is acylic with respect to $\mathcal{F}$, then

$$\check{H}^{\bullet}(\mathcal{U};\mathcal{F}) \cong \check{H}^{\bullet}(X;\mathcal{F}) \cong H^{\bullet}(X;\mathcal{F})\,. \tag{8.19}$$

Example 8.4.14 (Good covers). Consider a topological space admitting a good cover (Definition 31.1.16). The Leray theorem applies since all intersections are contractible and higher Čech cohomology vanishes on contractible spaces. Accordingly, for all topological spaces admitting a good cover (e.g. finite CW complexes or Riemannian manifolds), the Čech and singular cohomologies coincide.

8.5 Non-Abelian sheaf cohomology

8.5.1 Čech cohomology

The issue with extending Č cohomology to the non-Abelian context is that the whole definition made heavy use of operations that only exist in Abelian categories, e.g. images, kernels and addition. However, this problem can solved.

As a first step, the case of a single group G will be considered. The (would-be) differential d is defined as before (now with a multiplicative convention):

1. For every 0-cochain $\{\phi_i : U_i \to G\}_{i\in I}$: $(\mathrm{d}\phi)_{ij} := \phi_j\phi_i^{-1}$.
2. For every 1-cochain $\{\psi_{ij} : U_i \cap U_j \to G\}_{i,j\in I}$: $(\mathrm{d}\psi)_{ijk} := \psi_{ij}\psi_{ik}^{-1}\psi_{jk}$.
3. ...

But now a major issue arises. With this definition, the maps d_k are not group morphisms and the sets $\ker(\mathrm{d}_k), \mathrm{im}(d_k)$ are not groups. To make matters worse, from $k = 2$ onwards, d stops being a differential altogether.

For $k = 0, 1$, the situation can be saved. If $\mathrm{d}\phi = e$ for some 0-cochain, e being the identity element in G, then $\phi_i = \phi_j$ on $U_i \cap U_j$. So, by the sheaf condition, one obtains

a genuine function $\phi : X \to G$, i.e. $\check{H}^0(X;G) \cong C(X,G)$. For $k = 1$, neither the cocycles nor coboundaries form a group, so forming a quotient is out of the question, but the 0-cochains do act on the 1-cocycles by conjugation

$$(\phi \cdot \psi)_{ij} := \phi_j \psi_{ij} \phi_i^{-1} \,. \tag{8.20}$$

So one can take the quotient $\check{H}^1(X;G)$, as a set, of the 1-cocycles by this action and define this to be the first cohomology set. In the Abelian case, this recovers the usual construction of $\check{H}^1(X;G)$.

Remark 8.5.1. The reason why higher cohomology $\check{H}^{\geq 2}(X;G)$ can only be defined for Abelian groups, is not straightforward when defined as above. However, when reformulating cohomology in terms of mapping spaces with values in deloopings (see Chapter 12), it becomes clear why this is the case.

@@ COMPLETE (e.g. book by Deligne) @@

8.6 Ringed spaces

Definition 8.6.1 (Ringed space). A topological space equipped with a sheaf of rings.

Definition 8.6.2 (Locally ringed space). A ringed space for which the stalk at every point is a local ring (Definition 2.4.42).

Definition 8.6.3 (Quasicoherent sheaf). Let $(X, \mathcal{O}_X)$ be a ringed space. A quasicoherent sheaf $\mathcal{F}$ on X is a $\mathcal{O}_X$-module that is locally the cokernel of a morphism of free modules.

This means that $\mathcal{F}$ is locally presentable in the sense that it fits into an exact sequence of the form

$$\mathcal{O}_X^{k_i}|_{U_i} \longrightarrow \mathcal{O}_X^{l_i}|_{U_i} \longrightarrow \mathcal{F}|_{U_i} \longrightarrow 0 \tag{8.21}$$

for some open cover $\{U_i\}_{i \in I}$ of X.

A sheaf is said to be **coherent** if it is of **finite type** (or **finitely generated**), i.e. every point has an open neighbourhood U admitting a surjective morphism

$$\mathcal{O}_X^m|_U \longrightarrow \mathcal{F}|_U \tag{8.22}$$

for $m \in \mathbb{N}$ and, for every open neighbourhood V and integer $n \in \mathbb{N}$, any morphism

$$\mathcal{O}_X^n|_V \longrightarrow \mathcal{F}|_V \tag{8.23}$$

has a kernel of finite type. All coherent sheaves are, in particular, quasicoherent.

Property 8.6.4. The coherent sheaves over a ringed space $(X, \mathcal{O}_X)$ form an Abelian category (Section 4.3).

Definition 8.6.5 (Locally free sheaf). Let $(X, \mathcal{O}_X)$ be a ringed space. An $\mathcal{O}_X$-module is said to be locally free if for every point $x \in X$ there exists an open neighbourhood $U \subseteq X$ such that $\mathcal{F}|_U$ is free. If they are free of rank $n \in \mathbb{N}$, the sheaf is also said to be locally free of **rank** n.

Property 8.6.6. Locally free sheaves are quasicoherent. Moreover, locally free sheaves of finite rank are coherent if and only if the structure sheaf $\mathcal{O}_X$ is coherent. (In algebraic geometry, this will for example be the case for *locally Noetherian varieties* and, in differential geometry, this is the case for *complex manifolds*.)

Definition 8.6.7 (Invertible sheaf). A locally free sheaf $\mathcal{F}$ over a ringed space $(X, \mathcal{O}_X)$ such that there exists another locally free sheaf $\mathcal{G}$ satisfying

$$\mathcal{F} \otimes \mathcal{G} \cong \mathcal{O}_X \,. \tag{8.24}$$

Property 8.6.8. Over suitable spaces, such as *schemes* (see Chapter 10), locally free sheaves are invertible if and only if they are of rank 1.

Remark 8.6.9. Because of the previous property, many authors simply use locally free of rank 1 as the definition of invertibility. Moreover, some authors even use the stronger condition of coherency instead of local freeness. By Property 8.6.6 above, this is only equivalent in the case of *locally Noetherian schemes* or *complex manifolds*.

Definition 8.6.10 (Picard group). The Abelian group of invertible sheaves over a ringed space. More generally, one can define the Picard group in any symmetric monoidal category (Definition 4.1.11).

Chapter 9

Metric Spaces

Chapter Contents

9.1 Definition 225
9.2 Topology 226
9.3 Examples 228
9.4 Completeness 229
9.4.1 Convex metric spaces 229
9.5 Compactness 229

9.1 Definition

Definition 9.1.1 (Metric). A metric (or **distance**) on a set M is a map $d : M \times M \to \mathbb{R}^+$ that satisfies the following properties for all $x, y, z \in M$:

1. **Nondegeneracy**: $d(x, y) = 0 \iff x = y$,
2. **Symmetry**: $d(x, y) = d(y, x)$, and
3. **Triangle inequality**: $d(x, z) \leq d(x, y) + d(y, z)$.

A set M equipped with a metric d is called a **metric space**.

Definition 9.1.2 (Diameter). The diameter of a subset U of a metric space (M, d) is defined as follows:

$$\operatorname{diam}(U) := \sup_{x,y \in U} d(x, y) \,. \tag{9.1}$$

Definition 9.1.3 (Bounded). A subset $U \subseteq M$ of a metric space is said to be bounded if $\operatorname{diam}(U) < +\infty$.

Definition 9.1.4 (Isometry). An isometry or **distance-preserving function** $f : (M, d) \to (M', d')$ between metric spaces is a function satisfying

$$d'\big(f(x), f(y)\big) = d(x, y) \tag{9.2}$$

for all $x, y \in M$.

Definition 9.1.5 (Metric retraction). Let (M, d) be a metric space. A function $f : M \to M$ is said to be a retraction of metric spaces if:

- f is idempotent, and
- f is nonexpansive, i.e. the following relation holds for all $x, y \in M$:

$$d\big(f(x), f(y)\big) \leq d(x, y)\,. \tag{9.3}$$

The image of f is called a (metric) retract of M.

9.2 Topology

Various topological notions can be (re)formulated in terms of a metric. The most important ones are given below.

Definition 9.2.1 (Open ball). An open ball with radius $R > 0$ centred at a point $x_0 \in M$ is defined as the following set:

$$B(x_0, R) := \{x \in M \mid d(x, x_0) < R\}. \tag{9.4}$$

Property 9.2.2 (Metric topology). Every metric space can be turned into a topological space by taking the open balls to be a basis.

Definition 9.2.3 (Closed ball). The closed ball $\overline{B}(x_0, R)$ is defined as the closure of the open ball $B(x_0, R)$:

$$\overline{B}(x_0, R) := \{x \in M \mid d(x, x_0) \leq R\}\,. \tag{9.5}$$

Definition 9.2.4 (Packing). Let (M, d) be a metric space. A ε-packing, for $\varepsilon > 0$, is a subset $\mathcal{P} \subseteq M$ such that

$$\sup_{x,y \in \mathcal{P}} d(x, y) \leq \varepsilon\,. \tag{9.6}$$

The **packing number** $N_P(\varepsilon, M, d)$ is defined as the greatest cardinality of a ε-packing of (M, d).

Definition 9.2.5 (Metrizable space). A topological space that is homeomorphic to a metric space (M, d), i.e. a topological space for which there exists a metric $d : X \times X \to \mathbb{R}$ that induces the topology on X.

Theorem 9.2.6 (Urysohn's metrization theorem). *Every second-countable T_3-space is metrizable.*

Definition 9.2.7 (Sequential convergence). A sequence $(x_n)_{n\in\mathbb{N}}$ in a metric space (M,d) is said to converge to a point $a \in M$ if

$$\forall \varepsilon > 0 : \exists N_0 \in \mathbb{N} : \forall n \geq N_0 : d(x_n, a) < \varepsilon\,. \tag{9.7}$$

Definition 9.2.8 (Continuity). Let (M,d) and (M',d') be two metric spaces. A function $f : M \to M'$ is said to be continuous at a point $x_0 \in \mathsf{dom}(f)$ if

$$\forall \varepsilon > 0 : \exists \delta_\varepsilon > 0 : \forall x \in \mathsf{dom}(f) : d(x_0, x) < \delta_\varepsilon \implies d'\big(f(x_0), f(x)\big) < \varepsilon\,. \tag{9.8}$$

Example 9.2.9. Let (M,d) be a metric space. The distance function $d : M \times M \to \mathbb{R}$ is a continuous function.

Definition 9.2.10 (Uniform continuity). Let (M,d) and (M',d') be two metric spaces. A function $f : M \to M'$ is said to be uniformly continuous if

$$\forall \varepsilon > 0 : \exists \delta_\varepsilon : \forall x, y \in \mathsf{dom}(f) : d(x,y) < \delta_\varepsilon \implies d'\big(f(x), f(y)\big) < \varepsilon\,. \tag{9.9}$$

This is clearly a stronger notion than that of continuity since the number ε is equal for all points $y \in \mathsf{dom}(f)$.

Uniform continuity is equivalent to the following notion.

Definition 9.2.11 (Modulus of continuity). A function $f : (M,d) \to (M',d')$ of metric spaces is said to admit a modulus of continuity $\omega : \overline{\mathbb{R}^+} \to \overline{\mathbb{R}^+}$ if and only if

$$d'\big(f(x), f(y)\big) \leq \omega\big(d(x,y)\big) \tag{9.10}$$

for all $x, y \in M$. Note that this implies that ω is continuous at zero.

Two explicit examples of moduli of continuity are.

Definition 9.2.12 (Lipschitz continuity). Let (M,d) and (M',d') be two metric spaces. A function $f : M \to M'$ is said to be Lipschitz continuous if there exists a constant $C > 0$ such that

$$d'\big(f(x), f(y)\big) \leq C d(x,y) \tag{9.11}$$

for all $x, y \in M$.

Definition 9.2.13 (Hölder continuity). Let (M,d) and (M',d') be two metric spaces. A function $f : M \to M'$ is said to be γ-Hölder continuous, the constant $\gamma > 0$ is called the **Hölder exponent**, if there exists a constant $C > 0$ such that

$$d'\big(f(x), f(y)\big) \leq C d(x,y)^\gamma \tag{9.12}$$

for all $x, y \in M$.

9.3 Examples

Example 9.3.1 (Product space). Consider the Cartesian product

$$M = M_1 \times M_2 \times \cdots \times M_n \,,$$

where (M_i, d_i) is a metric space for all $i \leq n$. Equipped with the distance function

$$d(x, y) := \max_{1 \leq i \leq n} d_i(x_i, y_i) \,, \tag{9.13}$$

this set becomes a metric space.

Property 9.3.2 (Projections determine convergence). Let M be a product metric space. Consider the projections associated with the sets M_j:

$$\mathsf{pr}_j : M \to M_j : (a_1, \dots, a_n) \mapsto a_j \,. \tag{9.14}$$

A sequence in M converges if and only if every component $(\mathsf{pr}_j(x_n))_{n \in \mathbb{N}}$ converges in (M_j, d_j).

Example 9.3.3 (Supremum distance). Let $K \subset \mathbb{R}^n$ be a compact set. The following function defines a metric on $C(K, \mathbb{C})$:

$$d_\infty(f, g) := \sup_{x \in K} |f(x) - g(x)| \,. \tag{9.15}$$

Example 9.3.4 (p-metric). For every $p \geq 1$, the L^p-norm on $\mathbb{R}^n$ is defined as follows:

$$d_p(x, y) := \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p} . \tag{9.16}$$

For $p = 1$, this is sometimes called the **taxicab metric**.

Example 9.3.5 (Chebyshev distance). The Chebyshev distance is defined similarly to the supremum distance:

$$d_\infty(x, y) := \max_{1 \leq i \leq n} |x_i - y_i| \,. \tag{9.17}$$

This metric is also sometimes called the **maximum metric** or L^∞-metric.

Remark 9.3.6. The Chebyshev metric is also an example of a product metric defined on the Euclidean product space $\mathbb{R}^n$. The notation d_∞, which is also used for the supremum distance, can be justified if the space $\mathbb{R}^n$ is identified with the set of maps $\{1, \dots, n\} \to \mathbb{R}$. Another justification is the following relation:

$$d_\infty(x, y) = \lim_{p \to \infty} d_p(x, y) \,, \tag{9.18}$$

which is also the origin of the name L^∞-metric.

9.4 Completeness

Definition 9.4.1 (Cauchy sequence). A sequence $(x_n)_{n\in\mathbb{N}}$ in a metric space (M,d) is said to be Cauchy (or to have the Cauchy property) if

$$\forall \varepsilon > 0 : \exists N \in \mathbb{N} : \forall m,n \geq N : d(x_m, x_n) < \varepsilon\,. \tag{9.19}$$

A metric space (M,d) is said to satisfy the **Cauchy criterion** if a sequence converges to a point in M if and only if it is Cauchy.

Definition 9.4.2 (Complete metric space). A metric space that satisfies the Cauchy criterion.

Property 9.4.3. Subsets of metric spaces have the following properties:

- Every closed subset of a complete metric space is complete.
- Every complete subset of a metric space is closed.

9.4.1 Convex metric spaces

Definition 9.4.4 (Convex space). A metric space (M,d) with the property that for every two points $x,y \in M$ there exists a third point $z \in M$ such that

$$d(x,z) = d(x,y) + d(y,z)\,. \tag{9.20}$$

Property 9.4.5 (Convex sets). A closed subset of a Euclidean space is a convex metric space if and only if it is a convex set (Definition 16.7.1).

9.5 Compactness

See Section 6.5 for the general theory of compact spaces.

Theorem 9.5.1 (Stone). *Every metric space is paracompact.*

Definition 9.5.2 (Totally bounded). A metric space M is said to be totally bounded if it satisfies the following equivalent statements:

- For every $\varepsilon > 0$ there exists a finite cover $\mathcal{F}$ of M with $\forall F \in \mathcal{F} : \mathsf{diam}(F) \leq \varepsilon$.
- For every $\varepsilon > 0$ there exists a finite subset $E \subset M$ such that $M \subseteq \bigcup_{x\in E} B(x,\varepsilon)$.

Definition 9.5.3 (Covering number). Let (M,d) be a metric space. The covering number $N_C(\varepsilon, M, d)$ is defined as the least number of ε-balls needed to cover M. M is totally bounded if the covering number is finite for every $\varepsilon > 0$. The logarithm of the covering number is sometimes called the **metric entropy**.

Property 9.5.4 (Boundedness). Every totally bounded set is, in particular, bounded and every subset of a totally bounded set is also totally bounded. Furthermore, every totally bounded space is second-countable.

The following theorem is a generalization of the Heine–Borel theorem for Euclidean spaces $\mathbb{R}^n$.

Theorem 9.5.5 (Heine–Borel). *For a metric space M, the following statements are equivalent:*

- *M is compact.*
- *M is sequentially compact.*
- *M is complete and totally bounded.*

Theorem 9.5.6 (Heine–Cantor). *Let M, M' be two metric spaces with M compact. Every continuous function $f : M \to M'$ is also uniformly continuous.*

Definition 9.5.7 (Equicontinuity). Let X be a topological space and let M be a metric space. A collection $\mathcal{F}$ of maps $X \to M$ is said to be equicontinuous at $x \in X$ if, for all $\varepsilon \geq 0$, there exists a neighbourhood U of x such that

$$\forall f \in \mathcal{F}, \forall x' \in U : d\big(f(x), f(x')\big) \leq \varepsilon\,. \tag{9.21}$$

More generally, when both X and Y are arbitrary topological spaces, the collection $\mathcal{F}$ is said to be (topologically) equicontinuous at $x \in X$ if, for every point $y \in Y$ and every open neighbourhood W of y, there exist neighbourhoods $U \ni x, V \ni y$ such that

$$\forall f \in \mathcal{F} : f(U) \cap V \neq \emptyset \implies f(U) \subset W\,. \tag{9.22}$$

Property 9.5.8. Let $I \subseteq \mathbb{R}$ be an open interval and let $\mathcal{F}$ be a collection of differentiable functions such that $\{f'(t) \mid f \in \mathcal{F}, t \in I\}$ is bounded. Then $\mathcal{F}$ is equicontinuous.

Theorem 9.5.9 (Arzelà–Ascoli). *Let K be a compact topological space and let M be a complete metric space. The following statements are equivalent for any collection $\mathcal{F} \subseteq C(K, M)$:*

- *$\mathcal{F}$ is compact with respect to the supremum distance (Example 9.3.3).*
- *$\mathcal{F}$ is equicontinuous, closed under uniform convergence and $\{f(x) \mid f \in \mathcal{F}\}$ is totally bounded for every $x \in K$.*

Chapter 10

Algebraic Geometry

References for this chapter are Gathmann (2019); Mumford (1999). For the basics on ring theory, ideals and algebraic dependence, see Sections 2.4 and 2.5.3. The section on tropical geometry is mainly based on Maclagan (2008). In order to not confuse the letters k and K, often used for fields, with various indices and dimensions, fields will be denoted by the symbol $\mathfrak{K}$ in this chapter. Moreover, the base field $\mathfrak{K}$ will also be assumed to be algebraically closed (unless noted otherwise).

Chapter Contents

10.1 Varieties **231**
10.1.1 Topology 234
10.1.2 Varieties 236
10.1.3 Projective varieties 236
10.2 Schemes **239**
10.2.1 Morphisms 242
10.2.2 Tangent space 244
10.3 Sheaf theory ♣ **245**
10.4 Algebraic groups & Galois theory **248**
10.5 Tropical geometry **248**

10.1 Varieties

For notational simplicity and to differentiate between $\mathfrak{K}^n$ as a vector space and as a set, the notion of an affine space is introduced.

Definition 10.1.1 (Affine space). $\mathbb{A}^n$ is defined as the underlying set of the vector space $\mathfrak{K}^n$:

$$\mathbb{A}^n := \left\{(a_1, \dots, a_n) \in K^n\right\}. \tag{10.1}$$

Definition 10.1.2 (Algebraic set). Consider a finite set of polynomials in $\mathfrak{K}[x_1, \dots, x_n]$. It is not hard to show that the zero locus of these polynomials depends only on the ideal spanned by them and, hence, one can define the algebraic set associated to an ideal $I \subseteq \mathfrak{K}[x_1, \dots, x_N]$ to be

$$V(I) := \left\{(a_1, \dots, a_n) \in \mathbb{A}^n \mid \forall f \in I : f(a_1, \dots, a_k) = 0\right\}. \tag{10.2}$$

A set $S \in \mathbb{A}^n$ is said to be an **(affine)** algebraic set if there exists an ideal I such that $S = V(I)$. An algebraic set $S \in \mathbb{A}^n$ is said to be **irreducible** if it is not the union of two strictly smaller algebraic sets. Irreducible algebraic sets are also called **affine varieties**.

Remark. Some authors such as Gathmann (2019) make no distinction between general algebraic sets and affine varieties.

Property 10.1.3. By Hilbert's basis theorem 2.5.2 one can express any algebraic set as the zero locus of a finite number of polynomials.

Given an algebraic set S, the set $I(S)$ is defined as the ideal of polynomials that vanish on S. The following theorem gives an important relation between algebraic sets and ideals.

Theorem 10.1.4 (Hilbert's Nullstellensatz). *Let J be an ideal in $\mathfrak{K}[x_1, \dots, x_n]$ and let $\sqrt{J}$ denote its radical (Definition 2.4.32).*

$$I\big(V(J)\big) = \sqrt{J} \tag{10.3}$$

Similar to the case of the weak Nullstellensatz, one obtains the following result.

Corollary 10.1.5. There exists a bijection between the algebraic subsets of $\mathbb{A}^n$ and the radical ideals in $K[x_1, \dots, x_n]$. The irreducible algebraic sets correspond to the prime ideals by the Lasker–Noether decomposition theorem 2.4.34.

Definition 10.1.6 (Rational point). Let $V \subseteq \mathfrak{K}^n$ be an affine variety and let $\{f_i\}_{i \leq r}$ be a set of generating polynomials. The rational points $\mathfrak{K}(V)$ are exactly the solutions of the system of equations

$$\begin{cases} f_1(x_1, \dots, x_n) = 0 \\ \vdots \\ f_r(x_1, \dots, x_n) = 0\,. \end{cases} \tag{10.4}$$

Definition 10.1.7 (Morphism of varieties). Let $V_1 \subset \mathbb{A}^{n_1}, V_2 \subset \mathbb{A}^{n_2}$ be two affine varieties. A morphism $\varphi : V_1 \to V_2$ is a function that can be expressed in the following way:

$$\varphi(x_1, \dots, x_{n_1}) = \left(f_1(x_1, \dots, x_{n_1}), \dots, f_{n_2}(x_1, \dots, x_{n_1})\right), \tag{10.5}$$

where $f_i \in \mathfrak{K}[x_1, \dots, x_{n_1}]$ for all $i \leq n_2$.

A closely related notion is that of rational maps.

Definition 10.1.8 (Rational map). Consider two affine varieties X, Y. A rational map $f : X \to Y$ is an equivalence class of pairs (U, f_U), where U is a nonempty open subset and $f_U : U \to Y$ is a morphism of varieties, under the following relation: $(U, f_U) \sim (V, f_V)$ if and only if $f_U = f_V$ on a nonempty subset of $U \cap V$.

A rational map is said to be **dominant** if, for one of its representatives (U, f), the image $f(U)$ is dense. Dominance of rational maps assures that their composition exists and is well-defined. A rational map $f : X \to Y$ is said to be **birational** if it is dominant and admits a rational inverse.

Definition 10.1.9 (Coordinate ring). Consider the polynomial ring $\mathfrak{K}[x_1, \dots, x_n]$ and let V be an algebraic set in $\mathbb{A}^n$. The coordinate ring (or **affine ring**) of V is defined as the following quotient:

$$\Gamma(V) := \mathfrak{K}[x_1, \dots, x_n]/I(V). \tag{10.6}$$

The elements of this ring are the $\mathfrak{K}$-valued polynomials in the coordinates on V.

If V is irreducible, it follows from the Nullstellensatz that $I(V)$ is a prime ideal and, hence, that $\Gamma(V)$ is an integral domain. This property allows to construct the field of fractions (Definition 2.4.6). This field is called the **function field** of V and is denoted by $\mathfrak{K}(V)$. The elements of $\mathfrak{K}(V)$ are called the **rational functions** on V. It can be shown that the rational functions are exactly the rational maps $V \to \mathbb{A}^1$.

It should be noted that every morphism of varieties induces a $\mathfrak{K}$-morphism on the associated affine rings by precomposition. This gives rise to the following property.

Property 10.1.10 (Finitely generated algebras). Γ gives an equivalence between the category of algebraic sets and the category of finitely generated (reduced) K-algebras. This equivalence passes to an equivalence between the subcategories on affine varieties and integral domains.

Definition 10.1.11 (Dimension). The dimension of an affine variety is given by the **Krull dimension** of its coordinate ring, i.e. the maximum length of a chain of prime ideals in $\Gamma(X)$:

$$\dim(X) := \sup_{n \in \mathbb{N}} (\exists\, \mathfrak{p}_0 \subsetneq \mathfrak{p}_1 \subsetneq \cdots \subsetneq \mathfrak{p}_n). \tag{10.7}$$

By the Nullstellensatz this is equivalent to the maximum length of chains of irreducible algebraic subsets.

The **local dimension** $\dim_p(X)$ of a point $p \in X$ is defined in a similar way, with the start of the chains fixed at $\{p\}$. One then obtains

$$\dim(X) = \max_{p \in X} \dim_p(X) \,. \tag{10.8}$$

10.1.1 Topology

A topology on an affine variety can be constructed in the following way.

Definition 10.1.12 (Zariski topology). A set in $\mathbb{A}^n$ is deemed closed exactly if it is an algebraic set. A basis for this topology is given by the complements of zero loci

$$B_f = \{x \in \mathbb{A}^n \mid f(x) \neq 0\} \tag{10.9}$$

for $f \in K[x_1, \dots, x_n]$. This topology turns an affine variety into an irreducible space.

On an algebraic subset $V \subset \mathbb{A}^n$, one defines the Zariski topology as the induced topology of the one on $\mathbb{A}^n$. A basis for this induced Zariski topology is given by the sets B_f as above, but where f is now an element of $\Gamma(V)$.

The following property shows that the Zariski topology is very different from the topologies that occur in, for example, analysis.

Property 10.1.13 (Density). Any open subset of an affine variety is dense.

By dualizing, one can focus on the coordinate rings and construct varieties as a derived notion. In this approach, the main tool is the structure sheaf of a variety and, accordingly, the content of Chapter 8 on sheaf theory will, from here on, become a prerequisite.

Definition 10.1.14 (Structure sheaf). Consider an affine variety X with its associated coordinate ring $\Gamma(X)$. For any point $x \in X$, one can consider the set of functions $m_x \subset \Gamma(X)$ that vanish at x. This is a prime ideal, so one can construct a localization:

$$\mathcal{O}_x := \Gamma(X)_{m_x} = \{f/g \mid f, g \in \Gamma(X) \wedge g(x) \neq 0\} \,. \tag{10.10}$$

For every open subset $U \subset X$, one can then define the ring of functions on U as follows:

$$\mathcal{O}_X(U) := \bigcap_{x \in U} \mathcal{O}_x \,. \tag{10.11}$$

$\mathcal{O}_X$ is a sheaf with stalks given by $\mathcal{O}_x$. By Property 2.4.44, all stalks $\mathcal{O}_x$ are local rings and, hence, $(X, \mathcal{O}_X)$ is a locally ringed space (Definition 8.6.2). The residue field of these local rings is equal to the base field K. The elements of $\mathcal{O}_X(U)$ are called the **regular functions** on U.

This construction can be made more explicit. A map $\varphi : X \to K$ is said to be **regular** at a point $x \in X$ if there exists an open neighbourhood $U \ni x$ and polynomials $f, g \in \Gamma(X)$ such that $g \neq 0$ and $\varphi = f/g$ on U. As for continuous functions, the map φ is said to be regular on X if it is regular at every point $x \in X$.

Property 10.1.15. Let $f \in \Gamma(X)$ be a function on X and consider the basis set B_f, i.e. the complement of the zero locus of f. The structure sheaf assigns localizations to basis sets:

$$\mathcal{O}_X(B_f) = \Gamma(X)_f\,. \tag{10.12}$$

In particular,

$$\Gamma(X, \mathcal{O}_X) = \Gamma(X)\,, \tag{10.13}$$

where, on the left-hand side, Γ denotes the global sections functor 8.2.9. This property explains the notation $\Gamma(X)$ introduced before.

Remark 10.1.16. Both $\mathcal{O}_X(U)$ and $\mathcal{O}_x$ are subrings of the function field $K(X)$.

Alternative Definition 10.1.17 (Affine variety). A topological space X equipped with a sheaf $\mathcal{F}$ of K-valued functions such that X is isomorphic to an irreducible algebraic set Σ and such that $\mathcal{F}$ is isomorphic to the structure sheaf $\mathcal{O}_X$. An open subset of an affine variety is sometimes called a **quasi-affine variety**.

Using the notion of a regular function, the definition of a morphism of affine varieties can be restated.

Alternative Definition 10.1.18 (Morphism). A continuous function between affine varieties $f : X \to Y$ such that precomposition by f preserves regular functions.

Property 10.1.19 (Identity theorem). If two regular maps coincide on a nonempty open subset, they are equal.

Definition 10.1.20 (Generic stalk). For the construction of the stalk of the structure sheaf over a point $x \in X$, one takes a direct limit over all open sets containing x. This way, the local ring $\Gamma(X)_{m_x}$ is obtained. Moreover, this is a subring of the field of fractions $K(X)$ of $\Gamma(X)$. Now, using a similar definition, one can recover all of $K(X)$.

Instead of taking a direct limit over the open sets containing a certain point $x \in X$, take a direct limit over all open sets in X:

$$\mathcal{O}_{\tilde{x}} := \varinjlim_{U \subset X} \mathcal{O}_X(U)\,. \tag{10.14}$$

This stalk is called the generic stalk of X and it is isomorphic to $K(X)$.

Definition 10.1.21 (Étale morphism over algebraically closed fields). A morphism $f : X \to Y$, where both X and Y are affine varieties over an algebraically closed field, is said to be formally étale if the induced maps $\widehat{\mathcal{O}}_{f(x)} \to \widehat{\mathcal{O}}_x$ on the completion (Definition 2.4.50) of local rings are isomorphisms.

10.1.2 Varieties

Definition 10.1.22 (Prevariety). A connected locally ringed space that admits a finite cover by affine spaces.

Remark 10.1.23. It can be shown that every prevariety X is irreducible and, hence, the open sets form a direct system. This way, one can define the **generic stalk** of an arbitrary sheaf $\mathcal{F}$, as in the case of affine varieties. For the structure sheaf $\mathcal{O}_X$, this generic stalk is again called the **function field** $K(X)$. It coincides with the function field of every open affine subset of X.

Construction 10.1.24 (Gluing). Consider two prevarieties X, Y together with an isomorphism $f : U \cong W$ between open subsets $U \subset X, V \subset Y$. The prevarieties can be glued together along f as follows. First, build the attaching space (Construction 6.3.9) $U \cup_f Y$ with its canonical topology. Then, define the regular functions on a subset to be those that come from regular functions on (subsets of) X and Y.

Definition 10.1.25 (Variety[1]). A prevariety X for which the diagonal Δ_X is closed in $X \times X$. It should be noted that every affine variety is a variety, but not the other way around.

Remark 10.1.26. The motivation for this definition is Definition 6.1.31. In general topology it is well-known that a lot of pathological spaces can be excluded by restricting to Hausdorff spaces, i.e. spaces where distinct points admit disjoint neighbourhoods. Because open subsets of irreducible spaces have nonempty intersections, this property is sadly enough not very useful in the study of varieties. However, the equivalent definition using closedness of the diagonal remains useful if one does not consider the product topology on $X \times X$, but instead uses the 'gluing'-topology from Construction 10.1.24 above.

The following two closure properties are very important.

Property 10.1.27. Consider a prevariety morphism $f : X \to Y$, where Y is a variety. The graph of f is closed in $X \times Y$.

Property 10.1.28. Consider two prevariety morphisms $f, g : X \to Y$ where Y is a variety. The set on which f and g coincide is closed in X.

10.1.3 Projective varieties

Definition 10.1.29 (Projective space). Consider the vector space $\mathfrak{K}^n$. The projective space $\mathbb{P}_{n-1}(\mathfrak{K})$ or $\mathfrak{K}\mathbb{P}^{n-1}$ is defined as the quotient of $\mathfrak{K}^n$ under the following equivalence relation:

$$(x_1, \dots, x_n) \sim (y_1, \dots, y_n) \iff \exists \lambda \in \mathfrak{K}^\times : \forall i \leq n : x_i = \lambda y_i \,. \tag{10.15}$$

[1]Sometimes also called a **separated prevariety**.

The equivalence class of a vector $(x_1, \dots, x_n)$ is denoted by $[x_1 : \cdots : x_n]$. Because the numbers characterizing an equivalence class are only determined up to a common factor, the coordinates $x_1, \dots, x_n$ are called **homogeneous coordinates**.

Consider the subset

$$\mathfrak{K}_{\text{hom}}[x_0, \dots, x_n] \subset \mathfrak{K}[x_0, \dots, x_n] \tag{10.16}$$

consisting of all homogeneous polynomials. The definition of projective space implies that

$$\forall f \in \mathfrak{K}_{\text{hom}}[x_0, \dots, x_n] : f(\lambda x_0, \dots, \lambda x_n) = 0 \iff f(x_0, \dots, x_n) = 0 \tag{10.17}$$

and, hence, that the zero loci of homogeneous polynomials are well-defined subsets of the projective space $\mathbb{P}_n(\mathfrak{K})$.

Definition 10.1.30 (Projective algebraic set). As in the case of affine algebraic sets one can define two operations. Let I be a homogeneous ideal, i.e. an ideal in $\mathfrak{K}[x_0, \dots, x_n]$ that is generated by homogeneous polynomials. The projective algebraic set $V_{\mathbb{P}}(I)$ is defined as the zero locus of I:

$$V_{\mathbb{P}}(I) := \left\{ x \in \mathbb{P}_n(\mathfrak{K}) \mid \forall f \in I : f(x) = 0 \right\}. \tag{10.18}$$

Given a projective algebraic set $V \in \mathbb{P}_n(\mathfrak{K})$, one can define the ideal $I_{\mathbb{P}}(V)$ as follows:

$$I_{\mathbb{P}}(V) := \left(f \in \mathfrak{K}_{\text{hom}}[x_0, \dots, x_n] \mid \forall x \in V : f(x) = 0 \right), \tag{10.19}$$

i.e. the ideal $I_{\mathbb{P}}(V)$ is generated by all homogeneous polynomials that vanish on V. The **Zariski topology** on $\mathbb{P}_n(\mathfrak{K})$ is defined such that the closed sets are exactly the projective algebraic sets.

Theorem 10.1.31 (Projective Nullstellensatz). *For all homogeneous ideals I, except $I_0 := (x_1, \dots, x_n)$, one has:*

$$I_{\mathbb{P}}\left(V_{\mathbb{P}}(I)\right) = \sqrt{I}. \tag{10.20}$$

Corollary 10.1.32. As before, this implies that there exists a bijection between the projective algebraic sets in $\mathbb{P}_n(\mathfrak{K})$ and the homogeneous radical ideals (except for I_0) in $\mathfrak{K}[x_0, \dots, x_n]$.

Definition 10.1.33 (Coordinate ring). As for affine algebraic sets, the coordinate ring of a projective algebraic set V is defined as the quotient

$$\Gamma(V) := \mathfrak{K}[x_0, \dots, x_n]/I_{\mathbb{P}}(V). \tag{10.21}$$

The construction of regular functions on affine varieties (Definition 10.1.14) cannot be extended to projective spaces in a straightforward manner. Consider for example a

polynomial $f \in \mathfrak{K}[x_0, \dots, x_n]$. This polynomial does not form a well-defined function on a projective algebraic set $V_{\mathbb{P}}(I) \subset \mathbb{P}_n(\mathfrak{K})$ even if f is homogeneous, since changing the homogeneous coordinates on $V_{\mathbb{P}}(I)$ changes the value of f (only the zero locus is invariant). However, the ratio of two homogeneous polynomials of the same degree does form a well-defined function on $V_{\mathbb{P}}(I)$.

Since the ideal I is homogeneous, the quotient $R := \mathfrak{K}[x_0, \dots, x_n]/I$ is a graded algebra. Denote by $\mathfrak{K}(X)$ the zeroth order part of the localization of R by the homogeneous elements:

$$\mathfrak{K}(X) := \{f/g \mid \exists n \in \mathbb{N} : f, g \in R_n\}. \tag{10.22}$$

Now, although an element $f \in R_n$ does not give a well-defined function on X, the property $f(x) \neq 0$ is preserved under rescaling. Hence, one can define a ring $\mathcal{O}_x$ as before:

$$\mathcal{O}_x := \{f/g \in \mathfrak{K}(X) \mid g(x) \neq 0\}. \tag{10.23}$$

This ring has a maximal ideal $I_x = \{f/g \in \mathfrak{K}(X) \mid f(x) = 0, g(x) \neq 0\}$ such that all elements in $\mathcal{O}_x$ are invertible and, by Property 2.4.43; $\mathcal{O}_x$ is a local ring. One can then construct a sheaf $\mathcal{O}_X$ using the same procedure as for affine varieties to turn a projective space into a locally ringed space:

$$\mathcal{O}_X(U) = \bigcap_{x \in U} \mathcal{O}_x. \tag{10.24}$$

Property 10.1.34 (Variety). For every projective variety $X \subset \mathbb{P}_n(\mathfrak{K})$ the pair $(X, \mathcal{O}_X)$ is locally isomorphic to an affine variety and as such every projective variety is in particular a variety in the sense of Definition 10.1.25.

Property 10.1.35 ($\mathbb{A}^n$ in $\mathbb{P}_n(\mathfrak{K})$). Consider the affine variety $\mathbb{A}^n$. This set admits a bijective mapping onto an open subset of $\mathbb{P}_n(\mathfrak{K})$ as follows:

$$\varphi : \mathbb{A}^n \to U_0 : (x_1, \dots, x_{,}n) \mapsto [1 : x_1 : \cdots : x_n]. \tag{10.25}$$

It can be shown that this map is a homeomorphism if both spaces are equipped with the Zariski topology.

Property 10.1.36 (Schubert decomposition). The projective space $\mathbb{P}_n(\mathfrak{K})$ admits a decomposition of the form

$$\mathbb{P}_n(\mathfrak{K}) = \bigcup_{i=0}^{n} \mathfrak{K}^i, \tag{10.26}$$

where the union is set-theoretic. However, one can refine this to a statement in topology. The projective space $\mathbb{P}_n(\mathfrak{K})$ admits the structure of a CW complex with one k-cell in every dimension (namely $\mathbb{A}^k$). These cells are called **Bruhat** or **Schubert cells**. (The precise distinction is of no relevance here.)

Example 10.1.37 (Finite fields). Consider a *finite field* $\mathbb{F}_q$ (see Definition 11.1.1). Using the above decomposition one can easily compute the cardinality of $\mathbb{P}_n(\mathbb{F}_q)$:

$$|\mathbb{P}_n(\mathbb{F}_q)| = \sum_{i=0}^{n} |\mathbb{F}_q^i| = \sum_{i=0}^{n} q^i \equiv [n+1]_q \,, \tag{10.27}$$

For example, the **Fano plane** $\mathbb{P}_2(\mathbb{F}_2)$ has cardinality 7.

Construction 10.1.38 (Blow-up). Consider an algebraic set $X \subseteq \mathbb{A}^n$ together with a set of regular functions $\{f_1, \dots, f_k\} \subset \Gamma(X)$. Define the subset $Y := X \backslash V(f_1, \dots, f_k)$. By definition these functions do not all vanish simultaneously on Y and, hence, there exists a well-defined map

$$f : Y \to \mathbb{P}_{k-1}(\mathfrak{K}) : x \mapsto \left[f_1(x) : \cdots : f_k(x)\right] . \tag{10.28}$$

The graph of this morphism is closed in $Y \times \mathbb{P}_{k-1}(\mathfrak{K})$ by Property 10.1.27, but not in $X \times \mathbb{P}_{k-1}(\mathfrak{K})$. Its closure in the latter is called the blow-up $\widetilde{X}$ of X at $f_1, \dots, f_k$. The projection map $\pi : \widetilde{X} \to X$ is sometimes also called the blow-up (map). The graph Γ_f is isomorphic to Y and its complement $\pi^{-1}(V(f_1, \dots, f_k))$ in $\widetilde{X}$ is called the **exceptional set** (of the blow-up).

Construction 10.1.39 (Explicit description). Consider an algebraic set $X \subseteq \mathbb{A}^n$ together with its blow-up $\widetilde{X}$ at $\{f_1, \dots, f_k\}$. One can prove that the following inclusion holds:

$$\widetilde{X} \subseteq \left\{(x, y) \in X \times \mathbb{P}_{n-1}(\mathfrak{K}) \mid \forall i, j \leq n : y_i f_j(x) = y_j f_i(x)\right\} . \tag{10.29}$$

In the case of $X = \mathbb{A}^n$ and $f_i(x) = x_i$ one can even prove that this inclusion is an equality. Since the zero locus of the coordinate functions is $\{0\}$, one finds that the exceptional set of this blow-up is exactly $\mathbb{P}_{n-1}(\mathfrak{K})$.

Property 10.1.40. If X is irreducible, there exists a birational morphism $X \to \widetilde{X}$.

10.2 Schemes

Many of the definitions in this section will be dual to the ones in Section 2.4.

Definition 10.2.1 (Spectrum). Let R be a commutative ring. The spectrum $\mathsf{Spec}(R)$ is defined as the set of prime ideals of R. This set can be turned into a topological space by equipping it with the **Zariski topology** whose closed subsets are of the form

$$V(I) := \{P \subseteq I \mid I \text{ is an ideal}, P \in \mathsf{Spec}(R)\} . \tag{10.30}$$

A basis for this topology is given by the sets

$$D_f := \{P \not\ni f \mid f \in R, P \in \mathsf{Spec}(R)\} . \tag{10.31}$$

Property 10.2.2. $\mathsf{Spec}(R)$ is a compact T_0-space.

Definition 10.2.3 (Structure sheaf). Given a spectrum $X = \mathsf{Spec}(R)$, one can define a sheaf[2] $\mathcal{O}_X$ by setting $\forall f \in R : \mathcal{O}_X(D_f) := R_f$, where R_f is the localization of R with respect to the monoid of powers of f.

Property 10.2.4. The spectrum $\mathsf{Spec}(R)$, together with its structure sheaf, forms a ringed space.

Definition 10.2.5 (Affine scheme). A locally ringed space isomorphic to the spectrum of a commutative ring.

Example 10.2.6 (Spectrum of a field). For any field $\mathfrak{K}$, its spectrum $\mathsf{Spec}(\mathfrak{K})$ consists of a single point with $\mathfrak{K}$ as the stalk over it.

Property 10.2.7 (Stalks). Consider an affine scheme $\mathsf{Spec}(R)$ and a point $x \in \mathsf{Spec}(R)$ corresponding to the prime ideal $\mathfrak{p}$. The stalk at x is given by the localization $R_{\mathfrak{p}}$.

Definition 10.2.8 (Scheme). A locally ringed space such that for every point there exists an open neighbourhood isomorphic to an affine scheme. A scheme X **over a ring** R means a morphism $X \to \mathsf{Spec}(R)$.

Property 10.2.9 (Integers). $\mathsf{Spec}(\mathbb{Z})$ is the terminal scheme, i.e. for every scheme X there exists only one morphism $X \to \mathsf{Spec}(\mathbb{Z})$. This spectrum itself consists of all the prime numbers, together with a generic point at infinity. The stalks over these elements are given by $\mathbb{Z}_p$ and $\mathbb{Q}$, respectively.

The following definition extends Definition 10.1.6 to schemes.

Definition 10.2.10 (Rational point). Let $\mathfrak{X} : X \to \mathsf{Spec}(\mathfrak{K})$ be a scheme over a field $\mathfrak{K}$. The rational points $X(\mathfrak{K})$ are the sections of $\mathfrak{X}$, i.e. the morphisms $p : \mathsf{Spec}(\mathfrak{K}) \to X$ such that $\mathfrak{X} \circ p = \mathbb{1}_{\mathsf{Spec}(\mathfrak{K})}$.

Definition 10.2.11 (Geometric point). A geometric point of a scheme X over a field K is a morphism $\mathsf{Spec}\left(\overline{\mathfrak{K}}\right) \to X$, where $\overline{\mathfrak{K}}$ is an algebraic closure of $\mathfrak{K}$ (Definition 2.5.26).

Property 10.2.12 (Affine schemes). There exists an adjunction

$$\mathbf{CRing}^{\mathrm{op}} \underset{\mathsf{Spec}}{\overset{\Gamma}{\underset{\longrightarrow}{\overset{\longleftarrow}{\perp}}}} \mathbf{Sch}\,. \tag{10.32}$$

By Property 3.2.31 this becomes an (adjoint) equivalence between $\mathbf{AffSch}$ and $\mathbf{CRing}^{\mathrm{op}}$, i.e. a scheme X is affine if and only if the unit $X \to \mathsf{Spec}\big(\Gamma(X, \mathcal{O}_X)\big)$ is an isomorphism or, equivalently, if the map

$$\mathbf{Sch}(Y, X) \to \mathbf{CRing}\big(\Gamma(X, \mathcal{O}_X), \Gamma(Y, \mathcal{O}_Y)\big) \tag{10.33}$$

[2] In fact this is merely a *B-sheaf* as it is only defined on the basis of the topology. However, every *B-sheaf* can be extended to a sheaf by taking appropriate limits (Mumford, 1999).

is bijective for all schemes Y.

Property 10.2.13 (Functor of points). The functor of points of a scheme X is simply the functor represented by X. By the Yoneda lemma, morphisms between such functors, i.e. morphisms between (generalized) points, are equivalent to morphisms between schemes. Moreover, similar to the case of manifolds (this will be explained in Chapter 44), restricting representable functors to the subcategory of affine schemes loses no information, i.e. schemes are characterized by morphisms from affine schemes.

Property 10.2.14 (Sober). Every scheme is sober (Definition 6.8.4).

Definition 10.2.15 (Integral scheme). A scheme X such that for every open affine subset $\mathsf{Spec}(R) \subseteq X$ the ring R is an integral domain (Definition 2.4.5).

Property 10.2.16. By the above property, every integral scheme has a unique generic point.

Definition 10.2.17 (Reduced scheme). A scheme for which all stalks of the structure sheaf are reduced.

Property 10.2.18. A scheme is integral if and only if it is irreducible and reduced.

Definition 10.2.19 (Rational function). In Definition 10.1.9, the field of rational functions, the function field, was constructed as the field of fractions of the ring of global sections. For an integral scheme, the rings of local sections are again integral domains and, hence, one can again construct these fields. Moreover, in an integral scheme all open affine subsets are dense and, hence, one can define the function field as $\mathfrak{K}(X) := \mathsf{Frac}(R)$ for any affine open subset $\mathsf{Spec}(R) \subseteq X$. On the other hand, because integral schemes have a unique generic point, the associated prime ideal in any affine neighbourhood is the zero ideal and, consequently, the associated stalk is isomorphic to the function field.

Definition 10.2.20 (Group scheme). A group object in **Sch**.

Example 10.2.21 (Additive group). The additive group $\mathbb{G}_a$ (over some field $\mathfrak{K}$) has as underlying object the affine line $\mathsf{Spec}(\mathfrak{K}[x])$ and as operations:

1. **Addition**: Dual of the ring morphism $\mathfrak{K}[x] \to \mathfrak{K}[x_1] \otimes \mathfrak{K}[x_2] : x \mapsto x_1 + x_2$.
2. **Unit**: Dual of $\mathfrak{K}[x] \to \mathbb{Z} : x \mapsto 0$.
3. **Inversion**: Dual of $\mathfrak{K}[x] \to \mathfrak{K}[x] : x \mapsto -x$.

Equivalently, it is the **Ab**-valued functor that sends a scheme to the Abelian group of global sections of its structure sheaf.

10.2.1 Morphisms

Definition 10.2.22 (Open immersion). An open immersion $f : X \to Y$ of locally ringed spaces is a topological embedding such that the induced map $\mathcal{O}_{f(x)} \to \mathcal{O}_x$ is an isomorphism.

Definition 10.2.23 (Quasicompact morphism). A morphism $f : X \to Y$ of schemes such that Y admits an affine open cover whose preimages are compact[3]. Equivalently, a morphism is quasicompact if the preimage of every compact open subset is compact.

Property 10.2.24. A scheme is compact if and only if it is the union of a finite number of open affine subschemes.

The following definition is analogous to the characterization of Hausdorff spaces (Definition 6.1.31).

Definition 10.2.25 (Quasiseparated morphism). A morphism $f : X \to Y$ such that the diagonal morphism $\Delta_f : X \to X \times_f X$ is quasicompact. In the case $Y = \mathsf{Spec}(\mathbb{Z})$, where the diagonal morphism is truly the diagonal map $x \mapsto (x, x)$, X is called a **quasiseparated scheme**. If 'quasicompact' is replaced by 'closed', the morphism is said to be **separated**.

Remark 10.2.26. In the early days of scheme theory, what is here called a scheme was known as a 'prescheme'. Proper 'schemes' were what are now called separated schemes. Note that Mumford (1999) still uses this terminology.

Definition 10.2.27 (Finite type). A scheme morphism $f : X \to Y$ is said to be **locally of finite type** if Y admits an affine open cover $\{\mathsf{Spec}(B_i)\}_{i \in I}$ such that every preimage has an affine open cover $\{\mathsf{Spec}(A_{ij})\}_{j \in J_i}$ where all induced morphisms $B_i \to A_{ij}$ are of finite type. It is said to be of finite type if the index sets J_i are finite for all $i \in I$.

Definition 10.2.28 (Finite presentation). A scheme morphism $f : X \to Y$ is said to be of **finitely presented at a point** $x \in X$ if there exists an affine open set $\mathsf{Spec}(A) \cong U \ni x$ and an affine open set $\mathsf{Spec}(B) \cong V \subset Y$, with $f(U) \subseteq V$, such that the induced ring morphism $B \to A$ is finitely presented (Definition 2.4.46). The morphism $f : X \to Y$ is said to be **locally finitely presented** if it is finitely presented at every point $x \in X$ and it is said to be finitely presented if it is locally finitely presented, quasicompact and quasiseparated.

Property 10.2.29. For morphisms with locally Noetherian codomain, morphisms of (locally) finite type and (locally) finite presentation coincide.

Definition 10.2.30 (Flat morphism). A morphism of schemes such that the induced morphism on local rings is flat. If it is also surjective, it is said to be **faithfully flat**.

[3] In algebraic geometry, compact topological spaces that are not Hausdorff are often called **quasicompact**. However, this convention is not adopted here.

Definition 10.2.31 (Étale morphism). In Definition 10.1.21, étale morphisms were introduced for affine varieties over algebraically closed fields. However, for more general fields or schemes, this definition does not give the right results. A better approach is to dualize Definition 2.4.48, i.e. a scheme morphism $f : X \to Y$ is said to be **formally étale** if for every commutative ring R and nilpotent ideal $I \subset R$ the following lifting exists and is unique:

$$\begin{tikzcd} \mathsf{Spec}(R/I) \arrow[r] \arrow[d] & X \arrow[d, "f"] \\ \mathsf{Spec}(R) \arrow[r] \arrow[ur, dashed] & Y \end{tikzcd}$$

In other words, formally étale morphisms have the unique right lifting property with respect to infinitesimal thickenings. Analogously to the case of rings, the morphism f is said to be **formally smooth** (resp. **formally unramified**) if there is at least (resp. at most) one lift.

Since this definition does allow for some pathological examples, étale morphisms are those formally étale momorphisms that are locally of finite presentation.

Remark. The term 'formally' should be interpreted in the sense of *formal geometry* (see Chapter 44). There, infinitesimal neighbourhoods of points (see 44.4.31), used for defining the local geometry, are induced by nilpotent ideals. This is further exemplified by Property 10.2.45.

Remark 10.2.32. As a side note, the general definition of étale morphism is related to Definition 10.1.21. The completion of the local ring $\mathcal{O}_x$ is defined by the inverse limit over the filtration $\mathcal{O}_x/\mathfrak{m}_x \leftarrow \mathcal{O}_x/\mathfrak{m}_x^2 \leftarrow \mathcal{O}_x/\mathfrak{m}_x^3 \leftarrow \cdots$. Hence, a morphism $\widehat{\mathcal{O}}_x \to \widehat{\mathcal{O}}_y$ between completions is characterized by consecutive lifting problems that are exactly the content of Definition 2.4.48, since the ideals $\mathfrak{m}_x$ are clearly nilpotent in rings of the form $\mathcal{O}_x/\mathfrak{m}_x^k$ for any $k \in \mathbb{N}_0$.

This is generalized to the schemes as follows.

Property 10.2.33. If Y is locally Noetherian and $f : X \to Y$ is locally of finite presentation, étale morphisms can also be characterized in terms of completions. f is étale if and only if the induced morphism on completions is formally étale in the adic topology.

Moreover, if the induced morphisms of residue fields are isomorphisms, f is étale if and only the induced morphisms of completions are isomorphisms (as in the case of affine varieties over algebraically closed fields).

Definition 10.2.34 (Finite étale morphism). A morphism of schemes $f : X \to Y$ for which there exists a covering $\{U_i\}_{i\in I}$ of Y by affine sets $U_i \equiv \mathsf{Spec}(A_i)$ such that $f^{-1}(U_i) \cong \mathsf{Spec}(B_i)$ is a free, separable A_i-algebra B_i for all $i \in I$.

Remark 10.2.35 (Finite étale covering). When a finite étale morphism $f : X \to Y$ exists, X is sometimes called a **finite étale cover(ing)** of Y. The full subcategory of the slice category $\mathsf{Sch}/_Y$ on finite étale coverings is often denoted by FEt_X.

Definition 10.2.36 (Étale fundamental group). Consider a connected scheme X. The (étale) fundamental group $\pi_1(X)$ is the unique (up to isomorphism) profinite group 6.8.7 such that FEt_X is equivalent to the category $\mathsf{Fin}\,\pi_1(X)\mathsf{Set}$ of finite of $\pi_1(X)$-sets.

Remark 10.2.37. Recall Corollary 7.1.24 for topological spaces. There, a stricter set of spaces was considered with the benefit that all $\pi_1(X)$-sets could be classified. By passing to profinite groups one can formulate a classification statement in the topological setting that is virtually identical to the above one.

10.2.2 Tangent space

Definition 10.2.38 (Tangent cone). Consider an affine variety $X = V(I)$. The tangent cone to X at the origin is defined as the zero locus of the 'initial ideal' of I:

$$C_0X := V\big(\{f^{\text{in}} \mid f \in I\}\big)\,, \tag{10.34}$$

where f^{in} denotes the **initial part** of f, i.e. the sum of the smallest degree monomials in f.

Definition 10.2.39 (Tangent space). Consider an affine variety X and choose a point $x \in X$. By working in a suitable affine chart, one can assume that $x = 0$. This implies that any polynomial $f \in I(X)$ has a vanishing constant term. The tangent space at x is defined as follows:

$$T_xX := V\big(\{f^{[1]} \mid f \in I(X)\}\big)\,, \tag{10.35}$$

where $f^{[1]}$ denotes the linear part of the polynomial f.

Property 10.2.40. For $x = 0$, one obtains that $I(0) = (x_1, \dots, x_n)/I(X)$. Moreover, there exists a natural isomorphism

$$I(0)/I(0)^2 \cong \mathsf{Hom}_K(T_0X, \mathfrak{K})\,. \tag{10.36}$$

The tangent space at 0 can accordingly also be obtained as the dual of $I(0)/I(0)^2$.

It is not so hard to prove that this property can in fact easily be transported to arbitrary points $x \in X$ if one replaces the ideal $I(0)$ by the maximal ideal of the structure sheaf $\mathcal{O}_X$ at x. Therefore, one can give the following general definition.

Definition 10.2.41 (Zariski tangent space). Consider a scheme X with structure sheaf $\mathcal{O}_X$. At every point $x \in X$, the stalk $\mathcal{O}_x$ is a local ring and, hence, one obtain a maximal ideal $\mathfrak{m}_x$. The quotient $\mathfrak{m}_x/\mathfrak{m}_x^2$ is a vector space over the residue field $\mathcal{O}_x/\mathfrak{m}_x$. It is called the Zariski cotangent space at $x \in X$. Its algebraic dual is called the Zariski tangent space at $x \in X$.

Property 10.2.42 (Inverse function theorem). In the smooth setting (see Chapter 31 and onwards), the inverse function theorem states that a tuple of functions forms a local coordinate system at a point if their *differentials* generate the *cotangent space* at that point.

Here, the ideal $\mathfrak{m}_x$ consists of the functions vanishing at $x \in X$ and the cotangent space is given by $\mathfrak{m}_x/\mathfrak{m}_x^2$. By Nakayama's lemma 2.4.66, a tuple of elements represent 'coordinate functions', i.e. generate $\mathfrak{m}_x$, if their 'differential', i.e. their residues in the Zariski cotangent space, generate the cotangent space.

Definition 10.2.43 (Regular scheme). A (Noetherian) scheme is said to be regular or **nonsingular** if the local dimension at any point is equal to the dimension of the Zariski tangent space at that point. Equivalently, an affine variety is smooth if at every point the tangent space is isomorphic to the tangent cone.

Alternative Definition 10.2.44 (Tangent cone: Schemes). Assume that X is a Noetherian scheme. The tangent cone $C_x X$ is given by the spectrum of the graded ring $\mathsf{Gr}(\mathcal{O}_x) = \bigoplus_{k\in\mathbb{N}} \mathfrak{m}_x^k/\mathfrak{m}_x^{k+1}$. If X is regular, the tangent cone coincides with the Zariski tangent space.

Recall Definition 10.1.21 of étale morphisms of affine varieties (over algebraically closed fields). The following property relates these to the infinitesimal thickening point-of-view.

Property 10.2.45. A morphism $f : X \to Y$ of varieties induces an isomorphism on completions $\widehat{\mathcal{O}}_{f(x)} \to \widehat{\mathcal{O}}_x$ if and only if it induces an isomorphism on tangent cones.

10.3 Sheaf theory ♣

This section also uses the content of Chapter 12.

Property 10.3.1 (Coherent sheaves). A sheaf on a locally Noetherian scheme is coherent (Definition 8.6.3) if and only if it is quasicoherent and of finite type.

Definition 10.3.2 (Zariski topology). The Zariski topology on the category of schemes is generated by sets $\{f_i : U_i \to X\}_{i\in I}$ of open immersions such that $X = \bigcup_{i\in I} f_i(U_i)$ in the set-theoretic sense. This latter condition is often called **joint surjectivity**. This gives the **big Zariski site** on an a scheme X. The **small Zariski site** of a scheme X is the subsite of $\mathbf{Sch}_{/X}$ on the open immersions.

Remark 10.3.3 (Size issues). There are some size issues with the category of schemes, it being a big category. The result is that in general there exists no sheafification functor, i.e. no general construction of the category of sheaves and, hence, this category can only be turned into a site, not a (Grothendieck) topos.[4] In the case of *topological* or

[4]One can actually also work with large sites to obtain topoi as long as they contain a *small topologically*

smooth manifolds, this problem is often ignored since one has a (small) dense subsite and no issues arise. For schemes, however, the situation is different. For example, the fpqc topology (to be introduced below) does not admit a sheafification functor. Two solutions exists in such a context, either one works within a suitable universe or one only considers a single (pre)sheaf at a time.

Property 10.3.4. The Zariski site is subcanonical.

The following topology is finer than the Zariski topology.[5]

Definition 10.3.5 (Étale cover). An étale cover of a scheme X is a jointly surjective family of étale morphisms that are locally of finite type[6]. The **big étale site** $\mathbf{Sch}_{/X,\text{ét}}$ is the slice category of X equipped with the coverage (see Definition 12.4.8) of étale covers. The **small étale site** is the subsite on étale morphisms.

Property 10.3.6. A presheaf satisfies descent with respect to the étale topology if and only if it satisfies descent with respect to the following covers:

- Zariski covers, and
- (single) faithully flat morphisms of affine schemes.

Definition 10.3.7 (Flat topologies). The finest topologies on the category of schemes (or slices thereof) are given by considering flat morphisms. The two most widely used flat topologies are the 'fppf-' and 'fpqc-topologies'. The **fppf-topology** (for *fidèlement plat de presentation finie*) is generated by jointly surjective families of flat morphisms that are locally of finite presentation. The **fpqc-topology** (for *fidèlement plat et quasi-compacte*) is generated by jointly surjective families of flat and quasicompact morphisms.

Remark 10.3.8. With the definition as above the fpqc-topology is technically not comparable to the Zariski topology since open immersions need not be quasicompact (unless one restricts to Noetherian schemes). A solution introduced by *Vistoli* and *Kleiman* is to replace quasicompactness by the following seemingly similar definition: There exists an open affine cover $\{U_i\}_{i\in I}$ of X such that each U_i is the image of a compact open subscheme. (If one drops the modifier 'locally' in the definition of the fppf-topology, a similar problem arises.)

Remark 10.3.9 (Standard covers). An equivalent approach to the definition of flat topologies is by first defining 'standard' fppf or fpqc covers on affine schemes. These are similar to the flat covers defined above, but where the families are finite and the codomains affine. The covers of general schemes are then generated by base change.

generating set or, even more general, satisfy the *WISC axiom*.

[5]There exists another topology in between the Zariski and étale topologies, the *Nisnevich topology*, but this is of no interest here.

[6]For X locally Noetherian this additional condition is superfluous.

By interpreting objects in the big topos (over a scheme) as generalized spaces in the sense of Chapter 44, the following definition is obtained.

Definition 10.3.10 (Algebraic space). An object $X \in \mathbf{Sh}(\mathbf{Sch}_{\text{fppf}})$ satisfying

1. The diagonal morphism $X \to X \times X$ is representable (see Property 12.4.29).
2. There exists a scheme morphism $Y \to X$ that surjective and étale. This is sometimes called an **atlas**.

Requiring representability of the diagonal can be argued as follows. The fibre product can be seen as a generalized notion of intersection, so representability means that the intersection of any two schemes is again a scheme. Moreover, representability of the diagonal also implies representability of the atlas morphism.

Definition 10.3.11 (Local property). Consider some property of scheme morphisms P, e.g. étale or smoothness. Relative to some topology on $\mathbf{Sch}$, the property P is said to be **local on the source** if for every cover $\{X_i \to X\}_{i \in I}$ and every morphism $f : X \to Y$ the following holds:

$$f \text{ satisfies } P \iff \forall i \in I : X_i \to Y \text{ satisfies } P\,. \tag{10.37}$$

Analogously, the property P is said to be **local on the base** or **local on the target** if for every cover $\{Y_i \to Y\}_{i \in I}$ and every morphism $f : X \to Y$ the following holds:

$$f \text{ satisfies } P \iff \forall i \in I : Y_i \times_Y X \to Y_i \text{ satisfies } P\,. \tag{10.38}$$

Inspired by locality, one can generalize many properties of scheme morphisms to morphisms of algebraic spaces.

Definition 10.3.12 (Property of algebraic spaces). Consider a representable morphism of algebraic spaces $F : X \to Y$. If P is a property of scheme morphisms satisfying:

1. it is preserved under base change, and
2. it is fppf-local on the target,

then f is said to have the property P if for all $T \in \mathrm{ob}(\mathbf{Sch})$ and all $G : h_T \to Y$ the morphism $S_G \to T$ has property P, where $S_G \in \mathrm{ob}(\mathbf{Sch})$ is a representing object of G, i.e. $h_{S_G} \cong h_T \times_Y F$.

Definition 10.3.13 (Étale morphism). A morphism $F : X \to Y$ of algebraic spaces such that for every étale morphism $G : h_S \to X$ the composition $f \circ g$ is also étale.[7]

[7] In a similar vein, one can define smooth morphisms of algebraic spaces.

10.4 Algebraic groups & Galois theory

Definition 10.4.1 (Linear algebraic group). A subgroup of $\mathsf{GL}(n, \mathfrak{K})$ defined by a (finite) set of polynomials in the matrix coefficients.

Property 10.4.2. From the definition, it is immediately clear that intersections of algebraic groups are again algebraic.

Definition 10.4.3 (Absolute Galois group). The Galois group $\mathsf{Gal}(\mathfrak{K}_S/\mathfrak{K})$ of the separable closure (Definition 2.5.35) of $\mathfrak{K}$.

The following property relates Galois theory to algebraic geometry.

Property 10.4.4 (Grothendieck's Galois theory). Consider a field $\mathfrak{K}$ with separable closure $\mathfrak{K}_S$.

$$\mathsf{Gal}(\mathfrak{K}_S/\mathfrak{K}) \cong \pi_1\big(\mathsf{Spec}(\mathfrak{K})\big)\,, \tag{10.39}$$

where π_1 is the étale fundamental group (the fundamental group at the topological level would not carry a lot of information since $\mathsf{Spec}(\mathfrak{K})$ is a one-point space).

Definition 10.4.5 (Abelian variety). A smooth, projective, algebraic group.

Theorem 10.4.6 (Mordell–Weil). *An Abelian variety X over a number field $\mathfrak{K}$ (Definition 2.5.16) such that its set of rational points $X(\mathfrak{K})$ (Definition 10.2.10) is a finitely generated Abelian group.*

10.5 Tropical geometry

The following definition generalizes Definition 2.4.1.

Definition 10.5.1 (Rig). Let R be a set equipped with two binary operations $+, \cdot$ (called[8] **addition** and **multiplication**). $(R, +, \cdot)$ is a rig if it satisfies the following axioms:

1. $(R, +)$ is a commutative monoid.
2. $(R, \cdot)$ is a monoid.
3. Multiplication is distributive with respect to addition.

Remark 10.5.2. Whereas a ring is a group object in **CRing**, a rig is a group object in **CMonoid**. (The monoidal structures on these categories should not be taken to be the Cartesian ones for this result).

Example 10.5.3 (Tropical rig). Consider the set $T := \mathbb{R} \cup \{+\infty\}$. This set becomes a rig when equipped with the following operations:

1. **Addition**: $x \oplus y = \min(x, y)$.

[8] This might become confusing when looking at the examples below.

2. **Multiplication**: $x \cdot y = x + y$.

Definition 10.5.4 (Tropicalization). In ordinary algebraic geometry, one considers polynomials over $\mathbb{R}$ as the basic building blocks. These also give rise to polynomials over T. To every polynomial

$$f = a + \sum_{i=1}^{k} \lambda_i x_i^{n_i} \tag{10.40}$$

with $a \in \mathbb{Z}$, one assigns its tropicalization

$$\mathrm{trop}(f) := \min(0, n_1 x_1, \dots, n_k x_k) . \tag{10.41}$$

Note that the tropicalization of any polynomial is a piecewise linear function.

A more formal definition goes as follows. Let K_P be the ring of **Puiseux series**, i.e. of elements of the form

$$a = \sum_{i=k}^{+\infty} \lambda_i x^{i/n} \tag{10.42}$$

for some $k, n \in \mathbb{N}$. These generalize Laurent series (Definition 17.4.1) to fractional exponents in that $K_P = \bigcup_{i \in \mathbb{N}_0} \mathbb{C}((x^{1/i}))$. Then, consider the valuation (Definition 2.4.20)

$$\mathrm{val} : K_P \to \mathbb{R} : a \mapsto \min\{i/k \mid \lambda_i \neq 0\} . \tag{10.43}$$

The tropicalization of

$$f = \sum_{u \in \mathbb{N}^n} \lambda_u x^u \tag{10.44}$$

with $n \in \mathbb{N}$ is given by

$$\mathrm{trop}(f) := \min_{u \in \mathbb{N}^n, \lambda_u \neq 0} \left(\mathrm{val}(\lambda_u) + \sum_{i=1}^{n} u_i x_i \right) . \tag{10.45}$$

@@ COMPLETE (link with mirror symmetry (Gross, 2011)) @@

Chapter 11

Number Theory

The section on modular forms and L-functions is based on Mustaţă (2011); Sutherland (2017).

Chapter Contents

11.1 Adic numbers **250**
11.1.1 Finite fields 250
11.1.2 Rational numbers 251
11.1.3 Number and global fields 252
11.2 Algebraic geometry ♣ **254**
11.2.1 Modular forms 254
11.2.2 Algebraic functions 256
11.3 L-functions **256**

11.1 Adic numbers

This section will make use of the content of Section 6.8.2.

11.1.1 Finite fields

Definition 11.1.1 (Finite field). The finite field (or **Galois field**) $\mathbb{F}_p$ is the field with a finite number of elements $p \in \mathbb{N}$. The number p is called the **order** of $\mathbb{F}_p$.

Property 11.1.2 (Prime order). If $p \in \mathbb{N}$ is prime, then $\mathbb{F}_p \cong \mathbb{Z}/p\mathbb{Z}$.

11.1.2 Rational numbers

Definition 11.1.3 (p-adic numbers). Consider a prime number $p \in \mathbb{N}$. For every $m, n \in \mathbb{N}$, one has that $\mathbb{Z}/p^n\mathbb{Z} \subseteq \mathbb{Z}/p^m\mathbb{Z}$ whenever $m \leq n$. The group $\mathbb{Z}_p$ of **p-adic integers** is given by the profinite group (Definition 6.8.7) obtained from this system of inclusions. The p-adic integers are of the form

$$\sum_{n=0}^{+\infty} a_n p^n \,, \tag{11.1}$$

with integers $0 \leq a_n < p$ for all $n \in \mathbb{N}$.

Just like $\mathbb{Q}$ is the field of fractions (Definition 2.4.6) of $\mathbb{Z}$ as in Example 2.4.8, the p-adic numbers $\mathbb{Q}_p$ can be obtained as the field of fractions of $\mathbb{Z}_p$. The p-adic numbers are of the form

$$\sum_{n=-N}^{+\infty} a_n p^n \,, \tag{11.2}$$

with $N \in \mathbb{N}$ and integers $0 \leq a_n < p$ for all $n \in \mathbb{N}$.

An alternative definition makes use of the notion of valuations (Definition 2.4.20).

Alternative Definition 11.1.4 (p-adic numbers). Consider a prime number $p \in \mathbb{N}$. The p-adic valuation of an integer $z \in \mathbb{Z}$ is defined as follows:

$$\nu_p(z) := \begin{cases} \max\{n \in \mathbb{N} : p^n \mid z\} & \text{if } z \neq 0 \,, \\ +\infty & \text{if } z = 0 \,. \end{cases} \tag{11.3}$$

This valuation extends to the rational numbers by taking

$$\nu_p\left(\frac{a}{b}\right) := \nu_p(a) - \nu_p(b) \,. \tag{11.4}$$

In turn, the p-adic valuation also induces an absolute value on $\mathbb{Z}$ (and on $\mathbb{Q}$) as in Example 2.4.21, given by

$$|z|_p := p^{-\nu_p(z)} \,. \tag{11.5}$$

The metric completions of $\mathbb{Z}$ and $\mathbb{Q}$ by these absolute values are the rings of p-adic integers and numbers, respectively.

Remark 11.1.5. The p-adic valuation of a rational number $q \in \mathbb{Q}$ can also be defined as the (unique) integer $k \in \mathbb{Z}$ such that

$$q = p^k \frac{m}{n} \tag{11.6}$$

for some integers $m, n \in \mathbb{Z}$ such that p^k, m and n are all coprime.

Property 11.1.6. If $b \bmod p \neq 0$, then $\frac{a}{b} \in \mathbb{Z}_p$ for all $a, b \in \mathbb{Z}$ and $p \in \mathbb{N}$.

The construction of the p-adic representation of a rational number $q \in \mathbb{Q}$ proceeds similar to the construction of the usual decimal representation, i.e. through long division:

1. Find the unique coprime integers satisfying $q = \frac{c}{d} p^k$ and, hence, $\nu_p(q) = k$.
2. $k - 1$ zeros have to be added to the p-adic representation.
3. Then, find the unique integers satisfying $q = ap^k + r$ with $0 \leq a < p$ and $\nu_p(r) > k$.
4. Then the next coefficients is a and the algorithm can be restarted with $q := r$.

Theorem 11.1.7 (Ostrowski). *The only nontrivial absolute values on $\mathbb{Q}$ are either the ordinary absolute value $|\cdot|$ or the p-adic absolute values $|\cdot|_p$.*

More generally, consider a complete field $\mathfrak{K}$ with respect to an absolute value. Then, either the absolute value is Archimedean and $\mathfrak{K} \cong \mathbb{R}, \mathbb{C}$, or the absolute value is non-Archimedean.

Definition 11.1.8 (Profinite integers). Similar to the construction of the p-adic integers, one can also construct the profinite completion (Definition 6.8.8) of $\mathbb{Z}$. Instead of taking the inverse limit over the integers modulo a prime power, one simply takes the inverse limit over all finite cyclic groups:

$$\widehat{\mathbb{Z}} := \varprojlim_{n \in \mathbb{N}} \mathbb{Z}/n\mathbb{Z} \,. \tag{11.7}$$

It can be shown that this is equivalent to taking the product of all p-adic integers:

$$\widehat{\mathbb{Z}} \cong \prod_{p \text{ prime}} \mathbb{Z}_p \,. \tag{11.8}$$

Definition 11.1.9 (Adèles). The ring of integral adèles is defined as the product

$$\mathbb{A}_{\mathbb{R}} := \mathbb{R} \times \widehat{\mathbb{Z}} \,. \tag{11.9}$$

To obtain the proper ring of adèles, the above ring is rationalized:

$$\mathbb{A}_{\mathbb{Q}} := \mathbb{Q} \otimes_{\mathbb{Z}} \mathbb{A}_{\mathbb{R}} \,. \tag{11.10}$$

Definition 11.1.10 (Idèles). The group of units $\mathbb{I}_{\mathbb{Q}} := \mathbb{A}_{\mathbb{Q}}^{\times}$.

11.1.3 Number and global fields

The product expression from the previous section can be reinterpreted by noting that the rationalization of the integral adèles can be written as a **restricted product**. Recall Definition 3.4.59 of products. Given an indexing set S and a collection of morphisms $\{\iota_s : X_s \to Y_s\}_{s \in S}$, the restricted product $\prod'_{s \in S} Y_s$ is given by the subset of the ordinary product $\prod_{s \in S} Y_s$ where all but a finite number of elements lie in the image of the morphisms ι_s.

Now, by the **prime factorization theorem** and Property 11.1.6, every element of the form

$$\frac{a}{b} \otimes c_p \tag{11.11}$$

with $a, b \in \mathbb{Z}$ and $c_p \in \mathbb{Z}_p$ will either be an element of $\mathbb{Z}_p$ if $b \bmod p \neq 0$ or an element of $\mathbb{Q}_p$ if $b \bmod p = 0$. Since there are only a finite number of prime factors of b, this shows that

$$\mathbb{A}_{\mathbb{Q}} \cong \mathbb{R} \times \prod_{p \text{ prime}}' \mathbb{Q}_p \,. \tag{11.12}$$

Moreover, taking $\mathbb{Q}_\infty := \mathbb{R}$ and recalling Ostrowski's theorem, one has

$$\mathbb{A}_{\mathbb{Q}} \cong \prod_{p \in \text{places}(\mathbb{Z})}' \mathbb{Q}_p \,. \tag{11.13}$$

Now, consider a number field $\mathfrak{K}$ (Definition 2.5.16). Consider the following notations:

- $\mathcal{O}$ denotes the ring of algebraic integers (Definition 2.5.20) of $\mathfrak{K}$.
- P denotes the set of places of $\mathfrak{K}$ with $S \subset P$ the Archimedean (or infinite) places.
- For every place $\nu \in P$, $\mathfrak{K}_\nu$ denotes the completion at ν.

By Ostrowski's theorem, the completions at the infinite place are isomorphic to $\mathbb{R}$ or $\mathbb{C}$. For each (*local*) field $\mathfrak{K}_\nu$, one can also construct the **valuation ring** as the subring $\mathcal{O}_\nu$ of elements of norm at most 1.

The **integral adèles** are defined as follows, in analogy to Eq. (11.9):

$$\mathbb{A}_{\mathcal{O}} := \widehat{\mathcal{O}} \times \prod_{\nu \in S} \mathfrak{K}_\nu \tag{11.14}$$

and the **adèle ring** is defined as its 'rationalization':

$$\mathbb{A}_{\mathfrak{K}} := \mathfrak{K} \otimes_{\mathcal{O}} \mathbb{A}_{\mathcal{O}} \,. \tag{11.15}$$

Note that the profinite completion $\widehat{\mathcal{O}}$ again admits a factorization over non-Archimedean places:

$$\widehat{\mathcal{O}} \cong \prod_{\nu \in P \setminus S} \mathcal{O}_\nu \,. \tag{11.16}$$

Combining these facts, one obtains the following expression:

$$\mathbb{A}_{\mathfrak{K}} \cong \prod_{\nu \in S} \mathfrak{K}_\nu \times \prod_{\nu \in P \setminus S}' \mathfrak{K}_\nu \cong \prod_{\nu \in P}' \mathfrak{K}_\nu \,. \tag{11.17}$$

This definition can now be generalized to global fields as is.

Definition 11.1.11 (Global field). Either a number field or a function field (Definition 10.1.9) over a finite field. Equivalently, a finite extension of either the rational numbers $\mathbb{Q}$ or the rational functions with coefficients in a finite field.

11.2 Algebraic geometry ♣

This section gives a relation between number theory and (algebraic) geometry (Chapter 10). The content of Chapters 2 and 17 will also be used throughout this section.

11.2.1 Modular forms

Definition 11.2.1 (Modular group). In the setting of number theory, the projective special linear group $\mathsf{PSL}(2,\mathbb{Z})$ is often called the modular group.[1] The modular group acts on the complex plane by **Möbius transformations**:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z := \frac{az+b}{cz+d}\,. \tag{11.18}$$

For this reason, $\mathsf{PSL}(2,\mathbb{C})$ is sometimes also called the **Möbius group**. The full modular group $\mathsf{SL}(2,\mathbb{Z})$ will be denoted by Γ.

Definition 11.2.2 (Modular form). A modular form of weight $k \in \mathbb{R}$ is a holomorphic function on the upper-half plane $f : \mathcal{H} \to \mathbb{C}$ satisfying the following two conditions:

1. **Automorphicity**: For $g \equiv \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathsf{SL}(2,\mathbb{Z})$, one has

$$f\big(g \cdot z\big) = (cz+d)^k f(z)\,, \tag{11.19}$$

   and

2. **Bounded growth**: $f(z)$ is bounded for $z \longrightarrow i\infty$.

If the modular form satisfies the stronger condition $f(z) \longrightarrow 0$ when $z \longrightarrow i\infty$, it is said to be **cuspidal** or it is simply called a **cusp form**. The spaces of modular forms and cusp forms of weight $k \in \mathbb{R}$ are, respectively, denoted by $M_k(\Gamma)$ and $S_k(\Gamma)$.

@@ MAYBE ADD automorphicity in full generality @@

Remark 11.2.3 (Arithmetic group). Modular forms can also be defined for subgroups of $\mathsf{SL}(2,\mathbb{Z})$ of finite index, such as the *congruence groups* (or their generalizations, the *arithmetic groups*).

Property 11.2.4. The generators of the modular group are given by

$$z \mapsto -\frac{1}{z} \qquad \text{and} \qquad z \mapsto z+1\,. \tag{11.20}$$

Invariance under the second generator shows that modular forms are, in particular, periodic and, hence, admit a Fourier expansion. Cusp forms are exactly those modular forms with vanishing constant Fourier coefficient.

[1] This name is also used for the special linear group $\mathsf{SL}(2,\mathbb{Z})$ since Möbius transformations are invariant under rescaling.

Example 11.2.5 (Eisenstein series). Consider a complex number $\tau \in \mathcal{H}$. The Eisenstein series of weight $2k \in 2\mathbb{Z}$, with $k > 2$, is defined as follows:

$$G_{2k}(\tau) := \sum_{(m,n)\in\mathbb{Z}^2\setminus(0,0)} \frac{1}{(m+n\tau)^{2k}} \,. \tag{11.21}$$

Remark 11.2.6 (Fourier expansion). Since $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is an element of $\mathsf{PSL}(2,\mathbb{Z})$, every modular form is periodic and, hence, admits a Fourier expansion (Formula 19.3.3):

$$f(z) = \sum_{n=-\infty}^{+\infty} a_n e^{2\pi i z} \,. \tag{11.22}$$

f can then be interpreted as a function on the punctured disk by making the transformation $z \longrightarrow q := e^{2\pi i z}$. By Riemann's theorem on removable singularities (Theorem 17.5.3), the growth condition on modular forms is equivalent to $f(q)$ admitting a holomorphic continuation to the origin. Moreover, it is a cusp form exactly if $a_0 = 0$. The growth condition also implies that the Fourier series can be truncated to positive integers:

$$f(z) = \sum_{n=0}^{+\infty} a_n e^{2\pi i z} \,. \tag{11.23}$$

Definition 11.2.7 (Hecke operator). For every $n \in \mathbb{N}_0$, let Γ_n denote the subgroup of Γ on matrices with determinant n. On $M_k(\Gamma)$, one can define the following averaging operators:

$$T_n f(z) := n^{k-1} \sum_{g\in\Gamma\backslash\Gamma_n} (cz+d)^{-k} f(g\cdot z) \,, \tag{11.24}$$

where $g \equiv \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. One can easily see that $[T_m, T_n] = 0$.

The Hecke operators also have an important relation to cusp forms. The subspace $S_k(\Gamma)$ is preserved under the action of the Hecke operators and, moreover the joint eigenforms of the Hecke operators satisfy the following relations on their Fourier coefficients:

$$a_n = \lambda_n a_1 \,, \tag{11.25}$$

where λ_n is the nth eigenvalue. $S_k(\Gamma)$ admits an inner product, the **Petersson inner product**, defined as follows:

$$\langle f \,|\, g \rangle := \int_{\mathcal{F}} f(z)\overline{g}(z) y^{k-2} \, dx \, dy \,, \tag{11.26}$$

where $\mathcal{F} \subset \mathcal{H}$ is any fundamental domain for Γ. With respect to this inner product, there exists a normalized eigenbasis of the Hecke operators for $S_k(\Gamma)$, where normalized means that the first Fourier coefficient is equal to 1 (and, hence, the Fourier coefficients correspond to the Hecke eigenvalues).

11.2.2 Algebraic functions

Definition 2.5.17 can be generalized to the functional setting.

Definition 11.2.8 (Algebraic function). Let R be a commutative ring. A function $f : R^n \to R$ is said to be algebraic if it is the solution of a polynomial equation with coefficients in $R[x_1, \dots, x_n]$.[2] If f is not algebraic, it is said to be **transcendental**.

11.3 L-functions

Definition 11.3.1 (Dirichlet series). A series of the form

$$\sum_{n=1}^{+\infty} \frac{a_n}{n^s}, \tag{11.27}$$

where $s \in \mathbb{C}$ and $(a_n)_{n\in\mathbb{N}} \subset \mathbb{C}$. If the coefficients satisfy the growth bound $|a_n| = O(n^\sigma)$, the series converges (absolutely) to a holomorphic function whenever $\Re(s) > \sigma$.

The Dirichlet series for unit coefficients is of major interest in mathematics.

Definition 11.3.2 (Riemann ζ-function).

$$\zeta(s) := \sum_{n=1}^{+\infty} \frac{1}{n^s} \tag{11.28}$$

for all $s \in \mathbb{C}$ for which $\Re(s) > 1$. For general s, an *analytic continuation* is used (see Definition 17.4.5).

Formula 11.3.3 (Euler product).

$$\zeta(s) = \prod_{p \text{ prime}} \left(1 - \frac{1}{p^s}\right)^{-1} \tag{11.29}$$

for all $s \in \mathbb{C}$ for which $\Re(s) > 1$. This is an example of an Euler product, a product over all prime numbers.

Definition 11.3.4 (Selberg ζ-function). A function $\zeta : \mathbb{C} \to \mathbb{C}$ satisfying the following conditions:

1. There exists a $\eta \in \mathbb{R}^+$ (usually taken to be 1) such that ζ admits a convergent series expansion of the form

$$\zeta(s) = \sum_{n=1}^{+\infty} \lambda^{-s}, \tag{11.30}$$

for some sequence of real numbers $(\lambda_n)_{n\in\mathbb{N}} \subset \mathbb{R}$, or admits a Euler product expansion.

[2] Often, the polynomial is required to be irreducible.

2. ζ analytic continuation (Definition 17.4.5) of the above series on all of $\mathbb{C}$.

3. There exists a 'completion' $\hat{\zeta}$ of ζ that admits a functional relation of the form

$$\hat{\zeta}(1-s) = \hat{\zeta}(s)\,. \tag{11.31}$$

4*. *Ramanajuan conjecture*: $a_1 = 1$ and $a_n = O(n^\varepsilon)$ for any $\varepsilon > 0$.

@@ CHECK AND COMPLETE @@

Example 11.3.5 (Differential operators). Consider a self-adjoint *elliptic differential operator* (see Definition 34.11.1) D whose spectrum lies in the positive real numbers $\mathbb{R}_0^+$.

$$\zeta_D(s) := \mathrm{tr}\left(D^{-s}\right). \tag{11.32}$$

It can be shown that this is equal to the *Mellin transform* (see Formula 19.3.27) of the *partition function* of D:

$$\zeta_D(s) = \int_0^{+\infty} t^{s-1}\left(\sum_{n=1}^{+\infty} \exp(-t\lambda_n)\right) dt\,. \tag{11.33}$$

Remark 11.3.6 (Bost–Connes system). The *Bost–Connes system*, a C^**-algebra* (see Definition 26.2.6) of the form $C^*\left(\mathbb{Q}[\mathbb{Q}/\mathbb{Z}] \rtimes \mathbb{N}^\times\right)$, admits a Hamiltonian whose associated ζ-function is the Riemann ζ-function. A physical realization of this system would be able to shed light on the *Riemann hypothesis*.

Definition 11.3.7 (Hasse–Weil ζ-function[3]). Let V be a nonsingular projective variety over the finite field $\mathbb{F}_p$ and, for all $k \in \mathbb{N}$, let $N_K := \left|\mathbb{F}_{p^k}(V)\right|$ be the number of rational points over $\mathbb{F}_{p^k}$.

$$Z(V,t) := \exp\left(\sum_{k=1}^{+\infty} N_k \frac{t^k}{k}\right) \in \mathbb{Q}[[t]]\,. \tag{11.34}$$

Remark 11.3.8 (Weil conjectures). The Weil conjectures are three related statements of the Hasse–Weil function of profound importance. The conjectures were:

1. $Z(V,t)$ is a rational function of t.

2. $Z(V,t)$ admits a functional relation as in Eq. (11.31).

3. Discrete *Riemann hypothesis*: The zeroes of $P_k(t)$, for all $k \in \mathbb{N}$, lie on the critical line $\Re(z) = k/2$, where the $P_k(t)$ are the integral polynomials in the rational form of $Z(V,t)$:

$$Z(V,t) = \frac{P_1(t)\cdots P_{2\dim(V)-1}(t)}{P_0(t)\cdots P_{2\dim(V)}(t)} \tag{11.35}$$

with $P_0(t) = 1 - t$ and $P_{2\dim(V)}(t) = 1 - p^n t$.

[3] Also called the **local Hasse–Weil function** or **local L-function**.

4. If X is a *good reduction* of nonsingular projective variety over a number field in the complex numbers, then $\deg(P_k)$ equals the kth Betti number (Definition 7.2.17) of the set of complex points of V.

Items 1, 2 and 4 were proven by *Grothendieck* using *étale cohomology*, whereas the discrete Riemann hypothesis was proven by *Deligne* using *l-adic cohomology*.

Definition 11.3.9 (L-function of varieties). Consider a nonsingular projective variety V over $\mathbb{Q}$. For almost all primes p, one obtains a good reduction V_p. The Dirichlet series

$$Z_{\mathbb{Q},V}(s) := \prod_{p \text{ good prime}} Z(V_p, p^{-s}) \tag{11.36}$$

is called the L-function of V (or sometimes also the Hasse–Weil function of V).

Example 11.3.10 (Riemann ζ-function). Consider $V = \operatorname{Spec} \mathbb{Q}$ with $V_p \cong \operatorname{Spec}(\mathbb{F}_p)$ for prime $p \in \mathbb{N}$. In this case,

$$Z(V_p, t) = \exp\left(\sum_{n=1}^{+\infty} \frac{t^n}{n}\right) = (1-t)^{-1} \tag{11.37}$$

and, hence,

$$Z_{\mathbb{Q},V}(s) = \prod_{p \text{ prime}} (1-p^{-s})^{-1} = \zeta(s) \tag{11.38}$$

by Formula 11.3.3.

Recall the cusp forms of Section 11.2.1.

Definition 11.3.11 (L-function of modular forms). Let $f : \mathcal{H} \to \mathbb{C}$ be a cusp form of weight $k \in \mathbb{R}$ and consider its Fourier series

$$f(z) = \sum_{n=1}^{+\infty} a_n e^{2\pi i z}. \tag{11.39}$$

The associated L-function is given by the following Dirichlet series:

$$L_f(s) := \sum_{n=1}^{+\infty} \frac{a_n}{n^s}. \tag{11.40}$$

The domain of convergence is bounded by $\Re(s) > 1 + \frac{k}{2}$.

Theorem 11.3.12 (Modularity theorem[4]). *Every Hasse–Weil L-function of an elliptic curve can be obtained as the L-function of a modular form.*

[4] Also called the **Taniyama–Shimura(–Weil) theorem/conjecture** (or any of its permutations).

Part III

Higher Set Theory

Table of Contents

12 Topos theory **262**

12.1 Elementary topoi . . . 262

12.2 Morphisms . . . 264

12.3 Internal logic . . . 266

12.4 Grothendieck topoi . . . 268

12.4.1 Grothendieck topologies . . . 268

12.4.2 Topological sheaves . . . 273

12.4.3 Lawvere–Tierney topology . . . 275

12.4.4 Diaconescu's theorem . . . 277

13 Model theory ♣ **281**

13.1 Simplicial sets . . . 282

13.1.1 Constructions . . . 283

13.1.2 Homological algebra . . . 285

13.2 Localization . . . 286

13.3 Model categories . . . 288

13.3.1 Model structures . . . 288

13.3.2 Monoidal structures . . . 290

13.3.3 Homotopy category . . . 291

13.3.4 Reedy model structure ♣ . . . 294

13.4 Simplicial spaces . . . 297

13.4.1 Kan complexes . . . 297

13.4.2 Simplicial localization . . . 298

13.4.3 Segal spaces . . . 299

13.4.4 Coherent nerve . . . 301

13.5 Cofibrantly generated categories . . . 303

13.5.1 Transfinite constructions . . . 303

13.5.2 Cofibrant generation . . . 305

13.6 Homotopy (co)limits . . . 307

13.6.1 Simplicially enriched diagrams . . . 308

14 Higher topos theory ♣ **310**
14.1 Higher category theory . . . 310
14.1.1 n-categories . . . 310
14.1.2 n-functors . . . 313
14.1.3 Higher (co)limits . . . 314
14.2 ∞-Categories . . . 315
14.2.1 Simplicial approach . . . 315
14.2.2 Quasicategories . . . 315
14.2.3 Categorical notions . . . 316
14.3 Stacks . . . 316
14.3.1 2-sheaves . . . 316
14.3.2 Stacks on a 1-site . . . 318
14.4 Higher topos theory . . . 320
14.5 Cohomology . . . 321

15 Logic and Type Theory **324**
15.1 Logic . . . 325
15.1.1 Languages . . . 325
15.1.2 Propositional logic . . . 326
15.1.3 Sequent calculus . . . 327
15.1.4 Predicate logic . . . 328
15.2 Introduction to type theory . . . 328
15.3 Basic constructions . . . 329
15.3.1 Functions . . . 329
15.3.2 λ-calculus . . . 330
15.3.3 Products . . . 330
15.3.4 Propositions as types . . . 334
15.3.5 Identity types . . . 334
15.4 Categorical semantics . . . 337
15.4.1 Inductive types . . . 337
15.4.2 Polynomial functors . . . 337
15.5 Homotopy type theory . . . 338
15.5.1 Introduction . . . 338
15.5.2 Transport . . . 339
15.5.3 Equivalences . . . 341
15.5.4 Equality types: revisited . . . 342
15.6 Modal logic . . . 343
15.6.1 Modalities . . . 343
15.6.2 Type theory . . . 344

Chapter 12

Topos theory

The main reference for this chapter is Caramello (2018, 2019); Johnstone (2014); Mac Lane and Moerdijk (1994). For an introduction to stacks and descent theory, see Vistoli (2004).

Chapter Contents

12.1 Elementary topoi **262**

12.2 Morphisms **264**

12.3 Internal logic **266**

12.4 Grothendieck topoi **268**

12.4.1 Grothendieck topologies 268

12.4.2 Topological sheaves 273

12.4.3 Lawvere–Tierney topology 275

12.4.4 Diaconescu's theorem 277

12.1 Elementary topoi

Definition 12.1.1 (Subobject classifier). Consider a finitely complete category (in fact, the existence of a terminal object suffices). A subobject classifier is a mono[1] $\texttt{true} : 1 \hookrightarrow \Omega$ from the terminal object such that for every mono $\phi : x \hookrightarrow y$ there exists a unique morphism $\chi : y \to \Omega$ that fits in the following pullback square:

Alternative Definition 12.1.2. Consider a well-powered category $\mathbf{C}$. The assignment of subobjects $\mathsf{Sub}(x)$ to an object $x \in \mathsf{ob}(\mathbf{C})$ defines functor $\mathsf{Sub} : \mathbf{C}^{op} \to \mathbf{Set}$. A subobject classifier Ω is a representation of this functor, i.e. the following isomorphism is natural

[1] The symbol for this morphism will become clear in Section 12.3.

$$\begin{tikzcd} x \arrow[r] \arrow[d, hook, "\phi"'] \arrow[dr, phantom, "\text{pb}"] & 1 \arrow[d, hook, "\texttt{true}"] \\ y \arrow[r, "\exists!\chi"'] & \Omega \end{tikzcd}$$

Figure 12.1: Subobject classifier.

in x:

$$\mathsf{Sub}(x) \cong \mathbf{C}(x, \Omega)\,. \tag{12.1}$$

Example 12.1.3 (Indicator function). The category **Set** has the 2-element set $\{\texttt{true}, \texttt{false}\}$ as subobject classifier. The morphism $\chi : S \to \Omega$ is the indicator function

$$\chi_S(x) = \begin{cases} \texttt{true} & \text{if } x \in S\,, \\ \texttt{false} & \text{if } x \notin S\,. \end{cases} \tag{12.2}$$

Definition 12.1.4 (Elementary topos). An elementary topos is a finitely complete, Cartesian closed category (Definition 4.1.20) admitting a subobject classifier. Equivalently, one can define an elementary topos as a finitely complete category that has all power objects exist.

The power object Px of $x \in \mathsf{ob}(\mathcal{E})$ is related to the subobject classifier Ω by the following relation:

$$Px = \Omega^x\,. \tag{12.3}$$

Remark 12.1.5 (Finite colimits). The original definition by *Lawvere* also required the existence of finite colimits. However, it can be proven that finite cocompleteness follows from the other axioms.

Theorem 12.1.6 (Fundamental theorem of topos theory). *Let $\mathcal{E}$ be an elementary topos. The slice category $\mathcal{E}_{/x}$ is also a topos for every object $x \in \mathsf{ob}(\mathcal{E})$. The subobject classifier is given by $\pi_2 : \Omega \times x \to x$.*

Property 12.1.7 (Balanced). All monos in a topos are regular. Hence, every mono arises as an equalizer. Since, by Property 3.4.63, every epic equalizer is necessarily an isomorphism, it follows that every topos is balanced (Definition 3.4.5).

Property 12.1.8 (Epi/mono factorization). Every morphism $f : x \to y$ in a topos factorizes uniquely as an epi followed by a mono:

$$x \overset{e}{\twoheadrightarrow} z \overset{m}{\hookrightarrow} y\,. \tag{12.4}$$

The mono is called the **image** of f.

12.2 Morphisms

Definition 12.2.1 (Base change). Consider a category $\mathbf{C}$ with pullbacks. For every morphism $f : x \to y$ one can define a functor $f^* : \mathbf{C}/y \to \mathbf{C}/x$. This functor acts by pullback along f.

Definition 12.2.2 (Dependent sum and product). Consider a base change functor f^*. The dependent sum and product functors are given by the right and left adjoints (if they exist):

$$\sum_f \dashv f^* \dashv \prod_f . \tag{12.5}$$

Remark 12.2.3. The dependent sum can be shown to exist over any category. In fact, as a functor, it is simply given by postcomposition with f. However, the interpretation is much less trivial. Any morphism can be interpreted as a "space" with fibres the preimages. For example, in **Set**, a morphism $g : Z \to X$ represents Z as follows:

$$Z \cong \bigsqcup_{x \in X} g^{-1}(x) \equiv \sum_{x:X} g^{-1}(x).$$

The dependent sum generalizes this construction in a fibrewise manner, i.e. the fibre over a point $y \in Y$ is given by

$$\sum_f g\big|_y = \bigsqcup_{x \in f^{-1}(y)} g^{-1}(x) . \tag{12.6}$$

Instead of combining all fibres into a single space, it combines those that lie in a single fibre of f.

Although the dependent sum always exists, the dependent product requires more structure (see e.g. Huang (2022)).

Property 12.2.4 (Locally Cartesian closed categories). In a locally Cartesian closed category (Definition 4.1.20), all dependent products exist. In fact, the existence of the adjoint triple (12.5) is equivalent to $\mathbf{C}$ being locally Cartesian closed. For example, in **Set**, the dependent product is given by

$$\prod_f g\big|_y = \prod_{x \in f^{-1}(y)} g^{-1}(x) . \tag{12.7}$$

When f is the terminal morphism, this represents the space of sections $X \to Z \equiv \sum_{x:X} g^{-1}(x)$ of g.

More generally, the functor is obtained through the product-exponential adjunction in slice categories. The pullback is equal to a product in a slice category. Since $\mathbf{C}$ is locally Cartesian, one obtains for all $p : a \to y$ and $q : b \to y$:

$$\mathbf{C}_{/x}(p \times f, q) \cong \mathbf{C}_{/y}(p, q^f) \tag{12.8}$$

for some morphism $q^f : b^f \to y$. Applying this exponential functor to the identity morphism $\mathbb{1}_f$, gives a morphism $s : \mathbb{1}_y \to f^f$ or $s : y \to x^f$ (this simply picks out the identity morphism in the exponential object internally). The dependent product is then defined as the pullback $\prod_f g := s^* g^f$, where g^f is interpreted as the morphism $(f \circ g)^f \to f^f$ in $\mathbf{C}_{/y}$.

The idea behind this pullback construction is that in **Set**, the pullback of $g^X : Z^X \to X^X$ along $s : \mathbb{1}_X \to X^X$ is equal to the set

$$\{h \in Z^X \mid h \circ g = \mathbb{1}_Z\}, \tag{12.9}$$

i.e. it consists of all sections of g. Hence, the pullback construction generalizes Eq. (12.7).

Definition 12.2.5 (Logical morphism). A functor $f : \mathcal{E} \to \mathcal{F}$ between elementary topoi is said to be logical if it preserves finite limits, exponential objects and the subobject classifier.

Property 12.2.6. If a logical morphism has a left adjoint if and only if it has a right adjoint.

Definition 12.2.7 (Geometric morphism). A geometric morphism $f : \mathcal{E} \to \mathcal{F}$ of elementary topoi consists of an adjunction

$$\mathcal{E} \underset{f_*}{\overset{f^*}{\underset{\longrightarrow}{\overset{\longleftarrow}{\perp}}}} \mathcal{F},$$

where the left adjoint is left exact. The right adjoint f_* is called the **direct image** part of f and the left adjoint f^* is called the **inverse image** part. If f^* itself has a left adjoint, then f is said to be **essential**.

Definition 12.2.8 (Geometric surjection). A geometric morphism for which the inverse image part is faithful or, equivalently, reflects isomorphisms, is said to be **surjective** or is called a geometric surjection.

Definition 12.2.9 (Geometric embedding). A geometric morphism for which the direct image part is fully faithful. If $\mathcal{E} \hookrightarrow \mathcal{F}$ is a geometric embedding, $\mathcal{E}$ is sometimes called a **subtopos** of $\mathcal{F}$. Moreover, if it is an essential geometric morphisms, the subtopos is itself said to be **essential**. Essential subtopoi $\mathcal{E} \hookrightarrow \mathcal{F}$ are also called **levels** of $\mathcal{F}$.

Property 12.2.10 (Characterization of geometric embeddings). Let $f : \mathcal{E} \hookrightarrow \mathcal{F}$ be a geometric embedding and let $W \subset \mathsf{hom}(\mathcal{F})$ be the collection of morphisms that are mapped to isomorphisms under f^*. $\mathcal{E}$ is both equivalent to the full subcategory of $\mathcal{F}$ on W-local objects (Definition 3.2.21) and the *localization* $\mathcal{F}[W^{-1}]$ at W (see Definition 13.2.5).

Property 12.2.11 (Base change). The base change functors on a topos are logical and

admit a left adjoint, the postcomposition functor. This implies that these functors can be refined to essential geometric morphisms.

Example 12.2.12 (Topological spaces). Every continuous function $f : X \to Y$ induces a geometric morphism

$$\mathbf{Sh}(X) \underset{f_*}{\overset{f^*}{\underset{\longrightarrow}{\overset{\longleftarrow}{\perp}}}} \mathbf{Sh}(Y), \tag{12.10}$$

where the direct image functor f_* is defined as

$$f_*F(U) := F(f^{-1}U) \tag{12.11}$$

for any sheaf $F \in \mathbf{Sh}(X)$ and any open subset $U \in \mathbf{Open}(Y)$. The inverse image functor f^* is defined using the equivalence between sheaves on topological spaces and étalé spaces. Consider a sheaf $E \in \mathbf{Sh}(Y)$ as an étalé space $\pi : E \to Y$. The inverse image of E along a continuous function $f : X \to Y$ is the pullback of π along f.

This example implies that the global elements $* \to X$ of a topological space induce geometric morphisms of the form $\mathbf{Sh}(*) \to \mathbf{Sh}(X)$. By noting that $\mathbf{Sh}(*) = \mathbf{Set}$, one obtains the following generalization.

Definition 12.2.13 (Point). A point of a topos $\mathcal{E}$ is a geometric morphism $\mathbf{Set} \to \mathcal{E}$.

Notation 12.2.14 (Category of topoi). The category of elementary topoi and geometric morphisms is a 2-category. It is denoted by **Topos**.

To obtain the structure of a 2-category, one needs to define an appropriate notion of 2-morphism. Because a geometric morphism consists of an adjunction, one can consider two distinct conventions. Either, one can choose the 2-morphisms in **Topos** to be the natural transformations $f^* \Rightarrow g^*$ (with associated transformations $g_* \Rightarrow f_*$) or, one can choose them to be the natural transformations $f_* \Rightarrow g_*$ (and associated transformations $g^* \Rightarrow f^*$). This chapter follows Johnstone (2014) and the 'inverse image convention' is used, i.e. a 2-morphism $f \Rightarrow g$ consists of natural transformations $f^* \Rightarrow g^*$ and $g_* \Rightarrow f_*$.

Property 12.2.15 (Balanced). The category **Topos** is balanced (Definition 3.4.5), i.e. every geometric morphism that is both a geometric embedding and a geometric surjection is an equivalence.

Theorem 12.2.16 (Factorization). *Every geometric morphisms can be factorized as a geometric surjection followed by a geometric embedding.*

12.3 Internal logic

In this subsection, finitely complete categories that admit a subobject classifier are considered (they do not have to be elementary topoi).

Definition 12.3.1 (Truth value). A global element of the subobject classifier, i.e. a morphism $1 \to \Omega$. The subobject classifier Ω is also sometimes called the **object of truth values**.

Property 12.3.2 (Heyting algebra). For all objects x in an elementary topos, the poset of subobjects $\mathsf{Sub}(x)$ has the structure of an internal meet-semilattice since forming pullbacks of monos gives a natural transformation

$$\cap_x : \mathsf{Hom}(x, \Omega \times \Omega) \to \mathsf{Hom}(x, \Omega)\,. \tag{12.12}$$

By the Yoneda lemma, this induces an internal map

$$\wedge : \Omega \times \Omega \to \Omega\,. \tag{12.13}$$

In fact, it can be shown that this gives the structure of an internal Heyting algebra (Definition 1.7.44) and, in particular, that of an internal locale (Definition 6.8.2). Hence, every topos canonically gives an external Heyting algebra, namely $\mathsf{Sub}(1)$. Furthermore, every power object is an internal Heyting algebra. This in particular includes the subobject classifier $\Omega = P1$.

Definition 12.3.3 (Mitchell–Bénabou language). Let $\mathcal{E}$ be an elementary topos with subobject classifier Ω.

1. **Type**: An object $x \in \mathsf{ob}(\mathcal{E})$.
2. **Variable** (of type x): An identity morphism $\mathbb{1}_x$.
3. **Term** (of type x in variables α_i of type x_i): a morphism $\prod_{i \in I} x_i \to x$.
4. **Formula** or **proposition**: a term of type Ω. Moreover, a formula is deemed true if it factors through $\mathtt{true} : 1 \to \Omega$.

The logical connectives are induced by the internal Heyting structure on Ω. Quantifiers are induced by the (internal) completeness of Ω.

Every type also comes equipped with two binary relations:

1. $=_x$ is obtained from the characteristic morphism of the diagonal inclusion.
2. $\in_x$ is obtained by using the evalution map of the exponential object together with Eq. (12.3).

Since the slices of an elementary topos are themselves elementary topoi, one can also define dependent types. An **indexed type** is an object of a slice category. **(Dependent) sums** and **products** of type families are goven by the left and right adjoints to the base change functor.

An interpretation of the Mitchell–Bénabou language in an elementary topos $\mathcal{E}$ goes as follows.

Definition 12.3.4 (Kripke–Joyal semantics). The **entailment** $U \vdash \phi(\alpha)$ for a term $\alpha : U \to X$ and a proposition $\phi(x) : X \to \Omega$ is said to hold if α factors through the pullback $\{x \in X \mid \phi(x)\}$.

Property 12.3.5 (Monotonicity). If $V \vdash \phi(\alpha)$ and $f : U \to V$, then $U \vdash \phi(\alpha \circ f)$.

Property 12.3.6 (Locality). If $f : U \to V$ is epic and $V \vdash \phi(\alpha \circ f)$, then $U \vdash \phi(\alpha)$.

12.4 Grothendieck topoi

12.4.1 Grothendieck topologies

Definition 12.4.1 (Sieve). Let $\mathbf{C}$ be a small category. A sieve S on $\mathbf{C}$ is a fully faithfull discrete fibration $S \hookrightarrow \mathbf{C}$.

A sieve S on an object $x \in \mathbf{C}$ is a sieve in the slice category $\mathbf{C}_{/x}$, i.e. S is a subset of $\mathsf{ob}(\mathbf{C}_{/x})$ closed under precomposition. In other words, if $y \to x \in S$ and $z \to y \in \mathsf{hom}(\mathbf{C})$, then $z \to y \to x \in S$.

All of this can be summarized by saying that a sieve on an object $x \in \mathsf{ob}(\mathbf{C})$ is a subfunctor of the hom-functor $\mathbf{C}(-, x)$.

Example 12.4.2 (Maximal sieve). Let $\mathbf{C}$ be a category. The maximal sieve on $x \in \mathsf{ob}(\mathbf{C})$ is the collection of all morphisms $\{f \in \mathsf{hom}(\mathbf{C}) \mid \mathsf{cod}(f) = x\}$ or, equivalently, all of $\mathsf{ob}(\mathbf{C}_{/x})$.

Example 12.4.3 (Pullback sieve). Consider a morphism $f : x \to y$. Given a sieve S on y, one can construct the pullback sieve f^*S on x as the sieve of morphisms in S that factor through f:

$$f^*S(x) = \left\{g \mid f \circ g \in S(y)\right\}. \tag{12.14}$$

Property 12.4.4 (Presheaf topos). Consider the presheaf category $\mathbf{Psh}(\mathbf{C})$ on an arbitrary (small) category $\mathbf{C}$. This category is an elementary topos, where the subobject classifier assigns sieves:

$$\underline{\Omega}(x) := \{S \mid S \text{ is a sieve on } x\}. \tag{12.15}$$

The action on a morphism $f : x \to y$ gives the morphism $\underline{\Omega}(f)$ that sends a sieve S to its pullback sieve f^*S.

The morphism $\mathtt{true} : \underline{1} \hookrightarrow \underline{\Omega}$ is defined as the natural transformation assigning to every object its maximal sieve. For every subobject $\underline{Y} \hookrightarrow \underline{X}$ the characteristic morphism $\chi_{\underline{Y}}$ is defined as follows. Consider an object $c \in \mathsf{ob}(\mathbf{C})$. The component $\chi_{\underline{Y}}|_c$ is then given by

$$\chi_{\underline{Y}}|_c(x) := \left\{f \in \mathbf{C}(d, c) \mid \underline{X}(f)(x) \in \underline{Y}(d)\right\}, \tag{12.16}$$

for all $x \in \underline{X}(c)$.

The following definition is due to *Giraud* (for the original definition using the notion of a *cover*, see the end of this section).

Definition 12.4.5 (Grothendieck topology). A Grothendieck topology on a category $\mathbf{C}$ is a map J assigning to every object a collection of sieves satisfying the following conditions:

1. **Identity**[2]: For every object $x \in \mathsf{ob}(\mathbf{C})$, the maximal sieve M_x is an element of $J(x)$ or, equivalently, all sieves generated by isomorphisms are in $J(x)$.
2. **Base change**: If $S \in J(x)$, then $f^*S \in J(y)$ for every morphism $f : y \to x$.
3. **Locality**: Consider a sieve S on $x \in \mathsf{ob}(\mathbf{C})$. If there exists a sieve $R \in J(x)$, such that for every morphism $(f : y \to x) \in R$ the pullback sieve $f^*S \in J(y)$, then $S \in J(x)$.

The sieves in J are called (J-)**covering sieves**. A collection of morphisms with codomain $x \in \mathsf{ob}(\mathbf{C})$ is called a **cover**[3] of x if the sieve generated by these morphisms is a covering sieve on x.

Example 12.4.6 (Topological spaces). These conditions have the following interpretation in the case of topological spaces:

- The collection of all open subsets covers a space U.
- If $\{U_i\}_{i\in I}$ covers U, then $\{U_i \cap V\}_{i\in I}$ covers $U \cap V$.
- If $\{U_i\}_{i\in I}$ covers U and, if for every $i \in I$ the collection $\{U_{ij}\}_{j\in J_i}$ covers U_i, then $\{U_{ij}\}_{i\in I, j\in J_i}$ covers U.

The canonical Grothendieck topology on $\mathbf{Open}(X)$ is given by the sieves $S = \{U_i \hookrightarrow U\}_{i\in I}$, where $\bigcup_{i\in I} U_i = U$. This topology is denoted by $J_{\mathbf{Open}(X)}$.

Definition 12.4.7 (Site). A (small) category equipped with a Grothendieck topology J.

A slightly weaker notion than that of a (Grothendieck) topology is the following.

Definition 12.4.8 (Coverage). Let $\mathbf{C}$ be a category. A coverage on $\mathbf{C}$ is a map that assigns to every object $x \in \mathsf{ob}(\mathbf{C})$ a collection of families $\{f : y \to x\} \subset \mathsf{hom}(\mathbf{C})$, the **covering families** or **covers**, satisfying the following condition. If $\{f : y \to x\}$ is a covering family on x, then for every morphism $g : x' \to x$ there exists a covering family $\{f' : y' \to x'\}$ on x' such that every composite $g \circ f'$ factors through some f.

Definition 12.4.9 (Matching family). Consider a presheaf $F \in \mathbf{Psh}(\mathbf{C})$ together with a sieve S on $x \in \mathsf{ob}(\mathbf{C})$. A matching family for S with respect to F is a natural transforma-

[2]The name itself stems from the fact that the maximal sieve is generated from the identity morphism.

[3]Sometimes this term is also used to denote any collection of morphism with common codomain x, i.e. without reference to a covering sieve.

tion $\alpha : S \Rightarrow F$ between S, regarded as a subfunctor of $\mathbf{C}(-, x)$, and F.

More explicitly, it is an assignment of an element $x_f \in Fy$ to every morphism $(f : y \to x) \in S$ such that

$$F(g)(x_f) = x_{f \circ g} \tag{12.17}$$

for all morphisms $g : z \to y$. Equivalently, a matching family for S with respect to F is a set of elements $\{x_f\}_{f \in S}$ such that for all covering morphisms $f : y \to x, g : z \to x \in S$ and all morphisms $f' : c \to y, g' : c \to z$ such that $f \circ f' = g \circ g'$ the following equations holds:

$$F(f')(x_f) = F(g')(x_g)\,. \tag{12.18}$$

Given such a matching family, one calls an element $a \in Fx$ an **amalgamation** if it satisfies

$$F(f)(a) = x_f \tag{12.19}$$

for all morphisms $f \in S(y)$.

The existence of such an element can also be stated in terms of natural transformations. Consider the obvious inclusion ι_S of S into the the hom-functor $\mathbf{C}(-, x)$. Every morphism with codomain x can be obtained from the identity morphism by precomposition and, hence, a natural transformation $\mathbf{C}(-, x) \Rightarrow F$ is determined by its action on the identity morphisms $\mathbb{1}_x$. The existence of an amalgamation is thus equivalent to the existence of an extension of S along ι_S.

Remark 12.4.10. If the base category has all pullbacks, for example if it is a topos on its own, one can restrict the above commuting diagrams to the pullback diagrams of morphisms in the sieve S.

Definition 12.4.11 (Sheaf). Consider a site $(\mathbf{C}, J)$. A presheaf F on $\mathbf{C}$ is called a J-sheaf if every matching family, for every covering sieve in J, admits a unique amalgamation[4] or, equivalently, if all sieves admit a unique extension to representable presheaves.

The category $\mathbf{Sh}(\mathbf{C}, J)$ of J-sheaves on the site $(\mathbf{C}, J)$ is the full subcategory of $\mathbf{Psh}(\mathbf{C})$ on the presheaves that satisfy the above condition.

This definition can also be restated in terms of local objects (Definition 3.2.21).

Alternative Definition 12.4.12 (Sheaf). By definition every covering sieve admits a morphism into the Yoneda embedding: $\eta : S \hookrightarrow \mathcal{Y}x$. If the collection of all these morphisms is denoted by $\mathcal{S}$, a presheaf is a sheaf if and only if it is $\mathcal{S}$-local, i.e. if the following morphism is an isomorphism for all $\eta \in \mathcal{S}$:

$$Fx \cong \mathbf{Psh}(\mathcal{Y}x, F) \xrightarrow{\mathbf{Psh}(\eta, F)} \mathbf{Psh}(S, F)\,. \tag{12.20}$$

[4] If there exists at most one amalgamation, the presheaf is said to be **separated**.

This is also called the **descent condition** for sheaves. In this context the collection of matching families $\mathsf{Match}(S, F) := \mathbf{Psh}(S, F)$ for a sieve S with respect to a presheaf F is often called the **descent object** of S with respect to F.

Example 12.4.13 (Topological spaces). The usual category of sheaves $\mathbf{Sh}(X)$ on a topological space X is obtained as the category of sheaves on the site $(\mathbf{Open}(X), J_{\mathbf{Open}(X)})$. A Grothendieck topos of this form is also called a **spatial topos**. Since the morphisms in the covering sieves are exactly the inclusion maps $U_i \hookrightarrow U$, the pullback of two such morphisms is given by the intersection $U_i \cap U_j$. Hence, the condition for a matching family, as formulated in Definition 12.4.9 above, gives the second part of Definition 8.2.1. The uniqueness of an amalgamation is equivalent to the first part of that definition.

For topological spaces, sheaves are easily represented visually. A matching family assigns to every set U_i of an open cover $\mathcal{U} \equiv \{U_i\}_{i \in I}$ of U an element $x_i \in FU_i$, such that the restrictions coincide on double overlaps, as in step (1) in the figure below.

$$
\begin{tikzcd}
 & \text{Cover} & & \text{Matching family} & & \text{Unique amalgamation} \\
 & U_i \cap U_j \arrow[dl, hook] \arrow[dr, hook] & & x_i|_{U_i \cap U_j} = x_j|_{U_i \cap U_j} \arrow[dl] \arrow[dr] & & x_i|_{U_i \cap U_j} = x_j|_{U_i \cap U_j} \arrow[dl] \arrow[dr] \\
U_i \arrow[dr, hook] & & U_j \arrow[dl, hook] \arrow[r, "(1)"] & x_i \arrow[dr] & x_j \arrow[dl] \arrow[r, "(2)"] & x_i = x|_{U_i} \arrow[dr] & x_j = x|_{U_j} \arrow[dl] \\
 & U & & & & & x
\end{tikzcd}
$$

The descent condition then states that for every such matching family, there exists a unique element x on U, such that the elements of the matching family are restrictions of x as in step (2) of the figure above.

The classical example would be the assignment of the set of continuous functions to open subsets of a topological space. When two functions, defined on two open sets, coincide on the intersection, there exists a unique continuous function defined on the union, such that it restricts to the given functions.

Above, in Definition 12.4.8, the notion of a coverage was introduced. It should be clear that every coverage generates a sieve. Furthermore, although coverages are weaker and easier to handle, they are in fact equivalent for the purpose of sheaf theory.

Property 12.4.14. Consider a covering family C and let S_C be the sieve it generates. A presheaf is a sheaf for C if and only if it is a sheaf for S_C.

Example 12.4.15 (Canonical topology). The canonical topology on a category is the finest Grothendieck topology for which all representable presheaves are sheaves. A **subcanonical topology** is then defined as a subtopology of the canonical one, i.e. any Grothendieck topology for which all representable presheaves are sheaves.

Example 12.4.16 (Minimal and maximal topologies). The minimal Grothendieck topology on a category is the one for which only the maximal sieves are covering sieves. In this topology all presheaves are sheaves. The maximal Grothendieck topology is the one for which all sieves are covering sieves. In this topology only the terminal element of the associated presheaf category is a sheaf.

Definition 12.4.17 (Grothendieck topos). A category equivalent to the category of sheaves on a (small) site. This site is often called the **site of definition** for the given topos.

Property 12.4.18. Every Grothendieck topos is an elementary topos.

Property 12.4.19. For every Grothendieck topos, there exists a (standard) site of definition for which the Grothendieck topology is (sub)canonical.

Property 12.4.20. Let $\mathcal{E}$ be a Grothendieck topos. The canonical topology J_{can} on $\mathcal{E}$ is the topology for which the covering sieves are jointly epimorphic families. Moreover,

$$\mathcal{E} \cong \mathbf{Sh}(\mathcal{E}, J_{\text{can}}) . \tag{12.21}$$

Construction 12.4.21 (Sheafification). Given a presheaf $\mathcal{F}$, one can construct a sheaf $\overline{\mathcal{F}}$ along the same lines of Construction 8.2.19.

Definition 12.4.22 (Global sections functor). Every Grothendieck topos $\mathcal{E}$ admits a geometric morphism to **Set**, where the right adjoint assigns to an object x its set of global elements:

$$\Gamma : \mathcal{E} \to \mathbf{Set} : x \mapsto \mathcal{E}(1, x) . \tag{12.22}$$

When $\mathcal{E}$ is the sheaf topos over a topological space, this is exactly the global sections functor (Example 8.2.9). The left adjoint assigns to every set S the copower $S \cdot 1 \equiv \bigsqcup_{s \in S} 1$. When $\mathcal{E}$ is a sheaf topos, this adjoint is exactly the constant sheaf functor. It is sometimes denoted by LConst (cf. Example 8.2.20).

A different approach for defining sheaf topoi is through an embedding of sheaves into presheaves.

Definition 12.4.23 (Local isomorphism). A system of local isomorphisms in $\mathbf{Psh}(\mathbf{C})$ is a class of morphisms in $\mathbf{Psh}(\mathbf{C})$ forming a *system of weak equivalences* (see Definition 13.2.1) closed under pullbacks along morphisms out of representable presheaves.

Property 12.4.24 (Local isomorphisms and Grothendieck topologies). A system of local isos induces a *system of local epis* in the following way. $f : X \to Y$ is a local epi if $\mathsf{im}(f) \to Y$ is a local iso. A Grothendieck topology is defined by declaring a presheaf $F \in \mathbf{Psh}(\mathbf{C})$ to be a covering sieve at $X \in \mathsf{ob}(\mathbf{C})$ if $F \hookrightarrow \mathcal{Y}X$ is a local epi.

Alternative Definition 12.4.25 (Sheaf topos). A category $\mathbf{Sh}(\mathbf{C})$ equipped with a geometric embedding into $\mathbf{Psh}(\mathbf{C})$.

Proof (Proof of equivalence). By Property 12.2.10, such a category is equivalent to the full subcategory on S-local presheaves for some system of local isomorphisms S and, therefore, also to a sheaf topos in the sense of Grothendieck by the property above. □

Remark 12.4.26 (Descent condition). This is essentially a restatement of the descent condition (cf. 12.4.12). Covering sieves, regarded as subfunctors, are in particular local isomorphisms. Stability of sieves under pullback, together with the co-Yoneda lemma 3.4.80, which says that every presheaf is a colimit of representables, generates the full collection of local isomorphisms.

The characterization of geometric embeddings and, hence, of Grothendieck topoi in terms of *localizations* (Property 12.2.10) is equivalent to a definition in terms of reflections (this is due to *Street*).

Corollary 12.4.27. There exists a bijection between the Grothendieck topologies on a small category $\mathbf{C}$ and the equivalence classes of left exact reflective subcategories of $\mathbf{Psh}(\mathbf{C})$.

Definition 12.4.28 (Representable morphism). A natural transformation of presheaves $F \to G$ is said to be representable if, for every representable presheaf h_X and every morphism $h_X \to G$, the pullback $h_X \times_G F$ is representable.

Property 12.4.29 (Diagonals). The diagonal morphism $\Delta_F : F \to F \times F$, for F a presheaf on a category with pullbacks, is representable if and only if any of the following two equivalent properties holds:

1. For every two representable presheaves h_X, h_Y and natural transformations $h_X \to F, h_Y \to F$, the pullback $h_X \times_F h_Y$ is representable.
2. Every natural transformation $h_X \to F$ from a representable presheaf is representable.

12.4.2 Topological sheaves

See Chapter 8 for the application of sheaves to topology.

Property 12.4.30 (Presheaf topos). Consider the presheaf category

$$\mathbf{Psh}(X) := \mathbf{Psh}\big(\mathbf{Open}(X)\big) \tag{12.23}$$

over a topological space (X, τ). Unpacking Property 12.4.4 shows that this category is an elementary topos where the subobject classifier is given by

$$\Omega(U) = \{V \in \tau \mid V \subseteq U\}\,. \tag{12.24}$$

Construction 12.4.31 (Sheaves and étale bundles). Let X be a topological space. The functor

$$I : \mathbf{Open}(X) \to \mathbf{Top}_{/X} : U \mapsto (U \hookrightarrow X) \tag{12.25}$$

induces the following adjunction:

$$\mathbf{Top}_{/X} \underset{\Gamma}{\overset{E}{\underset{\longrightarrow}{\overset{\longleftarrow}{\perp}}}} \mathbf{Psh}(X)\,. \tag{12.26}$$

The slice category on the left-hand side is the category of (topological) *bundles* (see Chapter 33) over X. Both directions of the adjunction have a clear interpretation. The right adjoint assigns to every bundle its sheaf of local sections and the left adjoint assigns to every presheaf its bundle of germs.

By restricting to the subcategories on which this adjunction becomes an adjoint equivalence, one obtains the **étale space** and **sheaf categories** respectively:

$$\mathbf{Et}(X) \cong \mathbf{Sh}(X)\,. \tag{12.27}$$

The category on the right-hand side is the category of sheaves on a topological space X. The category on the left is the full subcategory on local homeomorphisms, i.e. the étale spaces (Definition 6.2.19).

Property 12.4.32 (Associated sheaf). The inclusion functor $\mathbf{Sh}(X) \hookrightarrow \mathbf{Psh}(X)$ admits a left adjoint, the **sheafification functor**, that assigns to every presheaf its associated sheaf. This functor is given by the composition $\Gamma \circ E$, which is simply Construction 8.2.15.

The fact that the counit of the adjunction (Construction 12.4.31) restricts to an isomorphism on the full subcategory $\mathbf{Sh}(X)$ is equivalent to the fact that the sheafification of a sheaf Γ is again Γ.

Definition 12.4.33 (Petit and gros topoi[5]). Consider a topological space X together with its category of opens $\mathbf{Open}(X)$. The petit topos over X is defined as the sheaf topos $\mathbf{Sh}(X) \equiv \mathbf{Sh}\big(\mathbf{Open}(X)\big)$. It represents X as a 'generalized space'. (By Construction 12.4.31, the objects in a petit topos are the étale spaces over the given base space.) Topoi equivalent to such petit topoi are sometimes said to be **spatial**. However, one can also build a topos whose objects are themselves generalized spaces. To this end, choose a site S of 'probes' and call the sheaf topos $\mathbf{Sh}(S)$ a gros topos. (See Section 44.4 for more information.)

Property 12.4.34 (Localic reflection). Mapping a topological space to its sheaf of continuous sections defines a functor $\mathbf{Sh} : \mathbf{Top} \to \mathbf{Topos}$ by Example 12.2.12. When

[5]For those that do not master French, '*petit*' and '*gros*' mean small and big, respectively.

restricted to the full subcategory of sober spaces (Definition 6.8.4), this functor becomes fully faithful. Generalizing to locales even gives a reflective inclusion (Definition 3.2.29).

This property states that no information is lost when regarding (sober) topological spaces as sheaf topoi. This also explains the name 'petit topos'.

Definition 12.4.35 (Localic topos). Multiple equivalent definitions exist:

1. A topos equivalent to a sheaf topos over a locale (Definition 6.8.2) equipped with the topology of jointly surjective morphisms.
2. A topos generated under colimits of subobjects of the terminal object 1.
3. A topos $\mathcal{E}$ for which the global sections functor $\Gamma : \mathcal{E} \to \mathbf{Set}$ is localic, i.e. every object in $\mathcal{E}$ is a subquotient of an object in the inverse image Γ^*

Given a geometric morphism to some base topos $\mathcal{S}$, one can define $\mathcal{S}$-localic topoi by generalizing the third point.

The following property shows that the locale in the first definition has a specific meaning.

Property 12.4.36. By Property 12.3.2, for every topos the poset $\mathsf{Sub}(1)$ is a locale. Every localic topos $\mathcal{E}$ satisfies $\mathcal{E} \cong \mathsf{Sh}\big(\mathsf{Sub}(1)\big)$, where $\mathsf{Sub}(1)$ is equipped with the topology of jointly surjective morphisms.

The equivalence between localic topoi and locales carries over to the notion of $\mathcal{S}$-localic topoi.

Property 12.4.37. The (2-)category of localic topoi over a base topos $\mathcal{S}$ is equivalent to the (2-)category $\mathbf{Loc}(\mathcal{S})$ of locales internal to $\mathcal{S}$.

Property 12.4.38. Given a locale X, the category $\mathbf{Loc}\big(\mathbf{Sh}(X)\big)$ is equivalent to the slice category $\mathbf{Loc}_{/X}$.

12.4.3 Lawvere–Tierney topology

This section gives an alternative approach to the construction of sheaf topoi, which is more closely linked to the logical aspect of topos theory.

Definition 12.4.39 (Lawvere–Tierney topology). As noted in Section 12.3 on the internal logic of elementary topoi, the subobject classifier Ω has the structure of an internal Heyting algebra and, in particular, that of a (bounded) meet-semilattice. This internal poset, viewed as an internal category, admits the construction of a closure op-

erator $j : \Omega \to \Omega$ (Definition 3.3.27) satisfying the following condition:

$$j \circ \wedge = \wedge \circ (j \times j)\,. \tag{12.28}$$

This condition states (in a nontrivial way) that j is (internally) order-preserving. More concisely, a Lawvere–Tierney topology is internally a left exact modality on Ω.

Remark 12.4.40. The condition satisfied by the unit morphism in the definition of a closure operator can also be reformulated in this context as follows:

$$j \circ \mathtt{true} = \mathtt{true}\,. \tag{12.29}$$

Example 12.4.41 (Double negation topology). As mentioned above, the subobject classifier in an elementary topos has the structure of an internal Heyting algebra and, hence, admits a negation morphism $\neg : \Omega \to \Omega$. It can be shown that double negation $\neg\neg : \Omega \to \Omega$ gives a Lawvere–Tierney topology.

Property 12.4.42. The Booleanization $\mathcal{E}_{\neg\neg} \hookrightarrow \mathcal{E}$ of an elementary subtopos is the smallest dense subtopos of $\mathcal{E}$ and the largest Boolean subtopos of $\mathcal{E}$, i.e. it is the unique subtopos such that:

1. **Dense**: $\mathcal{E}_{\neg\neg}$ contains the initial object $\perp$.
2. **Boolean**: Ω is a Boolean algebra.

As noted above, Lawvere–Tierney operators induce, internally, *universal closure operators* on all posets $\mathsf{Sub}(x)$ in the topos. Given an object x and a subobject $u \in \mathsf{Sub}(x)$, one defines the closure $j_*(u) \in \mathsf{Sub}(x)$ as the subobject classified by the characteristic morphism $j \circ \chi_u : x \to \Omega$.

Definition 12.4.43 (Dense object). Given a Lawvere–Tierney topology $j : \Omega \to \Omega$, a subobject $u \in \mathsf{Sub}(x)$ is said to be dense (in x) if it satisfies $j_*(u) \cong x$. An object is said to be **closed** if $j_*(u) \cong u$.

Example 12.4.44. On a presheaf topos, objects are dense exactly if they are covering sieves (Definition 12.4.5).

The following definition allows to generalize sheaves from topoi to finitely complete categories.

Alternative Definition 12.4.45 (Sheaf). Given a Lawvere–Tierney topology $j : \Omega \to \Omega$ on a topos $\mathcal{E}$, one calls an object $s \in \mathsf{ob}(\mathcal{E})$ a j-sheaf if it is local with respect to all dense morphisms (Definition 3.2.21), i.e. for all dense morphisms $u \hookrightarrow x$ the induced map

$$\mathcal{E}(x, s) \to \mathcal{E}(u, s)$$

is a bijection. If this map is only a monomorphism, the object is said to be **j-separated**.

Property 12.4.46. Lawvere–Tierney topologies on $\mathcal{E}$ are in correspondence with subtopoi of $\mathcal{E}$, given by the sheaves relative to the chosen topology. As before, the left adjoint to the embedding is the sheafification functor.

In the case of (pre)sheaf topoi, Lawvere–Tierney topologies recover Grothendieck topologies.

Corollary 12.4.47. For the presheaf topos on a small category $\mathbf{C}$, the Grothendieck topologies on $\mathbf{C}$ are equivalent to Lawvere–Tierney topologies on $\mathbf{Psh}(\mathbf{C})$.

Proof (Sketch of proof). Since a Grothendieck topology assigns to every object a collection of sieves, Property 12.4.4 implies that $J(x) \subseteq \Omega_{\mathbf{Psh}}(x)$ for all $x \in \mathsf{ob}(\mathbf{C})$. By the base change condition of Grothendieck topologies, this relation is natural in x and, hence, J is a subobject of $\Omega_{\mathbf{Psh}}$. One thus finds a characteristic morphism $j : \Omega_{\mathbf{Psh}} \to \Omega_{\mathbf{Psh}}$ that can be proven (by the other conditions of Grothendieck topologies) to define a Lawvere–Tierney topology on $\mathbf{Psh}(\mathbf{C})$. Conversely, a Lawvere–Tierney topology is a morphism $j : \Omega \to \Omega$ and, hence, determines a unique subobject of $\Omega_{\mathbf{Psh}}$, i.e. a unique collection of sieves for every object $x \in \mathsf{ob}(\mathbf{C})$. From the conditions of Lawvere–Tierney topologies, one can then prove that this collection satisfies the conditions of a Grothendieck topology. □

12.4.4 Diaconescu's theorem

The following concept weakens the notion of left exact functors to categories where not all (finite) limits exist.

Definition 12.4.48 (Flat functor). A functor $F : \mathbf{C} \to \mathbf{Set}$ is said to be flat if the opposite of the category of elements $\mathsf{El}(F)^{\mathrm{op}}$ (Definition 3.3.15) is filtered (Definition 3.4.50). The category of flat functors on $\mathbf{C}$ is denoted by $\mathbf{Flat}(\mathbf{C}, \mathbf{Set})$.

A general functor $F : \mathbf{C} \to \mathbf{D}$ is said to be **(representably) flat** if, for every object $d \in \mathsf{ob}(\mathbf{D})$, the opposite comma category $(d/F)^{\mathrm{op}}$, which is equivalently the opposite category of elements $\mathsf{El}\big(\mathbf{D}(d, -) \circ F\big)^{\mathrm{op}}$, is filtered.

The definitions above can be stated more explicitly as follows (cf. Definition 3.4.64). The functor $F : \mathbf{C} \to \mathbf{D}$ is representably flat if for every $d \in \mathsf{ob}(\mathbf{D})$:

1. There is at least one object $c \in \mathsf{ob}(\mathbf{C})$ and a morphism $f : d \to Fc$. Hence, at least one inhabited set in the image of F.

2. For every two objects $c, c' \in \mathsf{ob}(\mathbf{C})$ and every two morphisms $f \in \mathbf{D}(d, Fc)$ and $g \in \mathbf{D}(d, Fc')$, there exists an object $x \in \mathsf{ob}(\mathbf{C})$ and morphisms $i \in \mathbf{C}(x, c)$, $j \in \mathbf{C}(x, c')$ and $h \in \mathbf{D}(d, Fx)$ such that $Fi \circ h = f$ and $Fj \circ h = g$.

3. For every two parallel morphisms $f, g \in \mathbf{C}(c, c')$ and morphism $h \in \mathbf{D}(d, Fc)$ such that $Ff \circ h = Fg \circ h$, there exists a morphism $i \in \mathbf{C}(x, c)$ and a morphism $j \in \mathbf{D}(d, Fx)$ such that $f \circ i = g \circ i$ and $Fi \circ j = h$.

Remark 12.4.49. Note that flatness and representable flatness are only equivalent for **Set**-valued functors when the domain **C** is finitely complete.

Property 12.4.50. Consider the **Yoneda extension** that maps a functor $F \in [\mathbf{C}, \mathbf{Set}]$ to a functor $\tilde{F} : [\mathbf{C}^{\mathrm{op}}, \mathbf{Set}] \to \mathbf{Set}$ through left Kan extension along the Yoneda embedding. F is flat if and only if $\tilde{F}$ is left exact.[6]

Property 12.4.51. A functor $F : \mathbf{C} \to \mathbf{Set}$ is flat if and only if it is the filtered colimit of representable functors or, equivalently, if F-weighted filtered colimits commute with finite limits. Note that, by Universal Property 3.4.56, this means that

$$\mathbf{Flat}(\mathbf{C}^{\mathrm{op}}, \mathbf{Set}) \cong \mathbf{Ind}(\mathbf{C}) . \tag{12.30}$$

The previous property implies the following statement, which is the **Set**-valued version of *Diaconescu's theorem*, which will be covered in its internal version below.

Theorem 12.4.52 (Diaconescu: Set version). *A functor $F : \mathbf{C} \to \mathbf{Set}$ is flat if and only if the induced adjunction $\mathbf{Set} \leftrightarrows [\mathbf{C}^{\mathrm{op}}, \mathbf{Set}]$ is a geometric morphism or, alternatively, if flat functors are equivalent to global points of the presheaf topos $[\mathbf{C}^{\mathrm{op}}, \mathbf{Set}]$.*

Definition 12.4.53 (Morphism of sites). A representably flat functor preserving covers.

Now, for $\mathcal{E}$ a cocomplete topos, a functor $F : \mathbf{C} \to \mathcal{E}$ is said to be **(internally) flat** if it representably flat with respect to the internal logic of $\mathcal{E}$ or, equivalently, if the Yoneda extension $\mathsf{Lan}_y F$ is left exact.

Property 12.4.54. A functor $F : \mathbf{C} \to \mathcal{E}$, with $\mathcal{E}$ a cocomplete topllos, is flat if and only if:

- The family of terminal morphisms in **C** is jointly epimorphic.
- For all $c, d \in \mathsf{ob}(\mathbf{C})$, the family of morphisms into $Ac \times Ad$ induced by spans $c \leftarrow b \to d$ is jointly epimorphic.
- For any two parallel morphisms f, g in **C**, the family of factorizations of morphisms through their equalizer is jointly epimorphic.

These conditions are equivalent to the following ones:

[6] By Definition 3.4.86, Kan extensions admit an expression is terms of adjoints. In this setting, the right adjoint has the form of a Hom-functor and, accordingly, the Yoneda extension is sometimes written as a tensor product $- \otimes_{\mathbf{C}} F$. The terminology 'flat' is then in analogy with Definition 5.4.12. This essentially coincides with the tensor product of functors Definition 4.2.20.

- For every object $x \in \mathcal{E}$, there exists a jointly epimorphic family $\{x_i \to x\}_{i\in I}$ and and a family $\{x_i \to F(c_i) \mid c_i \in \mathsf{ob}(\mathbf{C})\}_{i\in I}$.
- For all $c, d \in \mathsf{ob}(\mathbf{C})$ and generalized elements $\lambda : x \to Fc \times Fd$, there exists an epimorphic family $\{\lambda_i : x_i \to x\}_{i\in I}$ and a family of spans $\{c \xleftarrow{f} b_i \xrightarrow{g} d\}_{i\in I}$ equipped with a family of generalized elements $\{\kappa_i : x_i \to Fb_i\}_{i\in I}$ such that
$$(Ff, Fg) \circ \kappa_i = \lambda \circ \lambda_i \,. \tag{12.31}$$
- For any two parallel morphisms $f, g : c \to d$ in $\mathbf{C}$ and any generalized element $\lambda : x \to Fc$ such that $Ff \circ \lambda = Fg \circ \lambda$, there exists an epimorphic family $\{\lambda_i : x_i \to x\}_{i\in I}$ and families of morphisms $\{f_i : c_i \to c\}$ and $\{\kappa_i : x_i \to Fc_i\}$ such that
$$\begin{aligned} f \circ f_i &= g \circ f_i \\ Ff_i \circ \kappa_i &= \lambda \circ \lambda_i \end{aligned} \tag{12.32}$$
for all $i \in I$.

Theorem 12.4.55 (Diaconescu: Topos version). *Consider a cocomplete topoi $\mathcal{E}$ and a (small) category $\mathbf{C}$. There exists an equivalence*
$$\mathbf{Topos}\big(\mathcal{E}, [\mathbf{C}^{\mathsf{op}}, \mathbf{Set}]\big) \leftrightarrows \mathbf{Flat}(\mathbf{C}, \mathcal{E}) \,, \tag{12.33}$$
where

- *flat functors are sent to geometric morphisms whose inverse image part is given by left Kan extension along the Yoneda embedding and direct image part is given by:*
$$F \mapsto \big(x \mapsto \mathcal{E}(F-, x)\big) \,, \tag{12.34}$$
and
- *geometric morphisms are sent to the precomposition of the inverse image part with the Yoneda embedding.*

To pass to the internal logic of topoi, Definition 3.4.50 needs to be generalized to internal category theory (Section 3.5).

Definition 12.4.56 (Internally filtered category). An internal category $\mathbf{C} \in \mathbf{Cat}(\mathcal{E})$ satisfying:[7]

1. The terminal morphism $C_0 \twoheadrightarrow 1$ is an epi.
2. For any two generalized elements $\lambda_1, \lambda_2 : U \to C_0$, there is an epi $\theta : V \twoheadrightarrow U$ and corresponding V-elements $\gamma_1, \gamma_2 : V \to C_0$ such that
$$\begin{aligned} d_1 \circ \gamma_1 &= d_1 \circ \gamma_2 \\ d_0 \circ \gamma_i &= \lambda_i \circ \theta \end{aligned} \tag{12.35}$$

[7] Some sources require all these morphisms to be regular epis.

for $i = 1, 2$.

3. For any two parallel generalized elements $\gamma_1, \gamma_2 : U \to C_1$ such that $d_0 \circ \gamma_1 = d_0 \circ \gamma_2$, there exists an epi $\theta : V \twoheadrightarrow U$ and a V-element $\kappa : V \to C_1$ such that

$$\begin{aligned} d_1 \circ \kappa = d_0 \circ \lambda_1 \circ \theta = d_0 \circ \lambda_2 \circ \theta \\ c(\kappa, \lambda_1 \circ \theta) = c(\kappa, \lambda_2 \circ \theta)\,. \end{aligned} \tag{12.36}$$

Chapter 13

Model theory ♣

General references for this chapter are Hovey (2007); Riehl (2019). For more on monoidal model categories, see Riehl (2013). A good reference for the section on simplicial spaces and, in particular, the theory of Segal spaces is Rezk (2001). For more on Reedy model structures, see Riehl and Verity (2013). A gentle introduction to the theory of homotopy (co)limits can be found in Lambrechts (2013); Riehl (2011). The content and language of Chapter 5 will be used throughout this chapter.

Chapter Contents

13.1 Simplicial sets **282**
13.1.1 Constructions 283
13.1.2 Homological algebra 285
13.2 Localization **286**
13.3 Model categories **288**
13.3.1 Model structures 288
13.3.2 Monoidal structures 290
13.3.3 Homotopy category 291
13.3.4 Reedy model structure ♣ 294
13.4 Simplicial spaces **297**
13.4.1 Kan complexes 297
13.4.2 Simplicial localization 298
13.4.3 Segal spaces 299
13.4.4 Coherent nerve 301
13.5 Cofibrantly generated categories **303**
13.5.1 Transfinite constructions 303
13.5.2 Cofibrant generation 305

13.6 Homotopy (co)limits 307
13.6.1 Simplicially enriched diagrams 308

13.1 Simplicial sets

Definition 13.1.1 (Simplex category). The simplex category $\mathbf{\Delta}$ has as objects the posets of the form $[n] := \{0 < \cdots < n\}$ and as morphisms the order-preserving maps.

Definition 13.1.2 (Simplicial set). The category **sSet** of simplicial sets is defined as the presheaf category $\mathbf{Psh}(\mathbf{\Delta})$. For all $n \in \mathbb{N}$, the set of **n-simplices in X** is defined as the set $X_n := X([n])$.

Definition 13.1.3 (Simplicial object). By internalizing the notion of a simplicial set, one obtains the definition of a simplicial object, i.e. a simplicial object in a category $\mathbf{C}$ is a $\mathbf{C}$-valued presheaf on $\mathbf{\Delta}$.

Remark 13.1.4. Note that the notion of **simplicial category** can mean two distinct things. In general, it will mean a category enriched in **sSet**. However, following the previous definition, it can also mean a simplicial object in the (2-)category **Cat**. It can be shown that all simplicially enriched categories are a specific kind of degenerate simplicial object in **Cat**, where the face and degeneracy maps are identity-on-objects.

Definition 13.1.5 (Standard simplex). For every $n \in \mathbb{N}$, the standard n-simplex $\Delta[n]$ is defined as the Yoneda embedding $\mathbf{\Delta}(-, [n])$. One can also define a functor $\Delta_{\text{top}} : \mathbf{\Delta} \to \mathbf{Top}$ that maps $[n]$ to the standard topological n-simplex Δ^n (Example 7.2.3).

Property 13.1.6. By the Yoneda lemma, there exists a natural bijection between the set of n-simplices of a simplicial set X and the set of maps $\Delta[n] \to X$.

Property 13.1.7 (Face and degeneracy maps). All morphisms in the simplex category $\mathbf{\Delta}$ are generated by morphisms of the following two types:

- For every n and $i < n$, the unique map $\delta_{n,i} : [n-1] \to [n]$ that misses the ith element.
- For every n and $i \leq n$, the unique map $\sigma_{n,i} : [n+1] \to [n]$ that duplicates the ith element.

Under the action of a presheaf, this gives the **face** and **degeneracy** maps $d_{n,i}$ and $s_{n,i}$. (If the index n is clear, it is often omitted.)

These morphisms satisfy the fundamental **simplicial identities**:

- $d_i \circ d_j = d_{j-1} \circ d_i$ for $i < j$,
- $d_i \circ s_j = s_{j-1} \circ d_i$ for $i < j$,

- $d_i \circ s_j = \mathbb{1}$ for $i = j$ or $i = j + 1$,
- $d_i \circ s_j = s_j \circ d_{i-1}$ for $i > j + 1$, and
- $s_i \circ s_j = s_{j+1} \circ s_i$ for $i \leq j$.

Definition 13.1.8 (Connected components). Consider a simplicial set X. Its set of connected components $\pi_0(X)$ is defined as the quotient of X_0 under the relation

$$X_1 \underset{d_1}{\overset{d_0}{\rightrightarrows}} X_0 \times X_0 \,. \tag{13.1}$$

This defines a fucntor $\pi_0 : \mathbf{sSet} \to \mathbf{Set}$. By base change 4.2.3, this also induces a functor on simplicial categories.

13.1.1 Constructions

Construction 13.1.9 (Nerve and realization). Consider a general functor $F : \mathbf{S} \to \mathbf{C}$ into a cocomplete category ($\mathbf{S}$ will often be a category of geometric shapes such as the simplex category $\mathbf{\Delta}$ or the *cube category* ⧈). Every such functor induces an adjunction

$$\mathbf{C} \underset{N}{\overset{|-|}{\underset{\longrightarrow}{\overset{\longleftarrow}{\perp}}}} \mathbf{Psh}(\mathbf{S}) \,. \tag{13.2}$$

The **realization functor** $|-|$ is defined as the left Kan extension $\mathsf{Lan}_y F$. The **nerve functor** $N : \mathbf{C} \to \mathbf{Psh}(\mathbf{S})$ is defined as $Nx := \mathbf{C}(F-, x) = \mathcal{Y}x \circ F$.

This definition can easily be generalized to the enriched setting. Furthermore, if one assumes that $\mathbf{C}$ is copowered over $\mathcal{V}$, the realization functor can be expressed as a coend:

$$|X| = \int^{s \in \mathbf{S}} Xs \cdot Fs \,. \tag{13.3}$$

Example 13.1.10 (Nerve of a category). To every small category $\mathbf{C}$, one can associate a simplicial set $N\mathbf{C}$ in the following way. The set $N\mathbf{C}_0$ is given by the set of objects in $\mathbf{C}$ and the set $N\mathbf{C}_1$ is given by the set of morphisms in $\mathbf{C}$. Now, for every two composable morphisms f, g, one obtains a canonical commuting triangle by composition. Let $N\mathbf{C}_2$ be the set of all these triangles. The higher simplices are defined analogously. Face maps act by composing morphisms or by dropping the exterior morphisms in a composable string. Degeneracy maps act by inserting an identity morphism.

Equivalently, one can define the **simplicial nerve functor** in the following way. Every poset $[n]$ admits a canonical category structure for which the order-preserving maps give rise to the associated functors. By the above construction, this inclusion $\mathbf{\Delta} \hookrightarrow \mathbf{Cat}$ induces a nerve functor

$$N : \mathbf{Cat} \to \mathbf{sSet} : \mathbf{C} \mapsto \mathbf{Cat}(-, \mathbf{C}) \,. \tag{13.4}$$

This way, one obtains $N\mathbf{C}_k = \mathbf{Cat}([k], \mathbf{C})$. This object is, by definition, equivalent to the collection of all strings of k composable morphisms in $\mathbf{C}$. It can be shown that the simplicial nerve functor is fully faithful.

Example 13.1.11 (Geometric realization). Consider a simplicial set X. From this object, one can construct a topological space as follows. First, take a point for every element in X_0. Then, glue 1-simplices between these points using the face maps. The higher (nondegenerate) simplices are attached analogously.

More abstractly, the geometric realization functor $|\cdot| : \mathbf{sSet} \to \mathbf{Top}$ is defined as a (left) Kan extension:

$$|\cdot| := \mathsf{Lan}_y \Delta_{\mathsf{top}} . \tag{13.5}$$

An application of the Yoneda lemma shows that the geometric realization can be expressed as a functor tensor product (Definition 4.2.8):

$$|X| = X \otimes_{\mathbf{\Delta}} \Delta_{\mathsf{top}} = \int^{n \in \mathbf{\Delta}} \Delta^n \cdot X_n . \tag{13.6}$$

This formula can easily be generalized to the category of simplicial topological spaces ($\mathbf{sSet}$ is a full subcategory obtained by endowing every set with the discrete topology). In $\mathbf{Top}$, the coend can be expressed as the quotient space

$$|X| := \bigsqcup_{n \in \mathbb{N}} X_n \times \Delta^n / \sim , \tag{13.7}$$

where the equivalence relation identifies the points $(x, f_* y)$ and $(f^* x, y)$ for all morphisms $f \in \mathsf{hom}(\mathbf{\Delta})$. The morphisms f^*, f_* are the ones induced by X and $\Delta_{\mathsf{top}} : \mathbf{\Delta} \hookrightarrow \mathbf{Top}$. As an immediate example one obtains

$$|\Delta[n]| = \Delta^n , \tag{13.8}$$

which shows that $\Delta[n]$ really deserves to be called the standard n-simplex.

Remark 13.1.12. Note that the composition of the geometric realization and (simplicial) nerve functors $|N \cdot| : \mathbf{Cat} \to \mathbf{Top}$ is also often called the **geometric realization** (of categories).

Example 13.1.13 (Singular set). Given a topological space X, one can define a simplicial set $\mathsf{Sing}(X)$. Its components are defined as the set of morphisms from the standard (topological) n-simplex to X:

$$\mathsf{Sing}(X)_n := \mathbf{Top}(\Delta^n, X) . \tag{13.9}$$

This is the object of relevance in the definition of singular (co)homology as given in Section 7.3.

Property 13.1.14 (Classifying space). For a (discrete) group G, one can construct two important objects: the delooping $\mathbf{B}G$ (Example 3.5.11) and the *classifying space* BG (see Definition 35.2.1). As their notations imply, there exists a relation between these space. By taking the geometric realization of $\mathbf{B}G$, one obtains BG. In fact, this method can be applied to any monoid A to obtain the so-called (two-sided) **bar construction**.

Construction 13.1.15 (Join). Consider two simplicial sets X, Y. An explicit construction of the join $X \star Y$ is as follows:

$$(X \star Y)_n := X_n \sqcup Y_n \sqcup \bigsqcup_{i+j=n-1} X_i \times Y_j \,. \tag{13.10}$$

The boundary maps on the first two components are those of X and Y. On the third component it acts by

$$\mathsf{d}_i(x,y) := \begin{cases} (\mathsf{d}_i x, y) & \text{if } i \leq j, j \neq 0\,, \\ (x, \mathsf{d}_{i-j-1} y) & \text{if } i > j, k \neq 0\,, \end{cases} \tag{13.11}$$

where $x \in X_j$ and $y \in Y_k$. For $j = 0$ and $k = 0$, the projections are used.

13.1.2 Homological algebra

In this section, simplicial sets are related to homological algebra (Chapter 5). A basic introduction is Master (2020).

Construction 13.1.16 (Alternating face map complex). From a simplicial Abelian group A, one can construct a connective chain complex as follows. For every $n \in \mathbb{N}$, define

$$(CA)_n := A_n \,. \tag{13.12}$$

The boundary maps δ_n are defined as the alternating sum of the face maps:

$$\delta_n := \sum_{i=1}^{n} (-1)^i dr_i \,. \tag{13.13}$$

Every group A_{n+1} contains a subgroup $D(A_n)$ generated by the degeneracy maps:

$$D(A_n) := \left\langle \bigcup_{i=1}^{n} s_i(A_n) \right\rangle . \tag{13.14}$$

If these degenerate simplices are quotiented out, the **normalized complex** is obtained.

This construction can be generalized to any simplicial group:

Construction 13.1.17 (Moore complex). Let G be a simplicial group. For every $n \in \mathbb{N}$, define

$$(NG)_n := \bigcap_{i=1}^{n} \ker(\mathsf{d}_i^n) \,. \tag{13.15}$$

The differential ∂_n is given by the zeroth face map d_0^n.

Property 13.1.18 (Equivalence). For simplicial Abelian groups, the Moore complex and normalized complex are isomorphic. Moreover, the inclusion of the normalized complex into the alternating face map complex is a quasi-isomorphism.

Theorem 13.1.19 (Dold–Kan correspondence). *The functor that maps simplicial Abelian groups to normalized chain complexes gives an equivalence of categories* $\mathbf{sAb} \cong \mathbf{Ch}^+(\mathbf{Ab})$.

13.2 Localization

Definition 13.2.1 (Category with weak equivalences). A category $\mathbf{C}$ with a subcategory $\mathbf{W}$ such that:

1. $\mathbf{W}$ contains all isomorphisms in $\mathbf{C}$ (in particular, $\mathbf{W}$ is wide).
2. $\mathbf{W}$ satisfies the '2-out-of-3 property': If any two of $\{f, g, f \circ g\}$ are in $\mathsf{hom}(\mathbf{W})$, so is the third.

Definition 13.2.2 (Weak factorization system). Consider a category $\mathbf{C}$. A pair (L, R) of classes of morphisms in $\mathbf{C}$ is called a weak factorization system (WFS) if it satisfies the following properties:

1. Every morphism in $\mathbf{C}$ factorizes as $g \circ f$ where $f \in L$ and $g \in R$.
2. L consists of exactly those morphisms in $\mathbf{C}$ that have the left lifting property (Definition 3.4.29) with respect to morphisms in R.
3. R consists of exactly those morphisms in $\mathbf{C}$ that have the right lifting property with respect to morphisms in L.

Remark. The original definition by *Quillen* only required that L and R satisfied the lifting properties with respect to each other, not that they were closed under this condition. This was enforced by a further condition that both L and R are closed under retracts in the arrow categories. It can be proven that this is equivalent to the definition above.

Definition 13.2.3 (Homotopical category). A category $\mathbf{C}$ equipped with a subcategory $\mathbf{W}$ such that:

1. $\mathbf{W}$ contains all identity morphisms in $\mathbf{C}$ (in particular, $\mathbf{W}$ is wide).
2. $\mathbf{W}$ satisfies the '2-out-of-6 property': If $f \circ g$ and $g \circ f$ are in $\mathsf{hom}(\mathbf{W})$, so are f, g, h and $f \circ g \circ h$.

It is not hard to see that every homotopical category is, in particular, a category with weak equivalences.

Definition 13.2.4 (Homotopical functor). A functor (between homotopical categories)

that preserves weak equivalences.

Definition 13.2.5 (Gabriel–Zisman localization). Consider a category $\mathbf{C}$ with a collection of morphisms $M \subset \mathsf{hom}(\mathbf{C})$. The localization of $\mathbf{C}$ with respect to M is constructed by adding for each morphism $f \in M$ a formal inverse to $\mathsf{hom}(\mathbf{C})$.

More specifically, the localization consists of a category $\mathbf{C}[M^{-1}]$ and a functor $F_M : \mathbf{C} \to \mathbf{C}[M^{-1}]$ inverting M, i.e. mapping all morphisms in M to isomorphisms, that is universal with respect to this property.

When $\mathbf{C}$ has finite products, the localization $C[x^{-1}]$ at an object $x \in \mathsf{ob}(\mathbf{C})$ is given by the localization at the projections $M_x := \{y \times x \to y \mid y \in \mathsf{ob}(\mathbf{C})\}$.

Definition 13.2.6 (Homotopy category I). When $\mathbf{C}$ is a category with weak equivalences W, the localization $\mathbf{C}[W^{-1}]$ is called the homotopy category $\mathbf{Ho}(\mathbf{C})$. In this context, the localization functor $\mathbf{C} \to \mathbf{Ho}(\mathbf{C})$ is sometimes denoted by $\gamma_{\mathbf{C}}$.

Remark 13.2.7 (Size issues). When $\mathbf{C}$ is small, so is its localization. However, even in the case where $\mathbf{C}$ is locally small, its localization might be large.

Property 13.2.8 (Reflective localization). Consider an adjunction $F \dashv G : \mathbf{C} \to \mathbf{D}$. The right adjoint is fully faithful, i.e. determines a reflective subcategory (Definition 3.2.29), if and only if F realizes $\mathbf{D}$ as a localization of $\mathbf{C}$. The essential image of G consists exactly of the local objects in C.

Definition 13.2.9 (Derived functor). Consider a homotopical functor $F : \mathbf{C} \to \mathbf{D}$ and consider the composition $\gamma := \gamma_{\mathbf{D}} \circ F$ with the localization functor $\gamma_{\mathbf{D}} : \mathbf{D} \to \mathbf{Ho}(\mathbf{D})$. The derived functor $\mathbf{Ho}(F) : \mathbf{Ho}(\mathbf{C}) \to \mathbf{Ho}(\mathbf{D})$ is obtained by applying the universal property of $\mathbf{Ho}(\mathbf{C})$ to γ.

This definition can be rephrased in terms of Kan extensions. Consider a homotopical functor $F : \mathbf{C} \to \mathbf{D}$. The left and right derived functors are defined as the following Kan extensions:

$$\begin{aligned} LF &:= \mathsf{Ran}_{\gamma_{\mathbf{C}}}(\gamma_{\mathbf{D}} \circ F)\,, \\ RF &:= \mathsf{Lan}_{\gamma_{\mathbf{C}}}(\gamma_{\mathbf{D}} \circ F)\,. \end{aligned} \tag{13.16}$$

In fact, one can drop the assumption that $\mathbf{D}$ has weak equivalences (here, F should map weak equivalences to isomorphisms). In this case, one simply has to replace $\gamma_{\mathbf{D}} \circ F$ by F in the above formulas.

Example 13.2.10 (Derived category). Consider an Abelian category $\mathbf{C}$ together with its category of chain complexes $\mathbf{Ch}(\mathbf{C})$. The derived category $\mathcal{D}(\mathbf{C})$ is defined as the localization of $\mathbf{Ch}(\mathbf{C})$ at the quasi-isomorphisms.

Remark 13.2.11. In this case, it can be shown that one can first restrict to the naive homotopy category $\mathbf{K}(\mathbf{C})$, consisting of chain complexes and chain maps up to chain homotopy, and then localize at the collection of quasi-isomorphisms.

13.3 Model categories

13.3.1 Model structures

Definition 13.3.1 (Model structure). Let $\mathbf{C}$ be a category. A (**Quillen**-)model structure on $\mathbf{C}$ consists of three classes of morphisms:

- **weak equivalences** W,
- **fibrations** Fib, and
- **cofibrations** Cof,

that satisfy the following two conditions:

1. W turns $\mathbf{C}$ into a category with weak equivalences (Definition 13.2.1).
2. $(\mathsf{Cof}, \mathsf{Fib} \cap W)$ and $(\mathsf{Cof} \cap W, \mathsf{Fib})$ are weak factorization systems (Definition 13.2.2).

The morphisms in $\mathsf{Fib} \cap W$ and $\mathsf{Cof} \cap W$ are said to be **acyclic** or **trivial**.

Remark. That W contains all isomorphisms follows from the property that any class of morphisms satisfying a lifting property contains all isomorphisms.

Definition 13.3.2 (Model category). A bicomplete category equipped with a model structure.[1]

Definition 13.3.3 (Proper model category). A model category is said to be left proper (resp. right proper) if weak equivalences are preserved by pushouts along cofibrations (resp. pullbacks along fibrations).

Definition 13.3.4 (Fibrant object). An object in a model category for which the terminal morphism is a fibration. Dually, an object in a model category is said to be cofibrant if the initial morphism is a cofibration.

Example 13.3.5 (Model structure on functor categories). In some cases, the functor category $[\mathbf{C}, \mathbf{D}]$, for $\mathbf{C}$ small and $\mathbf{D}$ a model category, admits two canonical model structures:

- **Injective model structure**: The weak equivalences are the natural transformations that are objectwise weak equivalences and the cofibrations are the natural transformations that are objectwise cofibrations.
- **Projective model structure**: The weak equivalences are the natural transformations that are objectwise weak equivalences and the fibrations are the natural transformations that are objectwise fibrations.

[1] *Quillen*'s original definition only required the existence of finite limits and finite colimits.

Property 13.3.6 (Resolution). In a model category, the bicompleteness property implies that the initial and terminal object always exist. The weak factorization property then implies that, for every object x, one can find a weakly equivalent fibrant replacement x^{f} and a weakly equivalent cofibrant replacement x^{cf} by suitably factorizing the initial and terminal morphisms. These replacements are sometimes called **resolutions** or **approximations**.

If the weak factorization system is functorial (Definition 3.3.5), (co)fibrant replacement defines an endofunctor that is weakly equivalent to the identity functor.

Definition 13.3.7 (Quillen adjunction). Let $\mathbf{C}, \mathbf{D}$ be two model categories. An adjunction

$$\begin{tikzcd} \mathbf{D} \arrow[r, shift right, "G"'] \arrow[r, phantom, "\perp"] & \mathbf{C} \arrow[l, shift right, "F"'] \end{tikzcd}$$

is called a **Quillen adjunction** if the left adjoint preserves cofibrations and acyclic cofibrations. The model category axioms imply that this is equivalent to requiring that the right adjoint preserves fibrations and acyclic fibrations. The adjoint functors are called (left and right) **Quillen functors**.

If $F \dashv G$ is a Quillen adjunction, such that, for all cofibrant objects x and fibrant objects y, the morphism $Fx \to y$ is a weak equivalence if and only if the adjunct $x \to Gy$ is a weak equivalence, then $F \dashv G$ is called a **Quillen equivalence**.

Quillen equivalences can also be characterized at the homotopical level.

Property 13.3.8. A Quillen adjunction $F \dashv G$ is a Quillen equivalence if and only if the left derived functor LF (or right derived functor RG) is an equivalence.

Property 13.3.9 (Derived adjunction). If $F \dashv G$ is a Quillen adjunction, the derived functors (LF, RG) also form an adjunction.

Property 13.3.10 (Doubly categorical interpretation). The map that sends a model category to its homotopy category and a Quillen functor to its derived functor is a *double pseudofunctor*. Amongst other things, this implies that the composition of derived functors is naturally weakly equivalent to the derived functor of the composition.

Example 13.3.11 (Topological spaces). A first example of model categories is the category **Top** of topological spaces (Chapter 6 and Chapter 7). This category can be endowed with a model structure by taking the weak equivalences to be the weak homotopy equivalences (Definition 7.1.32) and by taking the fibrations to be the Serre fibrations (Definition 7.1.61).

Example 13.3.12 (Simplicial sets). As a second example, consider the category **sSet**. This category can be turned into a model category by taking the weak equivalences to be the morphisms that induce weak homotopy equivalences between geometric re-

alizations and by taking the fibrations to be Kan fibrations. With this structure, the fibrant objects are the Kan complexes and the cofibrations are the monomorphisms, i.e. the levelwise injections. This model structure is generally called **Quillen's model structure** and the resulting model category is denoted by $\mathbf{sSet}_{\text{Quillen}}$.

Property 13.3.13 (Quillen). The adjoint pair of geometric realization and singular set functors (Example 13.1.11 and Example 13.1.13) gives a Quillen equivalence between the above model categories. This result allows to regard simplicial sets as if they were spaces and vice versa. Consequently, most of homotopy theory can be done in either category.

13.3.2 Monoidal structures

Definition 13.3.14 (Pushout product). Let $(\mathbf{M}, \otimes)$ be a closed symmetric monoidal category and consider two morphisms $f : a \to b$ and $g : x \to y$ in $\mathbf{M}$. By taking suitable tensor products, one can form the span $a \otimes y \leftarrow a \otimes x \to b \otimes x$. If the pushout of this diagram exists, the pushout product $f \square g$ is defined as the unique morphism from this pushout to $b \otimes y$ defined by the obvious diagram.

The dual concept, where one of the arguments is contravariant, is sometimes called a **pullback hom**, **pullback exponential** or **pullback power** depending on which bifunctor is used in the definition. In fact, the requirement that $\mathbf{M}$ carries a monoidal structure can be dropped and any bifunctor $\circledast : \mathbf{C} \times \mathbf{D} \to \mathbf{E}$ can be used. This more general construction is sometimes called the **Leibniz construction** and the case where the bifunctor is a tensor bifunctor is then called the **Leibniz tensor**.

If $\mathbf{M}$ is also a model category, it is said to satisfy the **pushout-product axiom** if the pushout product of two cofibrations is again a cofibration, acyclic if any of the input cofibrations is.

Definition 13.3.15 (Quillen bifunctor). A bifunctor satisfying the pushout-product axiom that preserves colimits in both variables is called a **(left) Quillen bifunctor**. It should be noted that the tensor product automatically satisfies this last property in the case of closed monoidal categories.

In fact, the natural setting for defining Quillen bifunctors is that of two-variable adjunctions (Definition 4.2.14). Consider such a triple of bifunctors $(\otimes, \mathsf{Hom}_L, \mathsf{Hom}_R)$.

- The bifunctor $- \otimes - : \mathbf{C} \times \mathbf{D} \to \mathbf{E}$ is said to be **left Quillen** if it satisfies the pushout-product axiom.
- The bifunctors $\mathsf{Hom}_L : \mathbf{C}^{op} \times \mathbf{E} \to \mathbf{D}$ and $\mathsf{Hom}_R : \mathbf{D}^{op} \times \mathbf{E} \to \mathbf{C}$ are said to be **right Quillen** if the Leibniz product of a cofibration and a fibration is a fibration and if the acyclicity of any of the inputs implies the acyclicity of the result.

It can be shown that, if one of these bifunctors is Quillen, the others are too.

Remark 13.3.16. The fact that for two-variable adjunctions one does not mention preservation of (co)limits follows from the property that left (resp. right) adjoints preserve colimits (resp. limits).

Definition 13.3.17 (Monoidal model category). A model category $\mathbf{M}$ that carries the structure of a closed symmetric monoidal category $(\mathbf{M}, \otimes, \mathbf{1})$ such that the following compatibility conditions are satisfied:

1. **Pushout-product**: The tensor bifunctor and internal homs define a Quillen two-variable adjunction.

2. **Unit**: For every cofibrant object $x \in \mathsf{ob}(\mathbf{M})$ and every cofibrant replacement $\mathbf{1}^c$ of $\mathbf{1}$, the induced morphism $\mathbf{1}^c \otimes x \to x$ is a weak equivalence.

Definition 13.3.18 (Enriched model category). Let $\mathcal{V}$ be a monoidal model category. A model category $\mathbf{M}$ is said to be $\mathcal{V}$-enriched if is satisfies the following conditions:

1. $\mathbf{M}$ is a $\mathcal{V}$-enriched category that is both powered and copowered over $\mathcal{V}$ (Definition 4.2.11).

2. The copower is a left Quillen bifunctor or, equivalently, the power is a right Quillen bifunctor.

Example 13.3.19 (Simplicial model category). An $\mathbf{sSet}_{\text{Quillen}}$-enriched model category. It can be shown that the full subcategory of a simplicial model category on fibrant-cofibrant objects is enriched over Kan-complexes.

13.3.3 Homotopy category

Definition 13.3.20 (Homotopy). Before being able to construct homotopies in general model categories, thereby generalizing the ideas from Section 7.1, one has to define the counterpart of the unit interval in a model category $\mathbf{M}$. To this end, consider an object $x \in \mathsf{ob}(\mathbf{M})$. By definition, one has the unique diagonal and codiagonal morphisms $\Delta : x \to x \times x$ and $\nabla : x \sqcup x \to x$. If these morphisms can be factorized in $\mathbf{M}$ through weak equivalences

$$x \to \mathsf{Path}(x) \qquad \text{and} \qquad \mathsf{Cyl}(x) \to x\,,$$

a **path** and **cylinder object** are obtained. By choosing the morphism to be a fibration (resp. cofibration), the notion of **good** cylinder (resp. path) object is obtained.

Using these objects, one can define left and right homotopies between parallel morphisms $f, g : x \to y$. A left homotopy between f and g is a morphism $\eta : \mathsf{Cyl}(x) \to y$ such that Diagram 13.1a commutes. Analogously, a right homotopy between f and g is a morphism $\lambda : x \to \mathsf{Path}(y)$ such that Diagram 13.1b commutes.

$$\begin{tikzcd} x \arrow[r] \arrow[dr, "f"'] & \mathsf{Cyl}(x) \arrow[d, "\eta"] & x \arrow[l] \arrow[dl, "g"] \\ & y & \end{tikzcd}$$

(a) A left homotopy.

$$\begin{tikzcd} & x \arrow[dl, "f"'] \arrow[d, "\lambda"] \arrow[dr, "g"] & \\ y & \mathrm{Path}(y) \arrow[l] \arrow[r] & y \end{tikzcd}$$

(b) A right homotopy.

Figure 13.1: Homotopies in a model category.

Example 13.3.21 (Topological spaces). Consider the category **Top**. A cylinder object for a topological space X is given by the product $X \times [0,1]$. The codiagonal map is factorized as the endpoint inclusion $X \sqcup X \rightrightarrows X \times [0,1]$ followed by the collapse $X \times [0,1] \xrightarrow{\pi_X} X$, where it is not hard to show that the collapse is a homotopy equivalence. A left homotopy with respect to $X \times [0,1]$ is exactly a homotopy in the sense of Definition 7.1.2.

A path object $\mathsf{Path}(X)$ is given by the mapping space $X^{[0,1]}$. The diagonal map is factorized by the basepoint inclusion $X \xrightarrow{\iota_0} X^{[0,1]}$ followed by the endpoint projections $X^{[0,1]} \rightrightarrows X \times X$. That the right homotopies with respect to $X^{[0,1]}$ give the same equivalence classes as the left homotopies is the content of Property 7.1.9.

Property 13.3.22. If the domain is cofibrant, every left homotopy induces a right homotopy. Dually, if the codomain is fibrant, every right homotopy induces a left homotopy.

Corollary 13.3.23 (Equivalence relation). Whenever the domain $x \in \mathsf{ob}(\mathbf{M})$ is cofibrant and the codomain $y \in \mathsf{ob}(\mathbf{M})$ is fibrant, the relations of being left-homotopic (or, equivalently, right-homotopic) coincide on $\mathbf{M}(x,y)$ and, in particular, define equivalence relations. The equivalence classes are denoted by $[x,y]$.

Property 13.3.24 (Stability under composition). Homotopies are preserved under both precomposition and postcomposition by arbitrary morphisms.

Property 13.3.25 (Weak equivalences). A morphism is a weak equivalence if and only if it is a homotopy inverse.

Definition 13.3.26 (Homotopy equivalence). Two objects $x, y \in \mathsf{ob}(\mathbf{M})$ in a model category are said to be homotopy equivalent if there exists morphisms $f : x \leftrightarrows y : g$ such that $f \circ g$ and $g \circ f$ are homotopic to the identity. The morphisms f, g are then said to be homotopy equivalences.

In Definition 13.2.5, it was shown that one can assign to every category with weak equivalences a homotopy category. When the category has the additional structure of a model category, one can construct an equivalent category.

Alternative Definition 13.3.27 (Homotopy category II). The homotopy category $\mathbf{Ho}(\mathbf{M})$ of a model category $\mathbf{M}$ is the category defined by the following data:

- **Objects**: $\mathrm{ob}(\mathbf{M})$, and
- **Morphisms**: $\mathbf{Ho}(\mathbf{M})(x, y) := [x^{\mathrm{cf}}, y^{\mathrm{cf}}]$.

In fact, it is easier to restrict to the subcategory $\mathbf{M}_{\mathbf{cf}}$ of $\mathbf{Ho}(\mathbf{M})$ on the fibrant-cofibrant objects due to the following property (when restricting to a subcategory, the resulting homotopy category is only equivalent and not isomorphic to the localization).

Property 13.3.28. The homotopy category of a model category is equivalent to those of the full subcategories on (co)fibrant objects:

$$\mathbf{Ho}(\mathbf{M}) \cong \mathbf{Ho}(\mathbf{M}_{\mathbf{f}}) \cong \mathbf{Ho}(\mathbf{M}_{\mathbf{c}}) \cong \mathbf{Ho}(\mathbf{M}_{\mathbf{cf}}). \tag{13.17}$$

The following theorem gives a stronger version of Property 13.3.25 above.

Theorem 13.3.29 (Whitehead). *A weak equivalence between objects that are both fibrant and cofibrant is a homotopy equivalence.*

Property 13.3.30. A Quillen equivalence of model categories induces an equivalence of homotopy categories.

Property 13.3.31 (Monoidal model categories). The homotopy category of a monoidal model category has a closed monoidal structure defined by the induced derived adjunction. The homotopy category of an enriched model category is enriched, powered and copowered over the homotopy category of its enriching category, where the enriched structure is again given by the induced derived adjunction.

In some cases, it is useful to consider categories that are strictly weaker than model categories but stronger than categories with weak equivalences. The prime example being the full subcategories of a model category on the (co)fibrant objects. These are often easier to handle in the setting of homotopy theory.

Definition 13.3.32 (Category of fibrant objects). A category $\mathbf{C}$ with weak equivalences $\mathbf{W} \hookrightarrow \mathbf{C}$ equipped with another subcategory $\mathbf{F} \hookrightarrow \mathbf{C}$ for which the morphisms are called **fibrations** such that the following conditions are satisfied:

1. $\mathbf{C}$ has finite products.
2. Fibrations and acyclic fibrations are preserved under arbitrary pullbacks.
3. Every object admits a good path object.
4. All objects are fibrant.

Theorem 13.3.33 (Factorization lemma). *In a category of fibrant objects, any morphism can be factorized as the right inverse of an acyclic fibration followed by a fibration.*

The following theorem is important for characterizing functors that preserve weak equivalences.

Theorem 13.3.34 (Ken Brown's lemma). *Let* $\mathbf{C}$ *be a category of fibrant objects and let* $\mathbf{D}$ *be a category with weak equivalences. If a functor* $F : \mathbf{C} \to \mathbf{D}$ *maps acyclic fibrations to weak equivalences, it preserves all weak equivalences.*

Remark 13.3.35. An analogous theorem exists for *categories of cofibrant objects*.

The above lemma allows to define derived functors for Quillen functors between model categories (the following definition is also better suited for working in the ∞-setting than Definition 13.2.9).

Alternative Definition 13.3.36 (Derived functor). Let $F : \mathbf{C} \to \mathbf{D}$ be a left (resp. right) Quillen functor. The left (resp. right) derived functors are obtained by precomposition with the cofibrant (resp. fibrant) replacement functors.

Property 13.3.37 (Derived functors are absolute). It can be shown that derived functors built using (co)fibrant replacement are given by absolute Kan extensions. So, even though homotopy categories often do not admit all (co)limits, the resulting Kan extensions do exist.

13.3.4 Reedy model structure ♣

Consider a bicomplete category $\mathbf{M}$ (this will often be $\mathbf{sSet}$). For any full subcategory inclusion $\mathbf{D} \hookrightarrow \mathbf{C}$, one obtains an induced **truncation** (or **restriction**) functor $\mathrm{tr} : \mathbf{M}^{\mathbf{C}} \to \mathbf{M}^{\mathbf{D}}$. The left and right adjoints of this functor are, respectively, called the **skeleton** and **coskeleton** functors.

Formula 13.3.38. The adjoint functors are defined by Kan extensions and, hence, can be expressed in terms of (co)ends and weighted (co)limits:

$$\begin{aligned} \mathrm{sk}(X)x &:= \int^{y \in \mathbf{D}} \mathbf{C}(y,x) \cdot Xy = \mathrm{colim}^{\mathbf{C}(-,x)} X\,, \\ \mathrm{cosk}(X)x &:= \int_{y \in \mathbf{D}} \big[\mathbf{C}(x,y), Xy\big] = \lim^{\mathbf{C}(x,-)} X. \end{aligned} \tag{13.18}$$

Definition 13.3.39 (Skeletal sets). Let $\mathbf{M} = \mathbf{sSet}$ and consider the inclusion $\mathbf{\Delta}_{\leq n} \hookrightarrow \mathbf{\Delta}$ of the full subcategory on the objects $\{[0], \dots, [n]\}$. The n-truncation functor tr_n discards all simplices of degree higher than n or, in other words, it 'truncates' a simplicial set at degree n.

The n-skeleton functor sk_n takes a simplicial set X of degree $\leq n$ and freely adds degenerate simplices in degrees $> n$, i.e. it is the smallest simplicial set containing X as a simplicial subset. The n-coskeleton functor cosk_n adds a simplex in degree $> n$ whenever all of its faces are present, i.e. the m-simplices in $\mathrm{cosk}_n X$ are given by the collection of all $(m+1)$-tuples of $(m-1)$-simplices that are compatible (along lower simplices).

Property 13.3.40 (Simplicial nerve). The nerve functor $N : \mathbf{Cat} \to \mathbf{sSet}$ from Example 13.1.10 is a fully faithful functor to the category of 2-coskeletal simplicial sets. This follows from the fact that, in ordinary categories, compositions of morphisms are unique and, hence, all higher-order tuples of composable morphisms are determined by composable pairs. In fact, this is just the characterization of (small) categories as categories internal to **Set** (under the isomorphism $C_2 \cong C_1 \times_{C_0} C_1$ which will be called the first *Segal condition* in Definition 13.4.19).

Definition 13.3.41 (Reedy category). A category $\mathbf{R}$ equipped with a **degree** function $\mathrm{ob}(\mathbf{R}) \to \alpha$, where α is an ordinal (Definition 1.7.21), and two wide subcategories $\mathbf{R}^{\pm}$ that satisfy the following conditions:

1. Nontrivial morphisms in $\mathbf{R}^+$ (strictly) increase the degree.
2. Nontrivial morphisms in $\mathbf{R}^-$ (strictly) decrease the degree.
3. All morphisms admit a unique factorization as a morphism in $\mathbf{R}^-$ followed by a morphism in $\mathbf{R}^+$. This factorization is sometimes called the **(canonical) Reedy factorization**.

Property 13.3.42 (Minimality). The Reedy factorization is the (unique) factorization of minimal degree, where the **degree** of a factorization $x \xrightarrow{f} y \xrightarrow{g} z$ is defined as the degree of y.

Property 13.3.43 (Isomorphisms are trivial). A morphism in a Reedy category is an isomorphism if and only if it is trivial.

Example 13.3.44. Some common examples of Reedy categories are discrete categories, finite posets, the simplex category $\mathbf{\Delta}$ and opposites of Reedy categories.

For Reedy categories, one can also define n-truncation, n-skeleton and n-coskeleton functors by restricting to the full subcategories on elements of degree $\leq n$.

Definition 13.3.45 (Matching and latching objects). Let $\mathbf{R}$ be a Reedy category and consider a diagram $X \in [\mathbf{R}, \mathbf{C}]$ with $\mathbf{C}$ small. Consider the skeleton monad and coskeleton comonad (often just called the skeleton and coskeleton functors) $\mathbf{sk}_n := \mathrm{sk}_n \circ \mathrm{tr}_n$ and $\mathbf{cosk}_n := \mathrm{cosk}_n \circ \mathrm{tr}_n$. The latching and matching objects of X at $r \in \mathrm{ob}(\mathbf{R})$ are defined as the restrictions of $\mathbf{sk}_{\deg(r)-1}$ and $\mathbf{cosk}_{\deg(r)-1}$ to the degree-$\deg(r)$ subcategory of $\mathbf{R}$, respectively. The counits of these adjunctions give rise to the **latching** and **matching** maps.

One can also define the latching and matching objects through (co)limits. Define the subcategory $\mathbf{R}^+(r)$ as the subcategory of $\mathbf{R}^+/r$ on all objects except the identity. The latching object $L_r X$ can be shown to be isomorphic to the colimit of X over $\mathbf{R}^+(r)$.

Example 13.3.46 (Simplicial objects). The above property allows to give a nice inter-

pretation to latching objects in the case of $\mathbf{R} = \mathbf{\Delta}^{\mathrm{op}}$. Using the *Eilenberg–Zilber lemma*, it can be shown that the nth latching object of a simplicial object is given by its collection of degenerate n-simplices.

Definition 13.3.47 (Boundary). The boundary of a representable presheaf $\mathbf{R}(-, r)$, denoted by $\partial\mathbf{R}(-, r)$, is defined as its latching object. It can be shown that $\partial\mathbf{R}(-, r)$ consists of exactly those morphisms that are not in $\mathbf{R}^{-}$.

Formula 13.3.48. One can show that the latching and matching objects can be obtained through (co)limits weighted by boundaries:

$$\begin{aligned} M_r X &\cong \lim^{\partial\mathbf{R}(r,-)} X\,, \\ L_r X &\cong \operatorname{colim}^{\partial\mathbf{R}(-,r)} X\,. \end{aligned} \tag{13.19}$$

From here on, $\mathbf{M}$ will be assumed to be a model category. In this case, a canonical model structure on the functor category $[\mathbf{R}, \mathbf{M}]$ for Reedy R can be defined.

Definition 13.3.49 (Relative matching and latching objects). Consider the (weighted) colimit bifunctor

$$\operatorname{colim} : \mathbf{Psh}(\mathbf{R}) \times [\mathbf{R}, \mathbf{M}] \to \mathbf{M}\,. \tag{13.20}$$

The Leibniz construction 13.3.14 allows to define the relative latching object of $f : X \to Y$ at $r \in \mathsf{ob}(\mathbf{R})$ as the Leibniz product of the boundary inclusion $\partial\mathbf{R}(-, r) \hookrightarrow \mathbf{R}(-, r)$ and f.

By Eqs. (4.30) and (13.19), the relative latching map is of the form $Xr \sqcup_{L_r X} L_r Y \to Yr$ and the relative matching map is of the form $Xr \to Yr \times_{M_r Y} M_r X$.

Property 13.3.50 (Reedy model structure). Let $\mathbf{R}$ be a Reedy category and let $\mathbf{M}$ be a model category. The functor category $[\mathbf{R}, \mathbf{M}]$ admits the following model structure:

- **Weak equivalences**: the objectwise weak equivalences.
- **Fibrations**: morphisms for which the relative matching map is a fibration (in $\mathbf{M}$) for all $r \in \mathsf{ob}(\mathbf{R})$
- **Cofibrations**: morphisms for which the relative latching map is a cofibration (in $\mathbf{M}$) for all $r \in \mathsf{ob}(\mathbf{R})$

Property 13.3.51 (Quillen (co)limit functors). Consider a $\mathcal{V}$-enriched model category $\mathbf{M}$ and a Reedy category $\mathbf{R}$. For every Reedy cofibrant functor W in $[\mathbf{R}, \mathcal{V}]$, the weighted limit and colimit functors are right and left Quillen, respectively.

@@ COMPLETE (this section is too abstract and difficult at this point, a more intuitive explanation of matching/latching objects is necessary) @@

13.4 Simplicial spaces

13.4.1 Kan complexes

Definition 13.4.1 (Horn). Consider the standard simplex $\Delta[n]$. For all $n \geq 1$ and $0 \leq k \leq n$, the (n,k)-horn $\Lambda^k[n]$ is defined as the subsimplicial set obtained by removing the kth face from $\partial\Delta[n]$. When $k = 0$ or $k = n$, the horn is said to be **outer**, otherwise it is said to be **inner**.

Definition 13.4.2 (Kan fibration). A morphism of simplicial sets that has the right lifting property with respect to all horn inclusions $\Lambda^k[n] \hookrightarrow \Delta[n]$.

Definition 13.4.3 (Kan complex). A simplicial set that has all horn fillers or, equivalently, a simplicial set for which the terminal morphism is a Kan fibration. The full subcategory of **sSet** on Kan complexes is denoted by **Kan**.

Property 13.4.4 (Horn filler condition). A simplicial set is the nerve of a (small) category if and only if all of its inner horns admit a unique filler (compositions are unique). If one requires all horns to admit a unique filler, the nerve of a groupoid is obtained.

By relaxing the above requirements, one can generalize the notion of a category (this is due to *Boardman* and *Vogt*).

Definition 13.4.5 (Quasicategory[2]). A simplicial set for which all inner horns have (not necessarily unique) fillers. This condition is sometimes called the **Boardman condition**.

At this point, a third instance of a 'homotopy category' can be defined.

Definition 13.4.6 (Homotopy category III). Let X be a quasicategory. The homotopy category $\mathbf{Ho}(X)$ consists of the following data:

- **Objects**: X_0.
- **Morphisms**: The quotient of X_1 under the relation $f \circ g \sim h$ if there exists a 2-simplex with (suitably oriented) edges f, g and h.

Property 13.4.7 (Fundamental category). If X is a quasicategory, its homotopy category is equivalent to its **fundamental category** $\pi_1(X)$, i.e. the image under the left adjoint of the (simplicial) nerve functor.

The following theorem is a restatement of Property 13.4.4.

Theorem 13.4.8 (Joyal). *A quasicategory is a Kan complex if and only if its homotopy category is a groupoid.*

[2]Some authors, such as *Joyal*, call these **logoi** (singular: **logos**).

13.4.2 Simplicial localization

Definition 13.4.9 (Homotopy category IV). Consider a simplicially enriched category $\mathbf{C}$. Its homotopy category $\pi_0(\mathbf{C})$ is defined as follows:

- **Objects**: $\mathsf{ob}(\mathbf{C})$, and
- **Morphisms**: $\pi_0\mathbf{C}(x, y)$.

Two morphism $f, g \in \mathbf{C}(x, y)$ are said to be **homotopic** if they are identified in $\pi_0(\mathbf{C})$. A morphism $f \in \mathbf{C}(x, y)$ is called a **homotopy equivalence** if it admits both a left and a right inverse in $\pi_0(\mathbf{C})$, i.e. if it becomes an isomorphism in the homotopy category.

Definition 13.4.10 (Dwyer–Kan equivalence I). Consider a simplicial functor $F : \mathbf{C} \to \mathbf{D}$ between two simplicially enriched categories. This functor is called a Dwyer–Kan equivalence if:

- F is **∞-fully faithful**, i.e. the induced map on hom-objects is a weak equivalence.[3]
- The induced map on connected components $\pi_0(F) : \pi_0(\mathbf{C}) \to \pi_0(\mathbf{D})$ is an equivalence (of categories). In fact, this condition can be relaxed to $\pi_0(F)$ being essentially surjective, since together with the previous condition this implies that $\pi_0(F)$ is an equivalence.

Construction 13.4.11 (Hammock localization). The hammock localization (or **simplicial localization**) $L^H\mathbf{C}$ of a category with weak equivalences $(\mathbf{C}, W)$ is the simplicially enriched category constructed as follows:

- **Objects**: $\mathsf{ob}(\mathbf{C})$, and
- **Morphisms**: $L^H\mathbf{C}(x, y)$ is the simplicial set defined as follows:

  First, for every $n \in \mathbb{N}$, one constructs a category with as objects the zigzags of morphisms in $\mathbf{C}$ relating x and y such that all left-pointing morphisms are in W, and as morphisms the endpoint-preserving 'natural transformations' (in the sense that all triangles/squares in the resulting diagram commute). Then, the coproduct is taken over the (simplicial) nerves of the categories for all $n \in \mathbb{N}$. Finally, the quotient is taken by the equivalence relations generated by:

  1. inserting or removing identity morphisms, and
  2. composing morphisms.

[3] This is a generalization of the ordinary definition for categories. (Simplicial categories will give models for ∞-*categories*.)

The following property relates all constructions of what could be called homotopy categories.

Property 13.4.12. Let $(\mathbf{C}, W)$ be a category with weak equivalences. It can be shown that

$$\mathbf{Ho}(L^H\mathbf{C}) \cong \mathbf{C}[W^{-1}]\,. \tag{13.21}$$

This construction gives a more explicit description of the homotopy category. Furthermore, if $\mathbf{M}$ is a simplicial model category, the categories $\mathbf{M}_{\mathbf{cf}}$ and $L^H\mathbf{M}$ are Dwyer–Kan equivalent.

Property 13.4.13. Quillen equivalent model categories have Dwyer–Kan equivalent simplicial localizations.

Property 13.4.14. Up to Dwyer–Kan equivalence, every simplicially enriched category can be obtained as the simplicial localization of a category with weak equivalences.

Definition 13.4.15 (Simplicial resolution). Consider an object $x \in \mathsf{ob}(\mathbf{M})$ in a model category $\mathbf{M}$. A (co)simplicial resolution of x is a Reedy (co)fibrant (co)simplicial object X together with a weak equivalence $x \simeq X_0$.

Every object in a model category admits a (co)simplicial resolution by taking a (co)fibrant replacement of the constant (co)simplicial object.

Definition 13.4.16 (Bergner model structure). The catgory of simplicially enriched categories admits the following model structure:

- **Weak equivalences**: Dwyer–Kan equivalences,
- **Cofibrant objects**: *simplicial computads*, and
- **Fibrant objects**: **Kan**-enriched categories.

Construction 13.4.17 (Free resolution). Consider a (small) category $\mathbf{C}$. From this category, we can construct a simplicial category $\mathfrak{C}\mathbf{C}$ consisting of the following data:

- **Objects**: $\mathsf{ob}(\mathbf{C})$, and
- **Morphisms**: $\mathsf{hom}(\mathfrak{C}\mathbf{C})_n := \mathsf{hom}(F^{n+1}\mathbf{C})$, where F is identity-on-objects and $F\mathbf{C}$ has as morphisms strings of composable morphisms in $\mathbf{C}$.

It can be shown that $\mathfrak{C}\mathbf{C}$ is a *simplicial computad* for any simplicially enriched category $\mathbf{C}$, $\mathfrak{C}$ acts as the cofibrant replacement functor in the Bergner model structure.

13.4.3 Segal spaces

Example 13.4.18 (Bisimplicial sets). Simplicial objects in **sSet** are, in particular, also **sSet**-enriched and, accordingly, such objects are often called bisimplicial sets or **sim-**

plicial spaces[4]. Since **sSet** is a model category, Example 13.3.5 above says that **ssSet** itself admits a model structure. It can furthermore be shown that the injective model structure on **ssSet** coincides with the Reedy model structure (a rather nontrivial statement).

Definition 13.4.19 (Segal space). Consider a fibrant object X in the injective model structure on **ssSet**. This bisimplicial set is called a Segal space if it satisfies the following weak form of the **Segal condition** for all $n \geq 1$:[5]

$$X_n \simeq \underbrace{X_1 \times_{X_0} \cdots \times_{X_0} X_1}_{n \text{ times}} . \tag{13.22}$$

These maps are called **Segal maps** (even when they are not weak equivalences). They are the morphisms induced by **spine** inclusions, i.e. inclusions of the union of edge 1-cells.

Definition 13.4.20 (Segal category). A bisimplicial set X is called a **Segal precategory** if X_0 is discrete. It is called a Segal category if, in addition, all its Segal maps are weak equivalences, i.e. if it is a Segal space with a discrete set of 'objects'.

Definition 13.4.21 (Mapping space). Consider a Segal space X. For every two points $x, y \in X_0$, the mapping space $\mathsf{Map}(x, y)$ is defined as the fibre of $(d_1, d_0) : X_1 \rightrightarrows X_0 \times X_0$ over the point (x, y). The identity at $x \in X_0$ is given by $s_0(x)$.

The following two definitions should be compared to Definitions 13.4.9 and 13.4.10.

Definition 13.4.22 (Homotopy category V). Consider a Segal category X. Its homotopy category $\mathbf{Ho}(X)$ is defined as follows:

- **Objects**: X_0, and
- **Morphisms**: $\pi_0 \mathsf{Map}(x, y)$.

Two points $f, g \in \mathsf{Map}(x, y)$ are said to be **homotopic** if they are identified in $\mathbf{Ho}(X)$. A point $f \in \mathsf{Map}(x, y)$ is called a **homotopy equivalence** if it admits both a left and a right inverse in $\mathbf{Ho}(X)$, i.e. if it becomes an isomorphism. The subspace $X_{\mathsf{hoequiv}} \subset X_1$ consists of the components that contain homotopy equivalences.[6]

Definition 13.4.23 (Dwyer–Kan equivalence II). A map of Segal spaces such that:

1. the induced map on mapping spaces is a weak equivalence, and
2. the induced map between homotopy categories is an equivalence (of categories).

[4]The latter name follows from the fact that topological spaces and simplicial sets are (Quillen-)equivalent.

[5]If the Reedy condition is omitted, the limit on the right-hand side has to be replaced by a *homotopy limit*.

[6]It should be noted that if any point in a component is a homotopy equivalence, all points in that component are homotopy equivalences.

Definition 13.4.24 (Complete Segal space). A Segal space X for which the map $s_0 : X_0 \to X_{\text{hoequiv}}$ is a weak equivalence.

Property 13.4.25 (Dwyer-Kan equivalence). A map of Segal spaces is a Dwyer–Kan equivalence if and only if it is a weak equivalence in the *complete Segal space model structure*. A map of complete Segal spaces is a Dwyer–Kan equivalence if and only if it is a levelwise weak equivalence.

13.4.4 Coherent nerve

The nerve and realization functors (Construction 13.1.9) can also be modified to incorporate the higher homotopical data present in a simplicially enriched category.

First, define a cosimplicial simplicially enriched category $C : \mathbf{\Delta} \to \mathbf{sSetCat}$ that assigns to every finite ordinal $[n]$, the category consisting of the following data:

- **Objects**: $[n] \equiv \{0, 1, \dots, n\}$, and
- **Morphisms**: $C[n](i, j) := N(P_{ij})$, where N is the ordinary nerve functor on **Cat** and P_{ij} is the poset consisting of all subsets of $\{i, \dots, j\}$ that contain both i and j.

The (homotopy) **coherent nerve functor** (or **simplicial nerve functor**[7]) is the nerve functor induced by this cosimplicial object. It not only knowns about all possible morphisms, it also knows about all possible ways how one can obtain this morphism.

Remark 13.4.26. The cosimplicial object $C : \mathbf{\Delta} \to \mathbf{sSetCat}$ is in fact just the free resolution functor $\mathfrak{C}$ in the Bergner model structure (Construction 13.4.17) restricted to the subcategory $\mathbf{\Delta}$.

Property 13.4.27. The homotopy coherent nerve functor N_Δ is uniquely determined by the following relation:

$$\mathbf{sSet}(\Delta[n], N_\Delta \mathbf{C}) = \mathbf{sSetCat}(C[n], \mathbf{C}) . \tag{13.23}$$

Property 13.4.28 (Simplicial realization). The left adjoint of the homotopy coherent nerve functor satisfies the following relation for all finite ordinals $[n]$:

$$|\Delta[n]| \cong C[n] . \tag{13.24}$$

Property 13.4.29 (Quasicategories). If $\mathbf{C}$ is **Kan**-enriched, its coherent nerve is a quasicategory. By Example 13.3.19, this allows to associate a quasicategory to any simplicial model category by passing to the coherent nerve of its fibrant-cofibrant subcategory.

At this point, it is interesting to reconsider the simplicial nerve construction (Example 13.1.10). Although the functor $N : \mathbf{Cat} \to \mathbf{sSet}$ is fully faithful, there is a problem

[7] This terminology was also used for the ordinary nerve functor taking values in **sSet**. In general, it should be clear from the context which one is meant.

in that weak equivalences of simplicial sets do not necessarily correspond to (weak) equivalences of categories. The issue comes from the definition of the canonical model structure (Example 13.3.12) on **sSet**. Here, the weak equivalences are those morphisms that become weak equivalences after geometric realization. During this step, information about the direction of morphisms is lost and it becomes impossible to distinguish categories from groupoids. Only when all morphisms are assumed to be invertible, i.e. the category is assumed to be a groupoid, does one recover the converse statement.

A first solution is to use a different model structure with less weak equivalences.

Definition 13.4.30 (Joyal model structure). **sSet** admits a model structure defined by the following data:

- **Weak equivalences**: *weak categorical equivalences*, i.e. maps $f : X \to Y$ such that the induced simplicial functor $C(f) : C(X) \to C(Y)$ is a Dwyer–Kan equivalence.
- **Cofibrations**: monomorphisms.
- **Fibrant objects**: quasicategories (Definition 13.4.5).

When characterizing those simplicial sets that are nerves of categories, groupoids and categories are distinguished by whether all horns admit a (unique) filler or only inner horns admit a (unique) filler. However, when the fibrant objects ar<e Kan complexes, those simplicial sets that admit all horn fillers, the homotopical structure only knows about nerves that come from a groupoid. By passing to the larger class of quasicategories, this problem is solved.

Instead of changing the model structure on **sSet** to overcome the issues with taking nerves of categories or groupoids, one can also change the construction of the nerve functor. An alternative approach was introduced by *Rezk*.

Definition 13.4.31 (Classifying diagram). Consider a (small) category $\mathbf{C}$ together with the functor category $\mathbf{C}^{[n]}$. The classifying diagram of $\mathbf{C}$ is the bisimplicial set $\widetilde{N}\mathbf{C}$ defined levelwise as follows:

$$\widetilde{N}\mathbf{C}_k := N\big(\mathsf{Core}(\mathbf{C}^{[k]})\big)\,, \tag{13.25}$$

where N is the ordinary nerve functor and Core denotes the core of a category (Definition 3.5.13).

The reason for why this construction is better for distinguishing categories stems from the fact that information about isomorphisms is already captured at degree 0, while information about noninvertible morphisms is only captured from degree 1 onwards.

Property 13.4.32. If $\mathbf{C}$ is small, then $\widetilde{N}\mathbf{C}$ is a complete Segal space.

@@ COMPLETE @@

13.5 Cofibrantly generated categories

13.5.1 Transfinite constructions

Before proceeding, some notions from ordinary category need to be specialized to the context of regular cardinals (Definition 1.7.32). In this section, the symbol κ will always denote such a regular cardinal.

Definition 13.5.1 (κ-filtered category). A category in which every diagram with less than κ arrows admits a cocone.

Definition 13.5.2 (κ-directed limit). Consider a poset I such that every subposet of cardinality less than κ has a lower bound (upper bound for directed colimits). Such a set is said to be κ-(co)directed. A limit of a diagram over this poset is called a κ-(co)directed (co)limit.

The following definition is a categorification of the previous one.

Definition 13.5.3 (κ-filtered limit). Consider a diagram $D : \mathbf{I} \to \mathbf{C}$. The limit (resp. colimit) of D is said to be κ-cofiltered (resp. κ-filtered) if $\mathbf{I}$ is a κ-cofiltered (resp. κ-filtered) category.

It should be noted that an analogue of Property 3.4.53 also holds in the κ-context, i.e. a category has all κ-directed limits if and only if it has all κ-filtered limits (and analogously for colimits).

Definition 13.5.4 (Small object). An object for which there exists a regular cardinal κ such that its covariant hom-functor preserves all κ-filtered colimits. These objects are also said to be **κ-compact** or **κ-presentable** (cf. Definition 3.4.58).

Definition 13.5.5 (Accessible category). A locally small category $\mathbf{C}$ for which there exists a regular cardinal κ such that $\mathbf{C}$ has all κ-filtered colimits and such that it is generated under κ-filtered colimits by a small subcategory $\mathbf{D}$ of κ-small objects, i.e. $\mathbf{C} \cong \mathsf{Ind}_\kappa(\mathbf{D})$.

Alternative Definition 13.5.6. A category equivalent to the category of (κ-)flat functors (Definition 12.4.48) on a small category or, by Property 12.4.51, the category of (κ)ind-objects on a small category.

Definition 13.5.7 (Locally presentable category). A cocomplete, accessible category. It can be shown that such categories are also complete and, hence, bicomplete.

Property 13.5.8. Every locally presentable category can be obtained as a full reflective subcategory of a presheaf category under an **accessible embedding**, i.e. a finitary embedding of accessible categories.

Property 13.5.9 (Gabriel–Ulmer duality). First, define the following two 2-categories. $\mathbf{LFP}$ consists of the following data:

- **Objects**: locally finitely presentable categories,
- **1-Morphisms**: finitary, right adjoint functors (Definition 3.4.54), and
- **2-Morphisms**: natural transformations.

Lex consists of the following data:

- **Objects**: small, finitely complete categories,
- **1-Morphisms**: left exact functors, and
- **2-Morphisms**: natural transformations.

There exists an equivalence $\mathbf{Lex}^{\mathrm{op}} \cong \mathbf{LPF}$ that sends a small, finitely complete category to its category of left exact copresheaves.

Theorem 13.5.10 (Adjoint functor theorem). *Consider a functor $F : \mathbf{C} \to \mathbf{D}$ between locally presentable categories. F admits a right adjoint if and only if it preserves all colimits. F admits a left adjoint if and only if it is accessible and if it preserves all limits.*

For ordinary categories, the axioms guarantee the existence of a (unique) composition of any finite number of (composable) morphisms. However, in some cases it is useful, or even necessary, to talk about the 'composite' of an infinite number of morphisms.

Definition 13.5.11 (Transfinite composition). Consider a category $\mathbf{C}$ with a collection of morphisms $I \subseteq \mathsf{hom}(\mathbf{C})$ and let α be an infinite ordinal. An α-indexed **transfinite sequence** of morphisms in I is a diagram of the form $D : \alpha \to \mathbf{C}$ such that:

1. Successor morphisms in α are mapped to elements of I.
2. D is 'continuous' in the sense that for every limit ordinal $\beta < \alpha$: $D\beta \cong \operatorname*{colim}_{\gamma<\beta} D\gamma$.

$D\lambda$ denotes the restriction of D to the (full) subdiagram $\{\gamma \mid \gamma < \lambda\}$ of α. The transfinite composition of this sequence is the induced morphism $D_0 \to D\alpha \equiv \mathsf{colim}\, D$.

Definition 13.5.12 (Cell complex). Consider a cocomplete category $\mathbf{C}$ with a collection of morphisms $I \subseteq \mathsf{hom}(\mathbf{C})$. A **relative** I-cell complex is a transfinite composition of pushouts (of coproducts[8]) of morphisms in I. An I-cell complex is an object such that the unique morphism from the initial object is a relative I-cell complex.

Notation 13.5.13 (Relative cell complexes). The set of all relative I-cell complexes is often denoted by $\mathsf{cell}(I)$.

[8] It can be shown that closure under coproducts follows automatically.

13.5.2 Cofibrant generation

Definition 13.5.14 (Saturated class). A collection of morphisms $I \subseteq \mathsf{hom}(\mathbf{C})$ in a category is said to be **(weakly) saturated** if it is closed under:

1. pushouts,
2. retracts (in the arrow category of $\mathbf{C}$), and
3. transfinite compositions.

Definition 13.5.15 (Cofibrantly generated WFS). A weak factorization system (R, L) on a category $\mathbf{C}$ (Definition 13.2.2) is said to be cofibrantly generated by a collection of morphisms $I \subseteq \mathsf{hom}(\mathbf{C})$ if

$$R = \mathsf{rlp}(I)\,. \tag{13.26}$$

The following is a famous result by *Quillen*.

Theorem 13.5.16 (Small object argument). *Let $\mathbf{C}$ be a locally presentable category with a set of morphisms $I \subseteq \mathsf{hom}(\mathbf{C})$. The WFS given by $\mathsf{llp}(I) = \mathsf{cell}(I)$ is cofibrantly generated.*

Remark 13.5.17. This theorem can be generalized to cocomplete categories where the domains of morphisms in I are small relative[9] to $\mathsf{cell}(I)$. Sets of morphisms with this property are said to **admit a small object argument**.

Definition 13.5.18 (Cofibrantly generated model category). A model category $\mathbf{C}$ is said to be cofibrantly generated by two sets of morphisms $I, J \subseteq \mathsf{hom}(\mathbf{C})$ if it satisfies the following conditions:

1. I and J both admit the small object argument.
2. The fibrations are given by $\mathsf{rlp}(J)$.
3. The acyclic fibrations are given by $\mathsf{rlp}(I)$.

It can be shown that the last two conditions are equivalent to the following ones:

2*. The cofibrations are the retracts (in the arrow category) of $\mathsf{cell}(I)$.

3*. The acyclic cofibrations are the retracts (in the arrow category) of $\mathsf{cell}(J)$.

The morphisms in I and J are called the **generating cofibrations** and **generating acyclic cofibrations**, respectively.

Sometimes, it is desirable to replace the model structure on a category $\mathbf{C}$ by one that has more weak equivalences (denote the new one by $\mathbf{C}_0$). If the cofibrations remain

[9] *Small relative to a set of morphisms* is defined just as ordinary smallness, but with general κ-filtered colimits replaced by those that start from morphisms in the given set.

the same, this new structure has some nice properties:

- The fibrations are a subclass of the original ones.
- The acyclic fibrations remain the same.
- The identity functors $\mathbb{1}_{\mathbf{C}_0} : \mathbf{C}_0 \leftrightarrows \mathbf{C} : \mathbb{1}_{\mathbf{C}}$ form a Quillen adjunction.
- Every object in $\mathbf{C}$ is weakly equivalent (in $\mathbf{C}_0$) to one in $\mathbf{C}_0$.

This procedure can be made explicit for a specific class of model categories. Let $\mathbf{C}$ be a left proper, cofibrantly generated, simplicial model category and consider a collection $S \subset \mathsf{hom}(\mathbf{C})$ of cofibrations with cofibrant domain. First, the notion of 'S-local objects' is introduced.

Definition 13.5.19 (Local object). A fibrant object $x \in \mathsf{ob}(\mathbf{C})$ is said to be S-local if, for all morphisms in S, the image under the Yoneda embedding of x is an acyclic Kan fibration. Analogously, a morphism is called an S-local weak equivalence if, for all S-local fibrant objects, its image under their Yoneda embeddings is an acyclic Kan fibration.

Property 13.5.20. Every weak equivalence is an S-local weak equivalence: $W \subset W_S$.

Construction 13.5.21 (Left Bousfield localization). Given a model category $\mathbf{C}$ with the same assumptions as before, the (left) Bousfield localization $L_S\mathbf{C}$ is defined as the same category but with the following model structure:

- **Cofibrations**: $\mathsf{cof}(\mathbf{C})$, and
- **Acyclic cofibrations**: cofibrations that are S-local weak equivalences.

If it exists, the Bousfield localization $L_S\mathbf{C}$ of $\mathbf{C}$ at S is the universal left Quillen functor $\gamma : \mathbf{C} \to L_S\mathbf{C}$ such that its left derived functor sends the image of S to isomorphisms in $\mathbf{Ho}(\mathbf{C})$. The identity functor gives a Quillen adjunction between $\mathbf{C}$ and $L_S\mathbf{C}$ and the induced adjunction on homotopy categories defines a reflective localization.

Remark 13.5.22. The notions of local object/morphisms can be slightly generalized (they give equivalent localizations). An **S-local morphism** $f : a \to b$ is a morphism such that for all objects $x \in \mathsf{ob}(\mathbf{M})$ for which the map

$$\mathbf{C}(g^{\mathrm{c}}, x^{\mathrm{f}}) : \mathbf{C}(d^{\mathrm{c}}, x^{\mathrm{f}}) \to \mathbf{C}(c^{\mathrm{c}}, x^{\mathrm{f}}) \tag{13.27}$$

is a weak equivalence for all $g \in S$ (x is said to be **S-local**), the map

$$\mathbf{C}(f^{\mathrm{c}}, x^{\mathrm{f}}) : \mathbf{C}(b^{\mathrm{c}}, x^{\mathrm{f}}) \to \mathbf{C}(a^{\mathrm{c}}, x^{\mathrm{f}}) \tag{13.28}$$

is also a weak equivalence. This is the homotopical version of Definition 3.2.21.

Definition 13.5.23 (Combinatorial model category). A locally presentable, cofibrantly generated model category.

The following theorem is an analogue of Property 13.4.14 for combinatorial model categories.

Theorem 13.5.24 (Dugger). *Every combinatorial model category is Quillen equivalent to a (left) Bousfield localization of a category of simplicial presheaves on a small category (with the global projective model structure 13.3.5).*

@@ FINISH @@

13.6 Homotopy (co)limits

Consider a category $\mathbf{C}$ with weak equivalences, together with diagrams $D, D' : \mathbf{I} \to \mathbf{C}$. Assume that there exists a weak equivalence between D and D', i.e. a natural transformation that consists of componentwise weak equivalences. Clearly, this induces a morphism between (co)limits, but it would be nice if the construction of (co)limits would also preserve the homotopical structure, i.e. weakly equivalent diagrams should have weakly equivalent (co)limits.

The main purpose of this section is to introduce a modification of the ordinary (co)limit functors that takes into account the homotopical structure of the underlying categories. A first step is the modification of (co)products.

Definition 13.6.1 (Homotopy (co)products). In ordinary categories, the universal property of a product (and, dually, for a coproduct) characterizes it as an object with an isomorphism

$$\mathbf{C}(-, c) \cong \prod_{i \in I} \mathbf{C}(-, c_i) \tag{13.29}$$

in that every collection of morphisms $\{f_i : x \to c_i\}_{i \in I}$ can be factorized as follows

$$\begin{tikzcd} x \arrow[d, dashed, "\exists!"'] \arrow[dr, "f_i"] & \\ c \arrow[r, "\pi_i"'] & c_i \,. \end{tikzcd}$$

To obtain a homotopical version, the commutativity of this diagram is relaxed up to a homotopy/path in $\mathbf{C}(x, c_i)$. This leads to the definition of a homotopy product as an object $c \in \mathrm{ob}(\mathbf{C})$ together with a natural weak (homotopy) equivalence

$$\psi_x : \mathbf{C}(x, c) \simeq \prod_{i \in I} \mathbf{C}(x, c_i) \,. \tag{13.30}$$

The homotopy (co)products are unusual in the sense that they can be obtained as (co)limits in the homotopy category $\mathbf{Ho}(\mathbf{C})$. More generally, one can define homotopy (co)limits in any homotopical category by passing to derived functors.

Definition 13.6.2 (Homotopy (co)limits). Let $\mathbf{A}$ be a homotopical category and let $\mathbf{B}$ be a small category. The homotopy limit and colimit functors are defined as the derived functors of the limit and colimit functors $(\mathsf{co})\mathsf{lim} : [\mathbf{B}, \mathbf{A}] \to \mathbf{A}$.

Remark 13.6.3. The reason for why homotopy (co)products can be obtained as ordinary (co)limits is related to the fact that their indexing categories are discrete. In this case, the adjoint of the derived functor is equivalent to the diagonal functor on the homotopy category.

Property 13.6.4 (Model categories). Consider the specific case where $\mathbf{C}$ is a model category. If $[\mathbf{D}, \mathbf{C}]$ admits an injective (resp. projective) model structure, the homotopy limit (resp. colimit) always exists and can be obtained through fibrant (resp. cofibrant) replacement as in Definition 13.3.36.

Example 13.6.5 (Reedy categories). Consider the general case of diagrams $D : \mathbf{R} \to \mathbf{C}$ with $\mathbf{R}$ Reedy. First, note that the constant functor $\Delta : \mathbf{C} \to [\mathbf{R}, \mathbf{C}]$ maps weak equivalences to (pointwise) weak equivalences.[10] If the Reedy structure is such that the constant functor preserves cofibrations, this functor is left-Quillen and Ken Brown's lemma 13.3.34 implies that its right Quillen adjoint, the limit functor, preserves weak equivalences. In this case, one can define the **homotopy limit** $\mathsf{holim}\, D$ as the functor $\mathsf{lim}(D \circ Q_{\mathrm{f}})$, where Q_{f} is the fibrant replacement-functor (which, in this case, acts pointwise). A dual construction gives rise to **homotopy colimits**.

13.6.1 Simplicially enriched diagrams

In the setting where diagrams are enriched over $\mathbf{sSet}$, one can define homotopy (co)limits in a more sophisticated way.

Definition 13.6.6 (Homotopy colimit). Consider a diagram $D : \mathbf{I} \to \mathbf{C}$ with $\mathbf{C}$ copowered over $\mathbf{sSet}$. The homotopy colimit of D is defined as the following tensor product (Definition 4.2.20):

$$\mathsf{hocolim}\, D := N(-/\mathbf{I}) \otimes_{\mathbf{I}} D = \int^{i \in \mathbf{I}} N(i/\mathbf{I}) \cdot Di\,, \tag{13.31}$$

where N is the nerve functor (Example 13.1.10).

A similar definition for homotopy limits can be used when the category is powered over $\mathbf{sSet}$.

Definition 13.6.7 (Homotopy limit). Consider a diagram $D : \mathbf{I} \to \mathbf{C}$ with $\mathbf{C}$ powered over $\mathbf{sSet}$. The homotopy limit of D is defined as the following hom-like object:

$$\mathsf{holim}\, D := \int_{i \in \mathbf{I}} [N(\mathbf{I}/i), Di]\,. \tag{13.32}$$

[10] This is also true when $\mathbf{R}$ is not Reedy.

This is exactly the characterization of the homotopy limit as the **sSet**-natural transformations between $N(\mathbf{I}/-)$ and D.

Remark 13.6.8 (Bousfield–Kan map). The expressions from the above formulas are also known as the **Bousfield–Kan formulas**. It should be noted that the above definitions are not strictly equivalent to the ones from the previous section. To be precise, the Bousfield–Kan formulas are only weakly equivalent to the general definitions if the objects in $\mathbf{C}$ are replaced by their (co)fibrant replacements, i.e. if in the above expressions D is postcomposed by a (co)fibrant replacement functor.

@@ COMPLETE (CHECK THESE STATEMENTS) @@

By Definition 13.3.27, one can construct morphisms in a homotopy category as morphisms from a cofibrant replacement to a fibrant replacement. This allows to define diagrams-up-to-homotopy (in two settings).

Definition 13.6.9 (Homotopy coherent diagram). A morphism in the homotopy category of the Bergner model category. When considering diagrams taking values in a **Kan**-enriched category, one can use the free resolution functor to characterize them as functors $D : \mathfrak{C}\mathbf{I} \to \mathbf{C}$. Since the simplicial realization functor extends the free resolution functor $\mathfrak{C}$ to simplicial sets, one can also define homotopy coherent diagrams on simplicial sets.

Natural transformations between such diagrams are given by homotopy coherent diagrams $\mathfrak{C}(\mathbf{I} \times \Delta[1]) \to \mathbf{C}$ in analogy with ordinary homotopies. However, these do no compose uniquely in the sense that one does not obtain a well-defined diagram on $\mathbf{I} \times \Delta[2]$. Therefore, one does not obtain a category of homotopy coherent diagrams. For simplicial sets I, it can be shown that the natural structure is that of a quasicategory:

$$\mathbf{CohDgrm}(\mathbf{I}, \mathbf{C}) \cong (N\mathbf{C})^{\mathbf{I}}, \tag{13.33}$$

where N is the simplicial nerve functor.

@@ COMPLETE @@

Chapter 14

Higher topos theory ♣

Chapter Contents

14.1 Higher category theory **310**
14.1.1 n-categories 310
14.1.2 n-functors 313
14.1.3 Higher (co)limits 314
14.2 ∞-Categories **315**
14.2.1 Simplicial approach 315
14.2.2 Quasicategories 315
14.2.3 Categorical notions 316
14.3 Stacks **316**
14.3.1 2-sheaves 316
14.3.2 Stacks on a 1-site 318
14.4 Higher topos theory **320**
14.5 Cohomology **321**

14.1 Higher category theory

This section gives an introduction to the theory of higher categories, in particular n-categories for finite $n > 0$. For the notion of ∞-categories, see Section 14.2.

14.1.1 n-categories

Definition 14.1.1 (n-category). A (strict) n-category consists of:

- objects (0-morphisms),
- 1-morphisms going between 0-morphisms,
- ...
- n-morphisms going between $(n-1)$-morphisms,

such that the composition of k-morphisms ($k \leq n$) is associative and satisfies the unit laws as required in an ordinary category. By generalizing this definition to arbitrary $n \in \mathbb{N}$, one can define the notion of a (strict) ∞-category.

If one relaxes the associativity and unit laws up to higher coherent morphisms, one obtains the notion of a **weak n-category**. Explicit definitions for such categories have been constructed up to **tetracategories** ($n = 4$). However, this construction by *Trimble* takes about 50 pages of diagrams.

Remark. n-morphisms are also called **n-cells**. This makes their relation to topological spaces (and, in particular, simplicial spaces) more apparent.

Example 14.1.2. The classical examples of a 1-category and 2-category are **Set** and **Cat**, respectively.

Property 14.1.3 (Composition in 2-categories). 2-morphisms can be composed in two different ways:

- **Horizontal composition**: Consider two 2-morphisms $\alpha : f \Rightarrow g$ and $\beta : f' \Rightarrow g'$, where $f' \circ f$ and $g' \circ g$ are well-defined. These 2-morphisms can be composed as

$$\beta \circ \alpha : f' \circ f \Rightarrow g' \circ g\,. \tag{14.1}$$

  This is sometimes called the **Godement product**.

- **Vertical composition**: Consider two 2-morphisms $\alpha : f \Rightarrow g$ and $\beta : g \Rightarrow h$, where f, g and h have the same domain and codomain. These 2-morphisms can be composed as

$$\beta \cdot \alpha : f \Rightarrow h\,. \tag{14.2}$$

As a consistency condition, the horizontal and vertical composition are required to satisfy the following **interchange law**:

$$(\alpha \cdot \beta) \circ (\gamma \cdot \delta) = (\alpha \circ \gamma) \cdot (\beta \circ \delta). \tag{14.3}$$

Definition 14.1.4 ((n,r)-category). A higher (∞-)category for which

- all parallel k-morphisms with $k > n$ are equivalent and, hence, trivial.

- all k-morphisms with $k > r$ are invertible (or equivalences in the fully weak ∞-sense).

Definition 14.1.5 (Weak inverse). Let $\mathbf{C}$ be a 2-category. A 1-morphism $f : x \to y$ is weakly invertible if there exist a 1-morphism $g : y \to x$ and 2-isomorphisms $g \circ f \Rightarrow \mathbb{1}_x$ and $f \circ g \Rightarrow \mathbb{1}_y$.

At this point, it should be obvious that the definition of a unit-counit adjunction (Definition 3.2.27) can be generalized to general 2-categories:

Definition 14.1.6 (Adjunction in 2-category). Let $\mathbf{C}$ be a 2-category. An adjunction in $\mathbf{C}$ is a pair of 1-morphisms $F : x \to y$ and $G : y \to x$ together with 2-morphisms $\varepsilon : F \circ G \Rightarrow \mathbb{1}_y$ and $\eta : \mathbb{1}_x \Rightarrow G \circ F$ that satisfy the zig-zag identities.

By looking at the defining relations of duals in a rigid monoidal category (Section 4.4), it should be clear that these are in fact the same as the defining relations of the unit and counit of an adjunction. This is a consequence of the fact that a 2-category with a single object can be regarded as a (strict) monoidal category where the composition in the 2-category becomes the tensor product in the monoidal category. Similarly, adjoint 1-morphisms in the 2-category become duals in the monoidal category. This is formalized as follows.

Property 14.1.7 (Monoidal categories). Consider a monoidal category $(\mathbf{C}, \otimes, \mathbf{1})$. From this monoidal category, one can construct the so-called **delooping** (bicategory) $\mathbf{BC}$ in the following way:

- There is a single object $*$.
- The 1-morphisms in $\mathbf{BC}$ are the objects in $\mathbf{C}$.
- The 2-morphisms in $\mathbf{BC}$ are the morphisms in $\mathbf{C}$.
- Horizontal composition in $\mathbf{BC}$ is the tensor product in $\mathbf{C}$.
- Vertical composition in $\mathbf{BC}$ is composition in $\mathbf{C}$.

Conversely, every 2-category with a single object comes from a monoidal category. Hence, the 2-category of (pointed) 2-categories with a single object and the 2-category of monoidal categories are equivalent. (This property and its generalizations are the content of the *delooping hypothesis*.)

In the same way one can deloop a braided monoidal category twice and find an identification with a one-object tricategory with one 1-morphism. However, this identification is not a trivial one as it makes use of the Eckmann–Hilton argument (Property 3.5.1) to identify different monoidal structures on this tricategory. (See also Section 4.10.)

14.1.2 n-functors

Definition 14.1.8 (2-functor). A 2-functor $F : \mathbf{C} \to \mathbf{D}$ (often called a **pseudofunctor**) is a morphism between bicategories. It consists of the following data:

- a function $F_0 : \mathsf{ob}(\mathbf{C}) \to \mathsf{ob}(\mathbf{D})$, and
- for every two objects $x, y \in \mathsf{ob}(\mathbf{C})$, a functor $F_{x,y} : \mathbf{C}(x,y) \to \mathbf{D}(Fx, Fy)$.

The function F_0 and the functors $F_{x,y}$ are also often denoted by F by abuse of notation. This data is required to satisfy some coherence conditions. These are specified by the following data:

1. **Associator**: For every pair of composable 1-morphisms $f \circ g$ in $\mathsf{hom}(\mathbf{C})$, a 2-isomorphism $\gamma_{f,g} : Ff \circ Fg \Rightarrow F(f \circ g)$ such that for every triple of composable morphisms $f \circ g \circ h$ in $\mathsf{hom}(\mathbf{C})$, the following identity holds:
$$\gamma_{f\circ g,h} \circ (\gamma_{f,g} \cdot \mathbb{1}_{Fh}) = \gamma_{f,g\circ h} \circ (\mathbb{1}_{Ff} \cdot \gamma_{g,h}). \tag{14.4}$$
2. **Unitor**: For every object $x \in \mathsf{ob}(\mathbf{C})$, a 2-isomorphism $\iota_x : \mathbb{1}_{Fx} \Rightarrow F\mathbb{1}_x$ such that for every morphism $f : x \to y$ in $\mathsf{hom}(\mathbf{C})$, the following identities hold:
$$\begin{aligned} \iota_y \cdot \mathbb{1}_{Ff} &= \gamma_{\mathbb{1}_y, f} \\ \mathbb{1}_{Ff} \cdot \iota_x &= \gamma_{f, \mathbb{1}_x}. \end{aligned} \tag{14.5}$$

Note that to be completely formal, one should have inserted the unitors and associators of the bicategories $\mathbf{C}, \mathbf{D}$.

Definition 14.1.9 (Lax natural transformation). Consider two 2-functors $F, G : \mathbf{C} \to \mathbf{D}$. A lax natural transformation $\eta : F \Rightarrow G$ consists of the following data:

1. For every object $x \in \mathsf{ob}(\mathbf{C})$, a 1-morphism $\eta_x : Fx \to Gx$.
2. For every 1-morphism $f : x \to y$ in $\mathsf{hom}(\mathbf{C})$, a 2-morphism $\eta_f : Gf \circ \eta_x \Rightarrow \eta_y \circ Ff$ such that the η_f are the components of a natural transformation $(\eta_x)^* \circ G \Rightarrow (\eta_y)_* \circ F$ and such that the assignment $f \mapsto \eta_f$ satisfies the 'obvious' identity and composition axioms.

Remark 14.1.10. As usual in the context of higher category theory, one can speak of lax 2-functors if the associator and unitors are merely required to be 2-morphisms, and of strict 2-functors if these morphisms are required to be identities. If the natural transformations between morphism categories in the definition of a lax natural transformation are all isomorphisms, this is called a **pseudonatural transformation**. If the 1-morphisms η are equivalences, they are called lax natural equivalences.

Definition 14.1.11 (Modification). Consider two 2-functors $F, G : \mathbf{C} \to \mathbf{D}$ and two parallel (lax) natural transformations $\alpha, \beta : F \Rightarrow G$. A modification $\mathfrak{m} : \alpha \Rrightarrow \beta$ maps

every object $x \in \mathrm{ob}(\mathbf{C})$ to a 2-morphism $\mathfrak{m}_x : \alpha_x \Rightarrow \beta_x$ such that $\beta_f \circ (\mathbb{1}_{Gf} \cdot \mathfrak{m}_x) = (\mathfrak{m}_y \cdot \mathbb{1}_{Ff}) \circ \alpha_f$.

This is generalized as follows.

Definition 14.1.12 (Transfor). A k-transfor[1] between two n-categories maps j-morphisms to $(j+k)$-morphisms (in a coherent way).

Example 14.1.13. The definitions for operations in bicategories above lead us to the following 'explicit' expressions for k-transfors (for small k):

- $k = 0$: n-functors,
- $k = 1$: (n-)natural transformations,
- $k = 2$: modifications, and
- $k = 3$: **perturbations**.

The following definition generalizes the notion of essential surjectivity (Definition 3.2.12) to higher category theory.

Definition 14.1.14 (n-surjective functor). An ∞-functor $F : \mathbf{C} \to \mathbf{D}$ is said to be n-surjective if for any two parallel $(n-1)$-morphisms f, g in $\mathbf{C}$ and n-morphism $\alpha : Ff \to Fg$ in $\mathbf{D}$, there exists an n-morphism $\tilde{\alpha}$ in $\mathbf{C}$ such that $F\tilde{\alpha} \cong \alpha$.

Definition 14.1.15 (Indexed category). Consider a category $\mathbf{I}$. An $\mathbf{I}$-indexed category is a pseudofunctor $\mathbf{C} : \mathbf{I}^{\mathrm{op}} \to \mathbf{Cat}$, i.e. a 2-presheaf on $\mathbf{I}$. Indexed functors and natural transformations are defined analogously.

Definition 3.2.32 can be generalized as follows.

Definition 14.1.16 (Universal morphism). Consider an n-functor $F : \mathbf{C} \to \mathbf{D}$ and an object $x \in \mathrm{ob}(\mathbf{D})$. A universal morphism from x to F is a pair $(d, f : x \to Fd)$ such that:

1. Every morphism $x \to Fd'$ factors through f, and
2. The precomposition $f^* : \mathbf{D}(Fd, Fd') \to \mathbf{D}(x, Fd')$ is fully faithful.

14.1.3 Higher (co)limits

Definition 14.1.17 (Weighted 2-limit). Consider 2-categories $\mathbf{I}, \mathbf{C}$ together with 2-functors $W : \mathbf{I} \to \mathbf{Cat}$ and $F : \mathbf{I} \to \mathbf{C}$. By direct generalization of the ordinary definition of weighted limits (Section 4.2.2), one says that $\lim^W F$ is the W-weighted (2-)limit of F if

[1] This name was first introduced by Crans (1998). A different name that is sometimes used is (n,k)-**transformation**, but this should not be confused with the natural transformations in the context of (n,r)-categories.

there exists a pseudonatural equivalence

$$\mathbf{C}(x, \lim^{W} F) \cong [\mathbf{I}, \mathbf{Cat}]\big(W, \mathbf{C}(x, F-)\big). \tag{14.6}$$

By restricting to the 2-category of strict 2-categories, strict 2-functors and strict natural transformations the resulting notion of a weighted 2-limit coincides with that of an ordinary weighted limit enriched in **Cat** (since strict 2-categories are simply **Cat**-enriched 1-categories.)

@@ COMPLETE @@

14.2 ∞-Categories

14.2.1 Simplicial approach

The first approach to ∞-category theory is the simplicial one. The motivation is Property 13.4.4, which relates the categorical structure to the existence of certain horn fillers. The generalization is then given by the notion of quasicategories (Definition 13.4.5).

Theorem 14.2.1 (Lurie). *An ∞-category is presentable if and only if it is equivalent to the coherent nerve of the fibrant-cofibrant subcategory of a combinatorial model category and, hence by Dugger's theorem 13.5.24, can be presented by simplicial presheaves.*

14.2.2 Quasicategories

Definition 14.2.2 (Overcategory). Consider a simplicial morphism $F : X \to \mathbf{C}$ where $\mathbf{C}$ is a quasicategory. The overcategory $\mathbf{C}_{/F}$, generalizing the comma category $\Delta \downarrow F$, is characterized by a natural isomorphism

$$\mathsf{Hom}_{\mathbf{sSet}}(Y, \mathbf{C}_{/F}) \cong \mathsf{Hom}_{X_{/\mathbf{sSet}}}(\iota_Y, F), \tag{14.7}$$

where $\iota_Y : X \hookrightarrow Y \star X$ is the join inclusion.

Remark 14.2.3 (1-categorical interpretation). Consider the case where X is the simplicial nerve of the one-object groupoid $\{*\}$. The functor F then singles out an object $c \in \mathsf{ob}(\mathbf{C})$. Since the image of F is 0-truncated, morphisms on the right-hand side are also 0-truncated.

@@ COMPLETE @@

Definition 14.2.4 (Terminal object). An object $1 \in \mathsf{ob}(\mathbf{C})$ such that

1. The projection $\mathbf{C}_{/1} \to \mathbf{C}$ is an acyclic fibration, and
2. for every object $d \in \mathsf{ob}(\mathbf{C})$ the right hom-Kan complex $\mathsf{Hom}^R(d, 1) := \mathbf{C}_{/1} \times_{\mathbf{C}} \{d\}$ is contractible.

Definition 3.4.37 can be generalized as follows.

Definition 14.2.5 (Limit). Consider an $(\infty,1)$-functor $F : \mathbf{C} \to \mathbf{D}$ of quasicategories. The limit of F is defined as the terminal object in the overcategory $\mathbf{C}_{/F}$.

14.2.3 Categorical notions

Definition 14.2.6 (Truncated object). An ∞-groupoid is said to be n-truncated for $n \in \mathbb{N}$ if it is an n-groupoid. An object $x \in \mathrm{ob}(\mathbf{C})$ of an $(\infty,1)$-category is said to be n-truncated if for all objects $y \in \mathrm{ob}(\mathbf{C})$, the hom-groupoid $\mathbf{C}(y,x)$ is n-truncated.

Definition 14.2.7 (Monomorphism). A morphism $f : x \to y$ in an $(\infty,1)$-category $\mathbf{C}$ that is (-1)-truncated as an object of the slice category $\mathbf{C}_{/y}$. Equivalently, the projection $\mathbf{C}_{/f} \to \mathbf{C}_{/y}$ is fully faithful.

Definition 14.2.8 (Epimorphism). A morphism $f : x \to y$ in an $(\infty,1)$-category $\mathbf{C}$ such that the induced morphism $\mathbf{C}(f,z) : \mathbf{C}(y,z) \to (x,z)$ is monic for all objects $z \in \mathrm{ob}(\mathbf{C})$.

14.3 Stacks

14.3.1 2-sheaves

An important subject, especially in the context of gauge theories in physics, is that of groupoid-valued (pre)sheaves. To this end, sites are generalized to higher category theory.

Definition 14.3.1 (2-presheaf). Consider a 2-category $\mathbf{C}$. A 2-presheaf on $\mathbf{C}$ is a pseudofunctor $F : \mathbf{C}^{\mathrm{op}} \to \mathbf{Cat}$. When $\mathbf{C}$ is the categorification of a 1-category, i.e. when it has discrete hom-categories, 2-presheaves are better known as **prestacks**.

Definition 14.3.2 (2-coverage). This is virtually the same as an ordinary coverage (Definition 12.4.8), but factorization is only required to exist up to an isomorphism. A 2-category equipped with a 2-coverage is called a **2-site**.

As for 1-sites, every coverage generates a unique sieve. It is the full subcategory on those morphisms that factor through a covering map in the given coverage (again up to isomorphism).

As in the case of ordinary categories (Definition 12.4.12), one can define 2-sheaves through a descent condition.

Definition 14.3.3 (2-sheaf). A 2-presheaf $F : \mathbf{C}^{\mathrm{op}} \to \mathbf{Cat}$ on a 2-site $(\mathbf{C}, J)$ is said to be a 2-sheaf with respect to J if for all sieves $S \in J$ the following functor is an equivalence:

$$Fc \cong \mathbf{Psh}_2(\mathcal{Y}c, F) \to \mathbf{Psh}_2(S, F)\,, \tag{14.8}$$

where the fist equivalence is just the 2-Yoneda lemma.

Remark 14.3.4. It should be noted that 2-(pre)sheaves can also be defined on ordinary (1-)sites. Sieves, regarded as subfunctors of the Yoneda embedding, take values in **Set**. By composing these with the embedding **Set** $\hookrightarrow$ **Cat** of sets as (discrete) categories, one obtains 2-subfunctors of the 2-Yoneda embedding. Often 2-sheaves over 1-sites are called **stacks** (although this terminology is also used for general 2-sites).

Definition 14.3.5 (Prestack of groupoids). Consider a category **C**. A prestack of groupoids on **C** is a **Grpd**-valued prestack on **C**.

The category of (groupoid-valued) prestacks becomes **Grpd**-enriched if one takes the hom-category between two prestacks F, G to consist of the following data:

- **Objects**: Natural transformations $\alpha : F \Rightarrow G$ (note that the components are themselves functors).
- **Morphisms**: 'strict modifications' in the sense that they map objects in **C** to natural transformations satisfying the whiskering condition (see also Definition 14.1.11)

$$\mathbb{1}_{Ff} \cdot \mathfrak{m}_b = \mathfrak{m}_a \cdot \mathbb{1}_{Gf} \,. \tag{14.9}$$

For ordinary sites and presheaves, descent was defined in terms of matching families. Since presheaves are now taking values in a 2-category, the matching families are a bit more complex. However, this structure is already familiar from differential geometry and algebraic topology, where it is known under the name of the *Čech nerve*.

Definition 14.3.6 (Čech groupoid). Consider a site $(\mathbf{C}, J)$. To every covering family $\mathcal{U} = \{f_i : x_i \to x\}_{i \in I}$, one can assign an internal groupoid in presheaves $C(\mathcal{U})$ consisting of the following data:

- **Objects**: $\bigsqcup_i \mathcal{Y}x_i$, and
- **Morphisms**: $\bigsqcup_{i,j} \mathcal{Y}x_i \times_{\mathcal{Y}x} \mathcal{Y}x_j$.

This is equivalent to the (**Grpd**-valued) presheaf that assigns to every object $y \in \mathsf{ob}(\mathbf{C})$ the groupoid consisting of the following data:

- **Objects**: Pairs $(i, g_i : y \to x_i)$ where $x_i \in \mathcal{U}$, and
- **Morphisms**: A unique arrow $(i, g_i) \to (j, g_j)$ if and only if $f_i \circ g_i = f_j \circ g_j$.

Comparing the definition of morphisms in the Čech groupoid to the condition for matching families in Definition 12.4.9, shows that one could presume that the Čech groupoid is related to the matching families. This intuition is indeed correct as explained by the following property.

Property 14.3.7 (Matching families). Any ordinary presheaf F can be considered to be **Grpd**-valued by postcomposing with the embedding **Set** $\hookrightarrow$ **Grpd**. For any covering

family $\mathcal{U}$, there exists an isomorphism

$$[\mathbf{C}^{\mathrm{op}}, \mathbf{Grpd}]\big(C(\mathcal{U}), F\big) \cong \mathbf{Psh}_2(\mathcal{U}, F)\,. \tag{14.10}$$

Because the Čech groupoid (co)represents a descent object, it is sometimes called a **codescent object**.

It is exactly this (co)descent property of the Čech groupoid that will be used in Chapter 44 to define (higher) smooth groupoids. Readers with some experience in algebraic topology will also notice that the Čech groupoid only contains the first degrees of the Čech complex. The full Čech complex can be obtained from the following construction.

Definition 14.3.8 (Čech nerve). Consider a morphism $f : y \to x$ in a category $\mathbf{C}$. The Čech nerve $C_\bullet(f)$ is the simplicial object (Definition 13.1.3) that contains, in degree $k \in \mathbb{N}$, the $(k+1)$-fold pullback of f along itself. For a covering family $\mathcal{U} \equiv \{f_i : x_i \to x\}$, the Čech nerve is defined as

$$C_\bullet(\mathcal{U}) := C_\bullet\left(\coprod_i x_i \to x\right). \tag{14.11}$$

For ∞-sheaves, the full Čech nerve will be used. However, for 2-sheaves and, in particular, stacks, only its 3-coskeleton is necessary. The extra information will encode the *cocycle condition* (33.1), well-known from the study of *fibre bundles*.

14.3.2 Stacks on a 1-site

For the definition of stacks, one needs the notions of fibred categories or, equivalently, pseudofunctors as defined in Section 3.3.1. The definitions are recalled here.

> Consider a functor $\Pi : \mathbf{A} \to \mathbf{B}$. A morphism f in $\mathbf{A}$ is said to be Π-Cartesian if, for every morphism φ in $\mathbf{A}$ and factorization of $\Pi\varphi$ through Πf in $\mathbf{B}$, there exists a unique factorization of φ through f. f is called the inverse image of Πf.
>
> A fibred category consists of a functor $\Pi : \mathbf{A} \to \mathbf{B}$ such that for each morphism $f : c \to d$ in $\mathbf{B}$ with $d \in \mathrm{im}(\Pi)$ and each lift $y \in \mathbf{A}_d$ there exists at least one inverse image in $(\tilde{f} : x \to y) \in \mathbf{A}$ of f. By the Grothendieck construction every fibred category gives rise to a pseudofunctor $F : \mathbf{B}^{\mathrm{op}} \to \mathbf{Cat}$ by sending objects to their fibres under Π and sending morphisms f to their pullback functors f^*.

Definition 14.3.9 (Descent datum). Consider a category $\mathbf{C}$ with a covering family $\mathcal{U} \equiv \{f_i : x_i \to x\}$ and a pseudofunctor $F : \mathbf{C}^{\mathrm{op}} \to \mathbf{Cat}$. The projections associated to the pullback $x_i \cap x_j := x_i \times_x x_j$ will be denoted by π_i and π_j, respectively (and analogously for iterated pullbacks). A descent datum for $\mathcal{U}$ with respect to F is a pair of families

$(\{g_i\}, \{f_{ij}\})_{i,j\in I}$, where $\{g_i\}$ is a matching family for $\mathcal{U}$ with respect to F and every f_{ij} is an isomorphism $\pi_i^* x_i \cong \pi_j^* x_j$. This data is required to satisify the following **cocycle condition**:

$$\pi_{ik}^* f_{ik} = \pi_{ij}^* f_{ij} \circ \pi_{jk}^* f_{jk} \,. \tag{14.12}$$

Morphisms $(\{g_i\}, \{f_{ij}\}) \to (\{g'_i\}, \{f'_{ij}\})$ between descent data are families of morphisms $\{\phi_i : g_i \to g'_i\}$ that satisfy

$$\pi_i^* \phi_i \circ f_{ij} = f'_{ij} \circ \pi_j^* \phi_j \,. \tag{14.13}$$

The category of descent data for $\mathcal{U}$ with respect to F will be denoted by $\mathsf{Descent}(\mathcal{U}, F)$.

Construction 14.3.10. Consider an object ξ in Fx. From this object, one can construct a descent datum as follows. The morphisms g_i are the pullbacks $f_i^*\xi$ and the isomorphisms $f_{ij} : \pi_i^* f_i^* \xi \cong \pi_j^* f_j^* \xi$ are obtained from the fact that both these objects are (Cartesian) pullbacks of the same morphisms. Arrows in Fx induce morphisms of descent data by (Cartesian) pullbacks along the covering maps. This construction defines a functor $Fx \to \mathsf{Descent}(\mathcal{U}, F)$. It can be shown that this construction is independent of a choice of cleavage up to equivalence.

Definition 14.3.11 (Stack). Consider a fibred category F over a site $(\mathbf{C}, J)$.

- F is called a **separated prestack** if for each covering family $\mathcal{U}$ on $x \in \mathsf{ob}(\mathbf{C})$, the functor $Fx \to \mathsf{Descent}(\mathcal{U}, F)$ is fully faithful.
- F is called a **stack** if for each covering family $\mathcal{U}$ on $x \in \mathsf{ob}(\mathbf{C})$ the functor $Fx \to \mathsf{Descent}(\mathcal{U}, F)$ is an equivalence.

This is a generalization of the descent condition in Definition 12.4.12. This can be seen by observing that $\mathsf{Descent}(\mathcal{U}, F) \cong \mathbf{Psh}_2(S(\mathcal{U}), F)$, where $S(\mathcal{U})$ is the sieve generated by $\mathcal{U}$ regarded as a fibred category. When F is fibred over groupoids, it is called a **stack of groupoids**. This forms the category $\mathbf{Sh}_{(2,1)}(\mathbf{C})$ of $(2,1)$-sheaves. In fact, it is this subcategory that is usually meant when considering stacks.

A more conceptual (although completely equivalent) generalization from (1-)sheaves to 2-sheaves can be obtained by starting from Property 14.3.7. There, it was shown that matching families for (1-)presheaves can be obtained as natural transformations from the Čech groupoid.

Property 14.3.12 (Descent data and Čech nerve). Let $C(\mathcal{U})$ denote the 3-coskeleton of the Čech nerve $C_\bullet(\mathcal{U})$. Pseudonatural transformations $C(\mathcal{U}) \Rightarrow F$ can be shown to be equivalent to tuples $(c, \{c_i\}, \{c_{ij}\}, \{c_{ijk}\})$, with $c_i \in Fx_i$, that fit into cubes lying in the image of $C_2(\mathcal{U})$ in which all edges consist of Cartesian morphisms. Arrows between such cubes are given by arrows between the vertices that make the 'obvious' diagrams commute.

By comparing these cubes to the previous definition of descent data, one obtains the following equivalence:

$$\mathsf{Descent}(\mathcal{U}, F) \cong [\mathbf{C}^{\mathrm{op}}, \mathbf{Cat}]\big(C(\mathcal{U}), F\big)\,. \tag{14.14}$$

@@ FINISH THIS @@

Remark 14.3.13 (1-sheaves). Although most of the above seems very abstract and complex compared to ordinary sheaves, it is not quite so. In fact, when restricting to pseudofunctors of the form $\mathbf{C}^{\mathrm{op}} \to \mathbf{Set}$, where the embedding $\mathbf{Set} \hookrightarrow \mathbf{Cat}$ sends sets to discrete categories, one obtains ordinary sheaves as a subcategory of stacks. For example, by the equivalence between pseudofunctors and Grothendieck fibrations, it is known that the Cartesian pullbacks f^* are in fact just the images of f under the pseudofunctor F. This way, the condition $\pi_1^* c_i \cong \pi_2^* c_j$ can be rewritten as $Ff_i'(c_i) = Ff_j'(c_j)$, which is nothing but the matching family condition (12.18).

14.4 Higher topos theory

In this section, the notion of topos is generalized from ordinary category theory to higher category theory. In particular, ∞-sheaves will be defined. This will require a suitable foundation for ∞-category theory. To this end, the language of (simplicial) model categories as introduced in Chapter 13 will be used.

Definition 14.4.1 (∞-groupoid). Objects of the full simplicial subcategory of $\mathbf{sSet}_{\mathrm{Quillen}}$ on Kan complexes. From Property 13.4.4, it is immediately clear how this generalizes the definition of ordinary groupoids. For groupoids one needs unique horn fillers (composition in ordinary categories is unique), while for ∞-groupoids this is allowed to be unique up to higher coherence.

Definition 14.4.2 ($(\infty,1)$-category). An $\infty\mathbf{Grpd}$-enriched category or, equivalently, a simplicially enriched category for which all hom-objects are Kan complexes. The functor category between $(\infty,1)$-categories is defined through the (simplicial) nerve and realization functors (Example 13.1.10):

$$[\mathbf{C}, \mathbf{D}] := |\mathbf{sSet}(N\mathbf{C}, N\mathbf{D})|\,. \tag{14.15}$$

Property 14.4.3 (Čech model structure). For any small category $\mathbf{C}$, the ∞-category of $\infty\mathbf{Grpd}$-valued ∞-sheaves can be represented by the category $[\mathbf{C}^{\mathrm{op}}, \mathbf{sSet}]$ of simplicial presheaves on $\mathbf{C}$ by a theorem of *Lurie* (Theorem 14.2.1), i.e. there exists an ∞-equivalence between $\mathbf{Sh}_{(\infty,1)}(\mathbf{C})$ and the full subcategory on fibrant-cofibrant objects of the (left Bousfield) localization of $[\mathbf{C}^{\mathrm{op}}, \mathbf{sSet}]$ at the Čech nerve projections. The resulting model structure is called the **Čech model structure**.

A presheaf X is fibrant in this model structure if the map

$$\mathsf{Hom}(M, X) \to \mathsf{Hom}\big(\mathcal{C}(\mathcal{U}), X\big) \tag{14.16}$$

is a weak equivalence for all open covers $\mathcal{U}$, i.e. exactly if X satisfies the descent condition and, hence, is an ∞-stack.

The most straightforward definition of an ∞-sheaf generalizes Definition 12.4.12.

Definition 14.4.4 (∞-sheaf). Consider an ∞-site $(\mathbf{C}, J)$ and let S denote the collection of monomorphisms in $\mathbf{Psh}_\infty(\mathbf{C})$ induced by the covering sieves. An ∞-presheaf on $\mathbf{C}$ is called a J-sheaf if it is S-local. A presheaf with values in an ∞-category $\mathbf{D}$ is called a sheaf if the representable presheaf $\mathbf{D}(x, F-)$ is a J-sheaf for all $x \in \mathrm{ob}(\mathbf{D})$.

In terms of the Čech nerve $\mathcal{C}$, the descent condition can be written as follows:

$$Fx \simeq \mathbf{Psh}_\infty\big(\mathcal{C}(\mathcal{U}), F\big) \tag{14.17}$$

for all covers $\mathcal{U}$ of x, where $\simeq$ denotes a weak equivalence.

Definition 14.4.5 (∞-stack). An $(\infty, 1)$-sheaf taking values in $\infty\mathbf{Grpd}$.

Definition 12.4.22 can be generalized as follows.

Property 14.4.6. For every ∞-topos $\mathbf{H}$, there exists a geometric morphism $(\mathrm{Disc} \dashv \Gamma) : \mathbf{H} \leftrightarrows \infty\mathbf{Grpd}$. Any morphism into a discrete object $\mathrm{Disc}(X)$ is constant.

The left adjoint is sometimes called the **discrete object functor**. This terminology stems from the case of the forgetful functor $\Gamma : \mathbf{Top} \to \mathbf{Set}$, where the (fully faithful) left adjoint equips a set with the discrete topology.

Example 14.4.7 (Sheaves on manifolds). One of the archetypal examples of ∞-topoi is the topos of sheaves over smooth manifolds. By the Yoneda embedding, one can regard a manifold as a sheaf and the global sections functor maps this representable sheaf to the manifold itself: $\Gamma(M) = M$. For a Lie group, one can construct the classifying stack $\mathcal{B}G$. The global sections functor maps this stack to the delooping groupoid $\mathbf{B}G$.

Definition 14.4.8 (Mapping stack). Consider two ∞-stacks $X, Y \in \mathbf{Sh}_{(\infty,1)}(\mathbf{C})$. The mapping stack is defined as follows:

$$[X, Y](U) := \mathbf{Sh}_{(\infty,1)}(\mathbf{C})(X \times U, Y)\,, \tag{14.18}$$

where on the right-hand side, U denotes the representable ∞-stack.

@@ FINISH (PERHAPS MOVE infinity-CATEGORY STUFF TO CHAPTER ON MODEL THEORY) @@

14.5 Cohomology

In this section, cohomology will be generalized to the ∞-categorical setting.

First, take a topological space X and an ∞-groupoid G. Geometric realization (Example 13.1.11) gives an equivalence $\infty\mathbf{Grpd} \cong \mathbf{Top}$ and, therefore, one can define the intrinsic cohomology of X with coefficients in G as follows:

$$H(X;G) := \pi_0 \mathbf{Top}(X, |G|)\,. \tag{14.19}$$

X can also be identified with its petit (∞-)topos $\mathbf{Sh}_{(\infty,1)}(X)$, in which X sits as the terminal object. From this point of view, the intrinsic cohomology of X with coefficients in G is

$$\overline{H}(X;G) := \pi_0 \mathbf{Sh}_{(\infty,1)}(X)(X, \mathsf{LConst}\, G) \cong \pi_0 \circ \Gamma \circ \mathsf{LConst}(G)\,. \tag{14.20}$$

This is the **cohomology with constant coefficients** of X with coefficients in G. If X is paracompact, the two cohomologies coincide: $H(X;G) \cong \overline{H}(X;G)$.

Now, it is time to pass to general cohomology.

Definition 14.5.1 (Intrinsic cohomology). Consider a $(\infty,1)$-category $\mathbf{H}$. For every two objects $X, A \in \mathbf{H}$, the hom-space $\mathbf{H}(X,A)$ is an ∞-groupoid. Define the following notions:

- The objects in $\mathbf{H}(X,A)$ are called **cocycles**.
- The morphism in $\mathbf{H}(X,A)$ are called **coboundaries**.
- The set of connected components

$$H(X;A) := \pi_0 \mathbf{H}(X,A) = \mathsf{Hom}_{\mathbf{Ho}(\mathbf{H})}(X,A)\,, \tag{14.21}$$

  where $\mathbf{Ho}(\mathbf{H})$ is the homotopy category 13.3.27 of $\mathbf{H}$, is called the intrinsic cohomology of X with coefficients in A.

If the object A admits an n-delooping $\mathbf{B}^n A$, the nth cohomology group of X is defined as

$$H^n(X;A) := H(X;\mathbf{B}^n A)\,. \tag{14.22}$$

Example 14.5.2 (Singular cohomology). Consider a topological space X. For every group G one can define the first delooping (Example 3.5.11), so one can also define the zeroth and first cohomology groups $H^{0,1}(X;G)$. Only when G is Abelian do higher deloopings exists (in fact, if G is Abelian all higher deloopings exist), and so in this case higher cohomology groups $H^{\geq 2}(X;G)$ can be defined. It can be shown that these coincide with the singular cohomology groups of X.

Example 14.5.3 (Group cohomology). Consider a (discrete) group G together with its delooping groupoid $\mathbf{B}G$. The cohomology of a group with coefficients in an Abelian group A (Section 5.6.1) is given by the intrinsic cohomology of $\infty\mathbf{Grpd}$ of the delooping groupoids:

$$H(G;A) \cong \pi_0 \infty\mathbf{Grpd}(\mathbf{B}G, \mathbf{B}A)\,. \tag{14.23}$$

By replacing the topos $\mathbf{H}$ by a slice topos $\mathbf{H}_{/X}$, one obtains twisted cohomology.

Definition 14.5.4 (Twisted cohomology). Consider a $(\infty, 1)$-topos $\mathbf{H}$ with some object $X \in \mathtt{ob}(\mathbf{H})$. The mapping space $\mathbf{H}(X, A)$, the cocycles of X with coefficients in A, is easily seen to be isomorphic to the mapping space $\mathbf{H}_{/X}(X, X \times A)$, where the second argument is equipped with the canonical projection morphism. Morphisms in this space are just sections of the trivial A-∞-bundle over X. General twisted cohomology can then be defined as the space of sections of an arbitrary A-∞-bundle over X.

By passing to classifying morphisms of bundles, one obtains the twist $\chi : X \to \mathbf{BAut}(A)$ and the universal bundle $\rho_A : A /\!\!/ \mathbf{Aut}(A) \to \mathbf{BAut}(A)$. χ-twisted cohomology is then given by (the connected components of) the following mapping space:

$$\mathbf{H}_{/\mathbf{BAut}(A)}(\chi, \rho_A)\,. \tag{14.24}$$

Chapter 15

Logic and Type Theory

The main reference for this chapter is The Univalent Foundations Program (2013). For a formal introduction to λ-calculus, see Selinger (2008).

In almost every section of this chapter (at least the ones about type theory), some cross-references to analogous definitions and propositions in other parts of this compendium could have been inserted (Chapter 3 on category theory in particular). However, to reduce the number of references, these relations will only be mentioned and the reader is encouraged to take a look at the relevant chapters whilst or after reading this chapter.

Chapter Contents

15.1 Logic **325**

15.1.1 Languages 325

15.1.2 Propositional logic 326

15.1.3 Sequent calculus 327

15.1.4 Predicate logic 328

15.2 Introduction to type theory **328**

15.3 Basic constructions **329**

15.3.1 Functions 329

15.3.2 λ-calculus 330

15.3.3 Products 330

15.3.4 Propositions as types 334

15.3.5 Identity types 334

15.4 Categorical semantics **337**

15.4.1 Inductive types 337

15.4.2 Polynomial functors 337

15.5 Homotopy type theory **338**
15.5.1 Introduction 338
15.5.2 Transport 339
15.5.3 Equivalences 341
15.5.4 Equality types: revisited 342
15.6 Modal logic **343**
15.6.1 Modalities 343
15.6.2 Type theory 344

15.1 Logic

15.1.1 Languages

Definition 15.1.1 (Language). An **alphabet** is a set of symbols. A **word** in the language is a string of symbols in the alphabet.

Consider an alphabet A. From this alphabet one can construct the free monoid A^* (the multiplication $*$ is sometimes called the **Kleene star**). This monoid represents the set of all words in A and a (formal) language is a subset $L \subseteq A^*$.

Definition 15.1.2 (Signature). Consider an alphabet A and a language L. A signature is a tuple (F, R, ar) that assigns a syntactic meaning to the symbols in A. F and R are respectively the sets of function symbols and relation symbols ($A = F \sqcup R$). The function $\mathsf{ar} : A \to \mathbb{N}$ assigns to every symbol its arity **arity**. Nullary function symbols are also called **constants**.

To give meaning to a language, some extra structure needs to be introduced.

Definition 15.1.3 (L-structure). Consider a (formal) language L. An L-structure consists of the following data:

1. A nonempty set U called the **universe**.
2. For each function symbol f, a function $\mathsf{ap}_f : U^{\mathsf{ar}(f)} \to U$. In particular, for each constant c, an element $u_c \in U$.
3. For each relation symbol $\in$, a set $R_\in \subseteq U^{\mathsf{ar}(\in)}$.

Definition 15.1.4 (L-term). A word in L, possibly containing new symbols (called **variables**), defined recursively as follows:

1. Every variable and every constant is a term.

2. For every n-ary function symbol f and terms $x_1, \dots, x_n$, $f(x_1, \dots, x_n)$ is also a term.

Definition 15.1.5 (L-formula). Consider a (formal) language L. An L-formula is a sentence consisting of terms in L together with parentheses and the following logical symbols (also called **logical connectives**):

- **Equality**: $=$,
- **Negation**: $\neg$,
- **Conjunction**: $\wedge$, and
- **Existential quantification**: $\exists$.

A variable is said to be **free** if it does not first appear next to a quantifier, otherwise it is said to be **bound**.

15.1.2 Propositional logic

Definition 15.1.6 (Proposition). A statement that is either *true* or *false* (not both).

Definition 15.1.7 (Paradox). A statement that cannot (consistently) be assigned a truth value.

Definition 15.1.8 (Contradiction). A statement that is always *false*.

Definition 15.1.9 (Tautology). A statement that is always *true*.

Notation 15.1.10 (Truth values). The truth values *true* and *false* are denoted by $\top$ and $\bot$ respectively.

Definition 15.1.11 (Logical connectives). The following logical operators are used in propositional logic:

- logical 'and' (**conjunction**): $P \wedge Q$,
- logical 'or' (**disjunction**): $P \vee Q$, and
- logical 'then' (**implication**): $P \implies Q$.

Using implication, one can also define the logical 'not' (**negation**): $\neg P :\equiv P \implies \bot$.

The basic inference rule is given by **modus ponens**:

$$\text{If } P \text{ and } P \implies Q, \text{ then } Q. \tag{15.1}$$

One could also use negation as a primitive connective and introduce implication as

$$P \implies Q :\equiv \neg P \wedge Q\,. \tag{15.2}$$

The general deductive system for propositional logic is obtained by combining this rule with the following axioms:

1. If P, then $Q \implies P$.
2. If $P \implies Q \implies R$, then $P \implies Q$ implies $P \implies R$.
3. If $P \wedge Q$, then both P and Q.
4. If P, then $P \vee Q$.
5. If Q, then $P \vee Q$.
6. If P, then Q implies $P \wedge Q$.
7. If $P \implies Q$, then $R \implies Q$ implies $P \vee R \implies Q$.
8. If $\perp$, then P. This principle is often called ***ex falso quodlibet***.

Property 15.1.12 (Boolean algebra). The set of propositions in classical logic admits the structure of a complete Boolean algebra (Definition 1.7.45).

Remark 15.1.13 (Intuitionistic logic). The above axioms (together with modus ponens) define a specific type of propositional logic, called intuitionistic or **constructive** (propositional) logic. The main difference with classic logic is that the *law of the excluded middle* or, equivalently, the *double negation elimination* principle was not added. The reason why this makes the logic *constructive* is that to prove a statement it is not sufficient anymore to exclude the possibility of the statement being false. One has to explicitly construct evidence for the truth of the statement.

As was remarked in the chapter on topoi, intuitionistic logic can be defined internal to any elementary topos. All one needs is a Heyting algebra (Definition 1.7.44).

@@ EXPLAIN THIS @@

15.1.3 Sequent calculus

Definition 15.1.14 (Sequent). A general sequent is of the form

$$P_1, \dots, P_m \vdash Q_1, \dots, Q_n \,. \tag{15.3}$$

In such expressions, the commas on the left-hand side indicate conjunction, whereas those on the right-hand side indicate disjunction, i.e. this sequent states "when every P_i holds, then at least one of the Q_j hold as well". The above sequent if (strongly) equivalent to

$$\vdash (P_1 \wedge \cdots \wedge P_m) \implies (Q_1 \vee \cdots \vee Q_n) \,. \tag{15.4}$$

@@ COMPLETE @@

15.1.4 Predicate logic

@@ ADD @@

15.2 Introduction to type theory

In ordinary set theory, the main objects are sets and their elements (and derived concepts such as functions). The framework in which to state and prove propositions is (in general) given by first-order logic. (See Section 1.1 for more on this.) In type theory, however, one puts all these notions on the same footing. That is, one considers all concepts such as functions, propositions, sets, etc. as specific instances of the general notion of *type*. A specific function, proof or element can then be seen as an *inhabitant* of a given type.

Definition 15.2.1 (Type judgement). A statement of the form $a : A$, saying that a has the type A, is called a type judgement. Objects having a certain type are in general called **terms** (of that type).

Method 15.2.2 (Type definition). The general method for defining a new type consists of 4 steps/rules:

1. **Formation rule**: This rule says when the new type can be introduced, given a collection of pre-existing types.
2. **Introduction rule**: This rule gives a **constructor** of the new type, i.e. a way to construct a term of the new type[1], in terms of the types required by the formation rule. The pre-existing terms from which a new term can be constructed is often called the **context**.
3. **Elimination rule**: This rule says how the new type can be used.
4. **Computation rule**: This rule says how the elimination and introduction rules interact, i.e. how the elimination rules can actually be applied to a term of the given type.

As in The Univalent Foundations Program (2013), a universe hierarchy à la Russell will be adopted, i.e. a sequence of universes $(\mathcal{U}_n)_{n\in\mathbb{N}}$ will be used where the terms of every universe are types and every universe is cumulative in the sense that $A : \mathcal{U}_n \implies A : \mathcal{U}_{n+1}$. In general, the subscripts will be omitted. However, one should take into account that every well-typed judgement should admit a formulation in which subscripts can be assigned in a consistent way.

In contrast to ordinary set theory, two kinds of equality will be considered. First, there

[1] As in object-oriented programming languages.

is the **judgemental** or **definitional equality**. This says, as the name implies, that two judgements are equal by definition and, as such, its validity lives in the metatheory (it is not a proposition and, hence, cannot be proven). For example, if $f(x)$ is defined as x^2, then $f(5)$ is, by definition, equal to 5^2. Equalities of this sort will be denoted by the $\equiv$ symbol (and, in definitions, $:\equiv$ will be used instead of $:=$). The second equality is the **propositional equality**. This states that two judgements are provably equal. Again, consider the function $f(x) :\equiv x^2$. In this case, the proposition $f(5) = 25$ is not true by definition and can be proven (it would, however, depend on the definition of the natural numbers). This sort of equality will be denoted by an ordinary equals sign $=$.

15.3 Basic constructions

15.3.1 Functions

Functions can be introduced in two ways. Either through a direct definition, such as in the case of the default example $f(x) :\equiv x^2$, or through *λ-abstraction*. Although the former one is clearly more useful during explicit calculations, the latter will often be used when working with abstract proofs. (For an introduction to λ-calculus, see the next section.)

Definition 15.3.1 (Function type). A general function type is introduced as follows:

- **Formation rule**: Given two types $A, B : \mathcal{U}$, one can form the function type $A \to B : \mathcal{U}$.
- **Introduction rule**: One can either define a function by an explicit definition $f(x) :\equiv \Phi$, where Φ is an expression possibly involving x, or by λ-abstraction $f :\equiv \lambda x.\Phi$.
- **Elimination rule**: If $a : A$ and $\lambda x.\Phi : A \to B$, then $\lambda x.\Phi(a) : B$.
- **Computation rule**[2]: $\lambda x.\Phi(a) :\equiv \Phi(a)$, i.e. function application is equivalent to the substitution of a for the variable x in the expression Φ. (To be completely correct, one should require the substitution to be *capture-avoiding*, i.e. free variables should remain free and distinct variables should not be assigned the same symbol.)

The **uniqueness principle** for function types should also be included in the definition, i.e. $\lambda x.f(x) \equiv f$. This says that every function is uniquely defined by its image.

An important generalization is obtained when the type of the output of a function is allowed to depend on the type of the input.

[2] In λ-calculus, this is often called β-reduction. (See the next section.)

Definition 15.3.2 (Dependent function types). Given a type $A : \mathcal{U}$ and a type family $B : A \to \mathcal{U}$, one can form the dependent function type

$$\prod_{a:A} B(a) : \mathcal{U} . \tag{15.5}$$

When B is a constant family, this type reduces to the ordinary function type $A \to B$. All other defining rules remain (formally) the same as in the nondependent setting.

Remark 15.3.3 (Scope). The Π-symbol scopes over all expressions to the right of the symbol unless delimited (similar to λ-calculus), e.g.

$$\prod_{a:A} B(a) \to C(a) \equiv \prod_{a:A} \big(B(a) \to C(a)\big) . \tag{15.6}$$

Example 15.3.4 (Polymorphic functions). An interesting example is obtained when the type A in the above definition is taken to be a universe $\mathcal{U}$ (this is a valid choice since universes are themselves types) together with $B(A) :\equiv A$. In this case, one obtains a function that takes a type as input and then acts on this type (or any other type constructed from it), e.g. the **polymorphic identity function**

$$\mathsf{id} : \prod_{A:\mathcal{U}} A \to A \tag{15.7}$$

defined by

$$\mathsf{id} :\equiv \lambda(A : \mathcal{U}).\lambda(a : A).a . \tag{15.8}$$

15.3.2 λ-calculus

@@ COMPLETE (e.g. Curry–Howard or even Curry–Howard–Lambek, typed vs. untyped calculus, ...) @@

15.3.3 Products

As in classic set theory, one of the basic notions is that of a product. This construction is ubiquitous throughout all of mathematics (and computer science). However, as opposed to set theory à la ZFC, products are not explicitly constructed as the set of all pairs of elements of its constituents. On the contrary, in type theory, one can prove that all elements necessarily have to be pairs.

Definition 15.3.5 (Product). First, the binary product of types is defined.

- **Formation rule**: Given any two types $A, B : \mathcal{U}$, one can form the product type $A \times B : \mathcal{U}$.
- **Introduction rule**: Given terms $a : A, b : B$, one can construct the term $(a, b) : A \times B$. This is called the **pairing** of the terms a and b.

- **Elimination and computation rules**: Functions out of a product $A \times B$ are defined through currying, i.e. given a function $A \to B \to C$, one can define a function $A \times B \to C$. Instead of giving an explicit definition every time one wants to construct a new function, a universal point of view is adapted: a single function that turns terms $f : A \to B \to C$ into terms $g : A \times B \to C$ is constructed. To this end, consider the **recursor**

$$\mathsf{rec}_{A\times B} : \prod_{C:\mathcal{U}} (A \to B \to C) \to A \times B \to C \tag{15.9}$$

with the constraint

$$\mathsf{rec}_{A\times B}\big(C, f, (a, b)\big) :\equiv f(a)(b)\,. \tag{15.10}$$

Example 15.3.6 (Projections). Analogous to the projection functions associated to the Cartesian product, one should have functions $\pi_1 : A \times B \to A$ and $\pi_2 : A \times B \to B$ that act on constructors as

$$\pi_1(a, b) :\equiv a \qquad \text{and} \qquad \pi_2(a, b) :\equiv b\,. \tag{15.11}$$

Using the recursor, one can define these functions by taking $C = A, f = \lambda a.\lambda b.a$ and $C = B, f = \lambda a.\lambda b.b$, respectively.

Definition 15.3.7 (Nullary product). One can also define a nullary product. In this case, it is called the **unit type 1**.

- **Formation rule**: $\mathbf{1} : \mathcal{U}$.
- **Introduction rule**: There is a unique nullary constructor $* : \mathbf{1}$.
- **Elimination and computation rules**: Since the constructor is a nullary operation, one does not expect to have projection maps and, likewise, one also does not expect function definition to be based on binary currying. Instead, the recursor is defined as follows:

$$\mathsf{rec}_{\mathbf{1}} : \prod_{C:\mathcal{U}} C \to \mathbf{1} \to C\,. \tag{15.12}$$

On the constructor $* : \mathbf{1}$, it is required to act trivially:

$$\mathsf{rec}_{\mathbf{1}}(C, c_0, *) :\equiv c_0\,. \tag{15.13}$$

Definition 15.3.8 (Dependent functions). One can easily generalize the above recursors to **inductors**, to allow for the definition of dependent functions out of product types (these functions are then said to be defined by an **induction principle**). In fact, one only has to change the type judgement of $\mathsf{rec}_{A\times B}$. This is accomplished by replacing $C : \mathcal{U}$ by a type family $C : A \times B \to \mathcal{U}$ and by replacing nondependent function types

by dependent function types (the form of the computation rules virtually remain the same):

$$\begin{aligned} \mathsf{ind}_{A\times B} &: \prod_{C:A\times B\to\mathcal{U}} \left(\prod_{a:A,b:B} C(a,b) \to \prod_{x:A\times B} C(x) \right), \\ \mathsf{ind}_{\mathbf{1}} &: \prod_{C:\mathbf{1}\to\mathcal{U}} C(*) \to \prod_{x:\mathbf{1}} C(x)\,. \end{aligned} \tag{15.14}$$

Property 15.3.9 (Uniqueness principle). Using the induction principle, one can prove that every term $x : A \times B$ is necessarily of the form (a, b) for some $a : A, b : B$. Furthermore, one can also prove that $* : \mathbf{1}$ is the unique term in $\mathbf{1}$.

One can also generalize products in such a way that the type of the second factor depends on the type of the first one (in classical set theory, this would correspond to an indexed disjoint union).

Definition 15.3.10 (Dependent pair type). As with function types, the definition is not given as explicit as for nondependent types. Suffice it to say that, given a type $A : \mathcal{U}$ and a type family $B : A \to \mathcal{U}$, one can form the dependent pair type

$$\sum_{a:A} B(a) : \mathcal{U}\,. \tag{15.15}$$

When B is a constant family, the type reduces to the ordinary product type $A \times B$. The recursion and induction functions are defined as in the product case, except for the obvious replacements, such as $A \times B \longrightarrow \sum_{a:A} B(a)$, needed to make everything consistent.

Remark 15.3.11. Dependent pair types are often called **Σ-types** (due to the notation).

Remark 15.3.12 (Scope). Like the Π-symbol, the Σ-symbol scopes over the entire expression to the right unless delimited.

Definition 15.3.13 (Coproduct). Here, a standalone definition is given. The relation with the ordinary product will be mentioned afterwards.

- **Formation rule**: Given two types $A, B : \mathcal{U}$, one can form the coproduct type $A+B : \mathcal{U}$.
- **Introduction rule**: Since in ordinary mathematics (and, in particular, category theory) the coproduct is dual to the product, one expects the projections to be replaced by **injections/inclusions**. In fact, these are taken to be the constructors of coproduct types, i.e. given terms $a : A$ and $b : B$, one can construct the terms $\iota_1(a) : A + B$ and $\iota_2(b) : A + B$.
- **Elimination rules**: Similar to the use of currying for the definition of functions out of a product, functions out of a coproduct are defined in steps. To this intent

the recursor and inductor are defined as follows:

$$\begin{aligned}
\mathsf{rec}_{A+B} &: \prod_{C:\mathcal{U}} (A \to C) \to (B \to C) \to A + B \to C\,, \\
\mathsf{ind}_{A+B} &: \prod_{C:A+B\to\mathcal{U}} \left(\prod_{a:A} C\big(\iota_1(a)\big) \right) \to \left(\prod_{b:B} C\big(\iota_2(b)\big) \right) \to \prod_{x:A+B} C(x)\,.
\end{aligned} \tag{15.16}$$

- **Computation rules**: The recursor acts on the constructors as follows (the inductor virtually has the same action):

$$\begin{aligned}
\mathsf{rec}_{A+B}\big(C, f_1, f_2, \iota_1(a)\big) &:\equiv f_1(a)\,, \\
\mathsf{rec}_{A+B}\big(C, f_1, f_2, \iota_2(b)\big) &:\equiv f_2(b)\,.
\end{aligned} \tag{15.17}$$

Definition 15.3.14 (Nullary coproduct). As was the case for products, one can also define a nullary version of the coproduct, the **empty type 0**:

- **Formation rule**: $\mathbf{0} : \mathcal{U}$.
- **Introduction rule**: There is no constructor for $\mathbf{0}$.
- **Elimination and computation rules**: Since there is no constructor for $\mathbf{0}$, one can always trivially 'construct' a function out of $\mathbf{0}$:

$$\begin{aligned}
\mathsf{rec}_{\mathbf{0}} &: \prod_{C:\mathcal{U}} \mathbf{0} \to C\,, \\
\mathsf{rec}_{\mathbf{0}} &: \prod_{C:\mathbf{0}\to\mathcal{U}} \prod_{x:\mathbf{0}} C(x)\,.
\end{aligned} \tag{15.18}$$

  This trivial function corresponds to the logical principle *ex falso quodlibet* as introduced in the section on logic above.

Since coproducts in set theory occur as binary disjoint unions, one could expect that there is a way to express coproducts in terms of dependent pair types.

Construction 15.3.15 (Coproducts as Σ-types). First, introduce the type $\mathbf{2} : \mathcal{U}$ (in set theory, this would be the 2-element set). The introduction rule constructs two terms $0, 1 : \mathbf{2}$. The elimination and computation rules say that one can use this type for binary indexing:

$$\mathsf{rec}_{\mathbf{2}} : \prod_{C:\mathcal{U}} C \to C \to \mathbf{2} \to C \tag{15.19}$$

with

$$\begin{aligned}
\mathsf{rec}_{\mathbf{2}}(C, c_0, c_1, 0) &:\equiv c_0\,, \\
\mathsf{rec}_{\mathbf{2}}(C, c_0, c_1, 1) &:\equiv c_1\,.
\end{aligned} \tag{15.20}$$

Using this type, one can prove that $A + B$ is judgementally equal to $\sum_{x:\mathbf{2}} \mathsf{rec}_{\mathbf{2}}(\mathcal{U}, A, B, x)$. The injections are given by pairing, i.e. $\iota_1(a) \equiv (0, a)$ and $\iota_2(b) \equiv (1, b)$. In a similar way, one can obtain binary products as dependent function types over $\mathbf{2}$.

15.3.4 Propositions as types

To conclude this section, an overview of all the concepts introduced above is given from a propositions-as-types perspective. In intuitionistic logic, this is often called the **Brouwer–Heyting–Kolmogorov interpretation** and, more specifically, it should be seen as an incarnation of the Curry–Howard correspondence.

- Types and their terms correspond to propositions and their proofs, respectively. In a proof-relevant context the fact that a type can have multiple terms makes it clear that, although distinct proofs eventually have the same result, the difference in their content can be important as well.
- Function types correspond to implications. A proof of the proposition $A \to B$ boils down to showing that every proof of A gives a proof of B.
- Π-types correspond to universal quantification, i.e. $\prod_{a:A} B(a)$ can be read as $\forall a \in A : B(a)$. Giving a proof of $\prod_{a:A} B(a)$ is the same as giving for every $a : A$ a proof of $B(a)$. This is indeed compatible with the fact that elements of Π-types are dependent functions, i.e. every element $a : A$ gives rise to a (possibly) distinct type/proposition.
- Σ-types correspond to existential quantification, i.e. $\sum_{a:A} B(a)$ can be read as $\exists a \in A : B(a)$. Giving a proof of $\sum_{a:A} B(a)$ is the same as giving a proof for some $(a, B(a))$. This is compatible with the fact that Σ-types can be identified with disjoint unions and hence every element can be associated with a specific constituent type.
- The logical connectives (conjunction and disjunction) correspond to the product and coproduct types.
- The truth values, *true* and *false*, correspond to the unit and empty types, respectively. Furthermore, if the negation of A is defined as the type $\neg A :\equiv A \to \mathbf{0}$, this indeed corresponds to the logical negation by the statements above.

15.3.5 Identity types

One of the most important, but at the same time most subtle, concepts in type theory (especially when moving on to extensions such as homotopy type theory) is the identity type. Since in predicate (and even propositional) logic the equality of two terms is a proposition, one could expect that to every two terms $a, b : A$ there corresponds an associated equality type $a =_A b : \mathcal{U}$. Note that the type of the terms is assumed to be the same since it does not make any sense to compare terms of different types.

[3]Sometimes called an **identity type**.

Definition 15.3.16 (Equality type[3]). The type corresponding to a propositional equality is defined by the following rules:

- **Formation rule**: Given terms $a, b : A$, one can form the equality type $a =_A b : \mathcal{U}$. When the type A is clear from the context, this is also often written as $a = b : \mathcal{U}$.
- **Introduction rule**: For every term $a : A$, there is a canonical identity element

$$\mathsf{refl}_a : a = a\,. \tag{15.21}$$

  The notation points to the fact that this term can be seen as a proof of the reflexivity of equalities.
- **Elimination and computation rules**: Here, the so-called **path induction principle** for equality types is presented, for the equivalent *based path induction principle* see The Univalent Foundations Program (2013).

  Given a type family

$$C : \prod_{a,b:A} a = b \to \mathcal{U}$$

  and a term

$$I : \prod_{a:A} C(a, a, \mathsf{refl}_a)\,,$$

  there exists a function

$$f : \prod_{a,b:A} \prod_{p:a=b} C(a, b, p) \tag{15.22}$$

  such that

$$f(a, a, \mathsf{refl}_a) :\equiv I(a) \tag{15.23}$$

  for all $a : A$.

Informally this principle says that all terms of the form (a, b, p), with $p : a = b$, are inductively generated by the 'constant' terms (a, a, refl_a). (See the section on homotopy type theory for a more geometric perspective).

Using the notion of identity types one can say when a given type resembles a proposition:

Definition 15.3.17 (Mere proposition). A type $A : \mathcal{U}$ for which

$$\mathsf{isProp}(A) :\equiv \prod_{a,b:A} a = b \tag{15.24}$$

is inhabited. This is also called an ***h*-proposition**.

Given a mere proposition $P : \mathcal{U}$, the related identity type is either uninhabited, if P itself is uninhabited, or has a unique term. Types of this form are said to be **contractible**.

Definition 15.3.18 (Contractible type). A type $A : \mathcal{U}$ for which

$$\mathsf{isContr}(A) :\equiv \sum_{a:A} \prod_{b:A} a = b \tag{15.25}$$

is inhabited.

Definition 15.3.19 (Homotopy level). The homotopy level or h**-level** of a type $A : \mathcal{U}$ is recursively defined as follows:

$$\mathsf{hasHLevel}(0, A) :\equiv \mathsf{isContr}(A) \tag{15.26}$$

$$\mathsf{hasHLevel}(n+1, A) :\equiv \prod_{a,b:A} \mathsf{hasHLevel}(n, a = b)\,. \tag{15.27}$$

A type of homotopy level $n + 2$ is also called a **homotopy** n**-type**.

Example 15.3.20. The lowest h-levels are given by

-2 : Contractible types,

-1 : mere propositions,

0 : sets,

1 : groupoids, ...

Because of the inductive nature of identity types, any homotopy n-type can be truncated to a homotopy k-type for $k < n$. In the case of a (-1)-truncation, the following notion is obtained.

Definition 15.3.21 (Bracket type). Consider a type $A : \mathcal{U}$. The bracket $[A] : \mathcal{U}$ or $[A] : \mathcal{U}$ is inhabited (and uniquely so) if and only if A is inhabited. It is defined by the following rules:

$$\mathsf{isInhab} : A \mapsto [A]\,, \tag{15.28}$$

$$\mathsf{propTrunc} : \prod_{a,b:A} \mathsf{isInhab}(a) = \mathsf{isInhab}(b)\,. \tag{15.29}$$

The term $\mathsf{propTrunc}$ exactly says that $[A]$ is a mere proposition.

Definition 15.3.22 (Homotopy fibre). Consider a function $f : A \to B$. For every term $b : B$, the homotopy fibre of f is defined as follows:

$$\mathsf{hfibre}(b, f) :\equiv \sum_{a:A} f(a) = b\,. \tag{15.30}$$

Definition 15.3.23 (Equivalence). A function $f : A \to B$ such that

$$\mathsf{isEquiv}(f) :\equiv \prod_{b:B} \mathsf{isContr}\big(\mathsf{hfibre}(b, f)\big) \tag{15.31}$$

is inhabited.

Slightly different notions exist. A **homotopy equivalence** is a function $f : A \to B$ such that there exists a function $g : B \to A$ and homotopies

$$p : \prod_{a:A} g\big(f(a)\big) = a \qquad \text{and} \qquad q : \prod_{b:B} f\big(g(b)\big) = b. \tag{15.32}$$

It is called an **adjoint equivalence** if it is a homotopy equivalence equipped with higher homotopies representing triangle identities for f and g.

Property 15.3.24. The types $\mathsf{isEquiv}(f)$, $\mathsf{isHomEquiv}(f)$ and $\mathsf{isAdjEquiv}(f)$ are co-inhabited.

15.4 Categorical semantics

15.4.1 Inductive types

Inductive types also admit semantics in category theory. The right concept for these is that of algebras over endofunctors (Definition 3.4.24). The recursion principle and accompanying computation rule of inductive types exactly state for every other F-algebra A there exists a morphism $T \to A$ and that this is a unique algebra morphism, i.e. inductive types are initial algebras over endofunctors. An induction principle assigns to every algebra morphism $B \to T$, where B should be interpreted as the total space $\sum_{x:T} B(t)$, a section $T \to B$. The computation rule then again says that this section is an algebra morphism.

Property 15.4.1. When working in sets, the recursion and induction principles are equivalent.

Remark 15.4.2. When passing from extensional to intensional type theory, one has to replace initial algebras by weakly initial algebras.

15.4.2 Polynomial functors

Definition 15.4.3 (Polynomial). A diagram of the form

$$W \xleftarrow{f} X \xrightarrow{g} Y \xrightarrow{h} Z\,. \tag{15.33}$$

In terms of ordinary polynomials, Y gives the index set of the terms, X as a subset of $\mathbb{N}$ gives the degrees of the terms and W gives the coefficients. If g is an identity morphism, one often speaks of **linear functors**.

Given a polynomial in a locally Cartesian closed category $\mathbf{C}$, the induced polynomial functor is given by the composition

$$\mathbf{C}/W \xrightarrow{f^*} \mathbf{C}/X \xrightarrow{\prod_g} \mathbf{C}/Y \xrightarrow{\sum_h} \mathbf{C}/Z\,, \tag{15.34}$$

where $\prod_g$ and $\sum_h$ are dependent product and sum functors (Definition 12.2.2).

Definition 15.4.4 (W-type). Consider a type A and a type family $B : A \to \mathcal{U}$. The W-type $W_{a:A}B(a)$ is obtained as the initial algebra for the polynomial functor

@@ COMPLETE @@

15.5 Homotopy type theory

15.5.1 Introduction

This section gives a reformulation or extension of the concept introduced before using the language of homotopy theory (and, more generally, algebraic topology). The relevant concepts can be found in Section 7.1 and Section 3.5.1. The resulting theory is called homotopy type theory or **HoTT**.

The general idea is to associate types with topological spaces and terms with points in those spaces. The main novelty is given by the identification of (propositional) equalities with paths between points. Since everything happens in a proof-relevant context, two equalities $p, q : a =_A b$ are not necessarily equal themselves and, hence, one can consider equalities between equalities (and so on). In the topological picture this gives rise to homotopies between paths. By going all the way and working out all coherence laws, one obtains the structure of a (weak) ∞-groupoid.[4]

It is also this interpretation that explains the name 'path induction' for the induction principle of equality types. Namely, what this induction principle says is that the free path space ΩA is *inductively generated* by constant loops (ranging over all possible points). This principle, however, sounds quite crazy. How can one build a path between two distinct points from (constant) loops? Here it is important to remind that everything only has to be equal up to homotopy and any path is indeed homotopy-equivalent to a constant loop if one retracts one of the endpoints along the path. It is thus important that one does not require the homotopies to act rel endpoints (as is often done in classical homotopy theory).

Definition 15.5.1 (Pointed type). A type $A : \mathcal{U}$ together with a distinguished term $a : A$, called the **base point**. Pointed types are often denoted by a pair (A, a). It should be clear that the type of pointed types $\mathcal{U}_\bullet$ is equal to $\sum_{A:\mathcal{U}} A$.

Definition 15.5.2 (Loop space). The loop space $\Omega(A, a)$ of a pointed type (A, a) is the pointed type $(a =_A a, \mathsf{refl}_a)$.

Now, the important aspect of HoTT is that the ∞-groupoid structure of a type can be

[4]This characterization is strongly related to the homotopy hypothesis (or theorem when using the right model for ∞-categories).

derived solely from the (path) induction principle of the equality types. Some examples are given.

Property 15.5.3 (Inversion). For every type $A : \mathcal{U}$ and terms $a, b : A$, there exists a function

$$p \mapsto p^{-1} : (a = b) \to (b = a) \tag{15.35}$$

such that $\mathsf{refl}_a^{-1} :\equiv \mathsf{refl}_a$ for all $a : A$.

Property 15.5.4 (Concatenation). For every type $A : \mathcal{U}$ and terms $a, b, c, d : A$, there exists a function

$$p \mapsto q \mapsto p \cdot q : (a = b) \to (b = c) \to (c = d) \tag{15.36}$$

such that $\mathsf{refl}_a \cdot \mathsf{refl}_a :\equiv \mathsf{refl}_a$ for all $a : A$. (Note that the composition does not follow the usual convention of right-to-left. This is why the symbol $\cdot$ and not $\circ$ was used.)

Property 15.5.5. The above operations satisfy the group relations (up to higher equalities):

- $p \cdot \mathsf{refl}_b = p$ and $\mathsf{refl}_a \cdot p = p$ for all $p : a = b$.
- $p \cdot p^{-1} = \mathsf{refl}_a$ and $p^{-1} \cdot p = \mathsf{refl}_b$ for all $p : a = b$.
- $(p^{-1})^{-1} = p$ for all $p : a = b$.
- $p \cdot (q \cdot r) = (p \cdot q) \cdot r$ for all $p : a = b, q : b = c, r : c = d$.

15.5.2 Transport

The relation with homotopy theory and category theory becomes even stronger when looking at function types:

Property 15.5.6. Given a function $f : A \to B$, there exists an **application function**

$$\mathsf{ap}_f : (a =_A b) \to \big(f(a) =_B f(b)\big) \tag{15.37}$$

such that $\mathsf{ap}_f(\mathsf{refl}_a) :\equiv \mathsf{refl}_{f(a)}$ for all $a, b : A$. Furthermore, this function behaves functorially in that it preserves concatenation, inverses and identities (again this should be interpreted in the full weak ∞-sense). From the topological perspective this can be interpreted as if all functions are 'continuous'.

Notation 15.5.7. Because functors in category theory are generally given the same notation when acting on objects or morphisms, the application function ap_f is also often denoted by f.

For dependent functions one can obtain a similar result. However, for this generalization, one needs some kind of 'parallel transport' since for two terms with $a = b$, it does not necessarily hold that $f(a)$ and $f(b)$ have the same type.

Property 15.5.8 (Transport). Given a type family $P : A \to \mathcal{U}$ and an equality $p : a =_A b$, there exists a **transport function**

$$p_* : P(a) \to P(b) \tag{15.38}$$

such that $(\mathsf{refl}_a)_* :\equiv \mathsf{id}(a)$ for all $a : A$. The pushforward notation is used since p_* can be (informally) interpreted as the pushforward of p along P.

From a topological perspective, this transport function allows to regard type families as fibrations (Definition 7.1.61). For every type family $P : A \to \mathcal{U}$, term $\alpha : P(a)$ and equality $p : a = b$, there exists a **lift**

$$\mathsf{lift}(p, \alpha) : (a, \alpha) = \big(b, p_*(\alpha)\big) \tag{15.39}$$

such that

$$\pi_1\big(\mathsf{lift}(p, u)\big) = p\,. \tag{15.40}$$

The equality $\mathsf{lift}(p, u)$ acts between terms of the Σ-type $\sum_{a:A} P(a)$, which can be interpreted as the total space of a **fibration** $\pi_1 : \sum_{a:A} P(a) \to A$. To take this terminology even further, one can could call functions $\sigma : \prod_{a:A} P(a)$ **sections** (of π_1).

Now, as mentioned before, for dependent functions one cannot just compare $f(a)$ and $f(b)$ if $a \not\equiv b$. However, the function $\mathsf{lift}(p, \cdot)$ gives a canonical path from one fibre to the other and every path between these fibres should factor through this canonical path essentially uniquely. Hence, one can define a path between α and β in the total space $\sum_{a:A} P(a)$, lying over $p : a = b$, to be a path $p_*(\alpha) = \beta$ (up to equivalence).

Property 15.5.9. Given a dependent function $f : \prod_{a:A} P(a)$, there exists a function

$$\mathsf{apd}_f : \prod_{p:a=b} p_*\big(f(a)\big) =_{P(b)} f(b)\,. \tag{15.41}$$

Again, with some abuse of notation, this function is also denoted by f

Since an ordinary function is a specific instance of a Π-type, one might expect that the application functions ap_f and apd_f are related in this case. The following property shows that this intuition is not unreasonable.

Property 15.5.10. Consider two types $A, B : \mathcal{U}$ and a function $f : A \to B$. For every equality $p : a =_A b$ and term $\alpha : P(b)$, there exists an equality $\tilde{p} : p_*(\alpha) =_{P(b)} \alpha$. Using this equality one can relate the application functions as follows:

$$\mathsf{apd}_f(p) = \tilde{p}\big(f(a)\big) \cdot \mathsf{ap}_f(p)\,. \tag{15.42}$$

15.5.3 Equivalences

In this paragraph the notions of equivalences and isomorphisms are considered in more detail. As is known from the chapter on category theory, the distinction between the various notions of similarity (or equality) is important yet subtle.

Lead by the intuition from topology a **homotopy** between functions is defined.

Definition 15.5.11 (Homotopy). Consider two sections $f, g : \prod_{a:A} P(a)$. A homotopy between f and g is a term of the type

$$f \sim g :\equiv \prod_{a:A} f(a) = g(a)\,. \tag{15.43}$$

It can be shown that homotopies induce equivalence relations on function types.

It has already been noted that functions can be regarded as functors between ∞-groupoids. Since homotopies act between functions, one might expect that these can be regarded as (weak) natural transformations between the (∞-)functors.

Property 15.5.12. Consider two sections $f, g : \prod_{a:A} P(a)$ and an equality $p : a = b$. If H is a homotopy between f and g, then

$$H(a) \cdot g(p) = f(p) \cdot H(b)\,. \tag{15.44}$$

Using the notion of homotopy one can introduce a first kind of 'equivalence'.

Definition 15.5.13 (Quasi-inverse). Given a function $f : A \to B$, a quasi-inverse of f is a triple (g, α, β), where $g : B \to A$ and

$$\alpha : f \circ g \sim \mathsf{id}_B \qquad\qquad \beta : g \circ f \sim \mathsf{id}_B\,. \tag{15.45}$$

From a homotopy theoretical perspective one would call the pair (f, g) a homotopy equivalence. The corresponding type is given by

$$\mathsf{qInv}(f) :\equiv \sum_{g:B\to A} (f \circ g \sim \mathsf{id}_B) \times (g \circ f \sim \mathsf{id}_A)\,. \tag{15.46}$$

Now, although this type may seem to give the right notion of equivalence, it is better to generalize it since it is in general not very well-behaved. (This is similar to the fact that adjoint equivalences between categories are better behaved than ordinary equivalences.)

In general an equivalence should satisfy three requirements:

1. For every function $f : A \to B$, there exists a function $\mathsf{qInv}(f) \to \mathsf{isEquiv}(f)$.
2. For every function $f : A \to B$, there also exists a function $\mathsf{isEquiv}(f) \to \mathsf{qInv}(f)$.

3. For every two terms $eq_1, eq_2 : \mathsf{isEquiv}(f)$, there exists an equality $eq_1 = eq_2$.

So, inducing an equivalence is logically equivalent to admitting a quasi-inverse and as such finding a quasi-inverse is sufficient to show that a function induces an equivalence.

15.5.4 Equality types: revisited

In the section on (intensional) type theory, equality types were introduced in a general and uniform way. The defining rules did not assume any specific structure on the underlying types. Although this made the technique of path induction widely applicable, it has the downside that one cannot leverage the internal structure of specific types to get more useful characterizations.

First, consider binary products (and by extension Σ-types). Can one express the equality of two elements $x, y : A \times B$ in terms of their projections? The answer is yes: there exists an equivalence

$$(x =_{A\times B} y) \simeq \big(\pi_1(x) =_A \pi_1(y)\big) \times \big(\pi_2(a) =_B \pi_2(y)\big)\,. \tag{15.47}$$

However, one should bear in mind that this is merely an equivalence. A term (resp. proof) of one side gives a term (resp. proof) of the other side, but it is not a judgemental equality (it is not even a propositional one). One could see this as a problem or defect of the theory and to resolve this kind of (apparent) issue the univalence axiom will be introduced at the end of this section. Still, one can leverage this equivalence to give a practical alternative[5] for the defining rules of the equality type in the case of product types:

Remark 15.5.14. The function $\big(\pi_1(a) = \pi_1(b)\big) \times \big(\pi_2(a) = \pi_2(b)\big) \to (a = b)$ associated to the above equivalence can be interpreted as an introduction rule of the equality type for binary products. At the same time one can take the application functions induced by the projections on $A \times B$ as elimination rules for the equality type. The homotopies associated to the equivalence in their turn induce the propositional computation rules and uniqueness principle.

One can also express the transport of properties along an equality $p : x =_{A\times B} y$ in terms of transport in the individual spaces:

Property 15.5.15. Consider two types $A, B : \mathcal{U}$ together with type families $P : A \to \mathcal{U}$ and $Q : B \to \mathcal{U}$. For every term α of the product family $(P\times Q)(x) :\equiv P\big(\pi_1(x)\big)\times Q\big(\pi_2(x)\big)$, the following equality type is inhabited:

$$p_*(\alpha) = \big(p_*(\pi_1(\alpha)), p_*(\pi_2(\alpha))\big)\,. \tag{15.48}$$

[5]Note that this is not a judgementally equal alternative. It is merely a convenient interpretation.

Note that all three occurrences of the pushforward p_* denote a different operation or, more precisely, the same operation but applied to different types.

One would intuitively expect that given two functions $f, g : A \to B$ that take the same value at all points, i.e. $f(a) = g(a)$ for all $a : A$, there exists an equality $f =_{A \to B} g$. However, this cannot be proven within the frame work of intensional type theory. This issue should also not come as a shock, since two functions that are defined differently might still take the same value at all points. To resolve this apparent gap in the theory, the following axiom is introduced.

Axiom 15.1 (Function extensionality). Given two functions $f, g : \prod_{a:A} P(a)$, there exists an equivalence $(f = g) \to \prod_{a:A} f(a) = g(a)$ that sends refl_f to $f(\mathsf{refl}_x)$.

Axiom 15.2 (Univalence axiom). Given two types $A, B : \mathcal{U}$, there exists an equivalence $(A =_{\mathcal{U}} B) \to (A \simeq B)$ that takes refl_A to id_A. A universe in which the univalence axiom holds is said to be univalent.

@@ COMPLETE @@

15.6 Modal logic

15.6.1 Modalities

The two most important or most well-known modalities are 'necessity' and 'possibility'. For any proposition p,

- $\Box p$ means that p is necessarily true, and
- $\Diamond p$ means that p is possibly true.

To formalize these modalities, a few axioms can be introduced.

Axiom 15.3 (K). $\Box$ preserves implication:

$$\Box(p \to q) \to (\Box p \to \Box q)\,. \tag{15.49}$$

In intuitionistic logic, a similar axiom has to be introduced for $\Diamond$:

$$\Box(p \to q) \to (\Diamond p \to \Diamond q)\,. \tag{15.50}$$

Axiom 15.4 (Necessitation). If p is true, then $\Box p$ is true or, equivalently:

$$\texttt{true} \to \Box\texttt{true}\,. \tag{15.51}$$

Axiom 15.5 (T).

$$\Box p \to p \tag{15.52}$$

As for Axiom K, a separate axiom has to be introduced for $\Diamond$ in intuitionistic logic:

$$p \to \Diamond p\,. \tag{15.53}$$

Axiom 15.6 (S4).

$$\Box p \to \Box\Box p \tag{15.54}$$

and

$$\Diamond\Diamond p \to \Diamond p \tag{15.55}$$

The categorical semantics for intuitionistic (propositional) logic is given by the internal language of Cartesian closed categories (Definition 4.1.20). When passing to (S4-)modal logic, one has to specialize these categories. The three axioms above imply that $\Box$ is a monoidal comonad. The analogous axioms for $\Diamond$ imply that it is a $\Box$-strong monad, i.e. it is a monad equipped with a natural transformation

$$\eta_{x,y} : \Box x \otimes \Diamond y \to \Diamond(\Box x \otimes y)\,. \tag{15.56}$$

15.6.2 Type theory

Definition 15.6.1 (Modality). Consider an $(\infty,1)$-topos $\mathbf{H}$. A modality or **modal operator** on $\mathbf{H}$ is an idempotent[6] (co)monad on $\mathbf{H}$ (see also Definition 3.3.27). For clarity, idempotent comonads will be called **comodalities** in this document (unless specifically mentioned, all statements can be dualized to pertain to comodalities).

Definition 15.6.2 (Modal type). A type $X \in \mathbf{H}$ is said to be modal for a modality $\Diamond$ if the unit $\eta_X : X \to \Diamond X$ is an equivalence.

Property 15.6.3. The modal types of a modality $\Diamond$ on $\mathbf{H}$ are exactly the $\Diamond$-algebras or, equivalently, the objects of the Eilenberg–Moore category $\mathbf{H}^{\Diamond}$ (Definition 3.3.20). This property allows to define modal types for monads that are not idempotent as their algebras. It also means that one can recover the monad objectwise as the composition of the unit and the inclusion.

Definition 15.6.4 (Opposite). An **adjoint cylinder** is an adjoint triple $\iota_! \dashv \iota^* \dashv \iota_*$ where the left and, equivalently, the right adjoint are fully faithful:

$$\mathbf{H} \underset{\hookrightarrow}{\overset{\hookrightarrow}{\leftarrow}} \mathbf{H}' \tag{15.57}$$

In the case where these functors are themselves modalities, these can be seen as full inclusions of modal types, turning $\mathbf{H}$ into an essential subtopos of $\mathbf{H}'$ (Definition 12.2.9), and such an adjoint cylinder represents **opposite** modalities. Such a situation also gives

[6] Sometimes, idempotency is not required.

rise to a new pair of adjoint modalities $\iota_! \iota^* \dashv \iota_* \iota^*$, with the left adjoint being a monad (the other adjunction is trivial because of the fully faithfulness).

A dual situation arises for an adjoint triple of the form

$$\mathbf{H} \underset{\leftarrow}{\overset{\leftarrow}{\hookrightarrow}} \mathbf{H}', \tag{15.58}$$

exhibiting $\mathbf{H}$ as a bireflective subcategory of $\mathbf{H}'$ (Definition 3.2.29). Here, the resulting adjoint modalities have as left adjoint a comonad.

Given an opposition of modalities (either of the two situations), one has a modality $\Diamond$ and a comodality $\Box$ and, hence, a unit and counit morphism:

$$\Box x \to x \to \Diamond x\,. \tag{15.59}$$

The composite map is said to represent the **unity** of the opposition in the Hegelian interpretation of *Lawvere*.

Remark 15.6.5 (Interpretation). Note that the adjoint triples in this definition have a slightly different interpretation due to there only being one inclusion arrow in the second case. If one interprets a modality as projecting out some kind of property of objects, the first opposition embodies the case where two related properties are considered, while the second case involves only one property but projected out in two different ways.

Remark 15.6.6 (Adjoint sequences). In some situations, longer sequences of adjoint functors arise (see, for example, Section 44.3). In the case of adjoint quadruples one gets induced adjunctions of the form

$$\mathsf{modality} \dashv \mathsf{comodality} \dashv \mathsf{modality}$$

and

$$\mathsf{comodality} \dashv \mathsf{modality} \dashv \mathsf{comodality}\,.$$

These are sometimes called **Yin** and **Yang triples**, respectively.

Notation 15.6.7. One can show that adjoint triples $F \dashv G \dashv H$ are equivalent to adjoint pairs in the 2-category having adjunctions as morphisms, hence to adjunctions of adjunctions. This inspires the following notations:

$$\begin{array}{ccc} F & \dashv & G \\ \perp & & \perp \\ G & \dashv & H \end{array}$$

Definition 15.6.8 (Negative). Consider a comodality $\bigcirc : \mathbf{H} \to \mathbf{H}$. The negative of a modal type X is obtained by removing its 'pure $\bigcirc$-part', i.e. its projection under $\bigcirc$. Categorically, this means that one takes the cofibre of the counit:

$$\overline{\bigcirc} X := \mathsf{cofib}(\bigcirc X \to X)\,. \tag{15.60}$$

Definition 15.6.9 (Aufhebung). Consider an inclusion of adjoint modalities (or, equivalently, an inclusion of essential subtopoi[7]):

$$\begin{array}{ccc} \Diamond_2 & \dashv & \Box_2 \\ \vee & & \vee \\ \Diamond_1 & \dashv & \Box_1 \end{array} \tag{15.61}$$

Level 2 is said to **resolve** level 1 if $\Diamond_1 < \Box_2$. The smallest resolution (if it exists) is called the (right) Aufhebung (of opposites).

Example 15.6.10 (*Dasein*). Consider the adjoint modality of '(pure) being' and 'nothing' (or non-being):

$$\emptyset \dashv *\,, \tag{15.62}$$

where the former is the constant comonad on the initial object and the latter the constant monad on the terminal object. The associated unity of opposites

$$\emptyset \cong \emptyset X \to X \to *X \cong *\,, \tag{15.63}$$

corresponds to *Hegel*'s "*there is nothing which is not an intermediate state between being and nothing*".

[7] One can show that these form a (complete) lattice and, hence, inclusion is well defined.

Part IV

Calculus

Table of Contents

16 Calculus **351**

16.1 General definitions . . . 351
16.2 Continuity . . . 352
16.3 Convergence . . . 353
16.4 Series . . . 353
16.4.1 Convergence tests . . . 353
16.5 Differentiation . . . 355
16.6 Integration theory . . . 357
16.6.1 Riemann integral . . . 357
16.6.2 Euler integrals . . . 359
16.6.3 Gaussian integrals . . . 360
16.6.4 Generalizations . . . 361
16.7 Convexity . . . 362
16.8 Trigonometry . . . 364

17 Complex Analysis **365**

17.1 Complex algebra . . . 365
17.2 Holomorphic functions . . . 366
17.3 Contour integrals . . . 367
17.4 Laurent series . . . 368
17.5 Singularities . . . 369
17.5.1 Poles . . . 369
17.5.2 Branch cuts . . . 370
17.5.3 Residue theorem . . . 371
17.6 Limit theorems . . . 372

18 Measure & Integration Theory **373**

18.1 Measure theory . . . 374
18.1.1 General definitions . . . 374
18.1.2 Lebesgue measure . . . 376
18.1.3 Measurable functions . . . 377

18.1.4 Limit operations . . . 379
18.1.5 Polish spaces . . . 381
18.2 Lebesgue integral . . . 381
18.2.1 Simple functions . . . 381
18.2.2 Measurable functions . . . 382
18.2.3 Integrable functions . . . 382
18.2.4 Convergence theorems . . . 383
18.2.5 Relation to the Riemann integral . . . 385
18.3 Space of integrable functions . . . 385
18.3.1 Distance . . . 385
18.3.2 Hilbert space L^2 . . . 386
18.3.3 L^p-spaces . . . 386
18.4 Product measures . . . 388
18.4.1 Construction . . . 388
18.4.2 Fubini's theorem . . . 389
18.5 Radon–Nikodym theorem . . . 389
18.6 Lebesgue–Stieltjes integral . . . 391
18.7 Signed measures . . . 393
18.8 Bochner integral . . . 395

19 Distributions **396**
19.1 Distributions . . . 397
19.1.1 Derivatives . . . 399
19.1.2 Support . . . 399
19.1.3 Examples . . . 400
19.1.4 Growth rates . . . 401
19.2 Convolutions and kernels . . . 402
19.3 Transformations . . . 403
19.3.1 Fourier series . . . 403
19.3.2 Fourier transform . . . 404
19.3.3 Wave front sets . . . 406
19.3.4 Laplace transform . . . 407
19.3.5 Integral representations . . . 407
19.4 Harmonic analysis ♣ . . . 408

20 Ordinary differential equations **410**
20.1 Boundary conditions . . . 410
20.2 Existence and uniqueness . . . 411
20.3 First-order ODEs . . . 411
20.4 Second-order ODEs . . . 412
20.4.1 General solution . . . 412

20.4.2 Constant coefficients 412
20.4.3 Frobenius' method 413
20.5 Fourier transform 414
20.6 Sturm–Liouville theory 415
20.7 Applications 416
20.7.1 Bessel functions 416

21 Partial differential equations 418
21.1 Important examples 418
21.2 General linear equations 419
21.3 First-order PDE 420
21.4 Second-order PDE 421
21.5 General methods 423
21.5.1 Separation of variables 423
21.6 Boundary conditions 423
21.6.1 Dirichlet problem 426
21.7 General 426
21.7.1 Symbols 426
21.8 Sobolev spaces 427

Chapter 16

Calculus

Chapter Contents

16.1 General definitions **351**
16.2 Continuity **352**
16.3 Convergence **353**
16.4 Series **353**
16.4.1 Convergence tests 353
16.5 Differentiation **355**
16.6 Integration theory **357**
16.6.1 Riemann integral 357
16.6.2 Euler integrals 359
16.6.3 Gaussian integrals 360
16.6.4 Generalizations 361
16.7 Convexity **362**
16.8 Trigonometry **364**

16.1 General definitions

Definition 16.1.1 (Domain). A connected, open subset of $\mathbb{R}^n$. (Not to be confused with the domain of a function as in Definition 1.5.1.)

Definition 16.1.2 (Factorial).

$$n! := n(n-1)\cdots 1\,, \tag{16.1}$$

where $n \in \mathbb{N}$. The convention is that $0! = 1$. (This, for example, agrees with the combinatorial result that there is a unique way to order zero objects.)

Definition 16.1.3 (Envelope). Consider a set $\mathcal{F}$ of real-valued functions with common domain X. An envelope (function) for $\mathcal{F}$ is any function $F : X \to \mathbb{R}$ such that

$$\forall f \in \mathcal{F}, x \in X : |f(x)| \leq F(x) . \tag{16.2}$$

16.2 Continuity

Definition 16.2.1 (Darboux function). A function that has the intermediate value property (Theorem 6.4.3).

Theorem 16.2.2 (Darboux). *Every differentiable function defined on a closed interval is Darboux.*

Corollary 16.2.3 (Bolzano). If $f(a) < 0$ and $f(b) > 0$ (or vice versa), there exists at least one point x_0 for which $f(x_0) = 0$.

Theorem 16.2.4 (Weierstrass's extreme value theorem). *Let $I = [a, b]$ be a closed interval and let $f : I \to \mathbb{R}$ be a continuous function. Then f attains a minimum and maximum at least once on I.*

Definition 16.2.5 (Absolute continuity). A function $f : \mathbb{R} \to \mathbb{R}$ is said to be absolutely continuous if for every $\varepsilon > 0$ there exists a $\delta_\varepsilon > 0$ such that for every finite collection of disjoint intervals $]x_i, y_i[$ satisfying

$$\sum_i (y_i - x_i) < \delta_\varepsilon , \tag{16.3}$$

the function f satisfies

$$\sum_i |f(y_i) - f(x_i)| < \varepsilon . \tag{16.4}$$

Property 16.2.6. The different types of continuity form the following hierarchy:

Lipschitz-continuous $\subset$ absolutely continuous $\subset$ uniformly continuous $\subset$ continuous .

Definition 16.2.7 (Function of bounded variation). A function f is said to be of bounded variation on the interval $[a, b]$ if the following quantity is finite:

$$V_{a,b}(f) := \sup_{P \in \mathcal{P}} \sum_{i=0}^{|P|-1} |f(x_{i+1}) - f(x_i)| , \tag{16.5}$$

where the supremum is taken over all partitions of $[a, b]$.

Property 16.2.8. Every function of bounded variation can be decomposed as the difference of two monotonically increasing functions.

Example 16.2.9. Every absolutely continuous function is of bounded variation.

16.3 Convergence

Definition 16.3.1 (Pointwise convergence). Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of functions. The sequence is said to converge pointwise to a limit function f if

$$\forall x \in \mathsf{dom}(f_n) : \lim_{n\to\infty} f_n(x) = f(x)\,. \tag{16.6}$$

Definition 16.3.2 (Uniform convergence). Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of functions. The sequence is said to converge uniformly to a limit function f if

$$\lim_{n\to\infty} \sup_{x\in\mathsf{dom}(f_n)} |f_n(x) - f(x)| = 0\,. \tag{16.7}$$

Definition 16.3.3 (Limit superior). Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of real numbers. The limit superior is defined as follows:

$$\limsup_{n\to\infty} x_n := \inf_{n\geq 1} \sup_{k\geq n} x_k\,. \tag{16.8}$$

Definition 16.3.4 (Limit inferior). Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of real numbers. The limit superior is defined as follows:

$$\liminf_{n\to\infty} x_n := \sup_{n\geq 1} \inf_{k\geq n} x_k\,. \tag{16.9}$$

Property 16.3.5. A sequence $(x_n)_{n\in\mathbb{N}}$ converges pointwise if and only if

$$\limsup_{n\to\infty} x_n = \liminf_{n\to\infty} x_n\,. \tag{16.10}$$

16.4 Series

16.4.1 Convergence tests

Property 16.4.1. A necessary condition for the convergence of a series $\sum_{i=1}^{+\infty} a_i$ is that

$$\lim_{n\to\infty} a_n = 0\,. \tag{16.11}$$

Property 16.4.2 (Absolute/conditional convergence). If $S' = \sum_{i=1}^{+\infty} |a_i|$ converges, so does $S = \sum_{i=1}^{+\infty} a_i$. In this case, S is said to be absolutely convergent. If S converges but S' does not, S is said to be conditionally convergent.

Definition 16.4.3 (Majorizing series). Let $S_a = \sum_{i=1}^{+\infty} a_i$ and $S_b = \sum_{i=1}^{+\infty} b_i$ be two series. The series S_a is said to majorize S_b if for every $k > 0$ the partial sums satisfy $S_{a,k} \geq S_{b,k}$, i.e.

$$\sum_{i=1}^{k} a_i \geq \sum_{i=1}^{k} b_i \tag{16.12}$$

for all $k \in \mathbb{N}$.

Method 16.4.4 (Comparison test). Let S_a, S_b be two series such that S_a majorizes S_b.

- If S_b diverges, then S_a diverges.
- If S_a converges, then S_b converges.
- If S_b converges, nothing can be said about S_a.
- If S_a diverges, nothing can be said about S_b.

Method 16.4.5 (Maclaurin–Cauchy integral test). Let f be a nonnegative, continuous and monotonically decreasing function defined on the interval $[n, +\infty[$ for some $n \in \mathbb{N}$. If $\int_n^{+\infty} f(x)\,dx$ is convergent, so is $\sum_{k=n}^{+\infty} f(k)$. On the other hand, if the integral is divergent, so is the series.

Remark 16.4.6. The function does not have to be nonnegative and decreasing on the complete interval. As long as it does on the interval $[N, +\infty[$ for some $N \geq n$, the statement holds. This can be seen by writing $\sum_{k=n}^{+\infty} f(k) = \sum_{k=n}^{N} f(k) + \sum_{k=N}^{+\infty} f(k)$ and noting that the first term is always finite (and similarly for the integral).

Property 16.4.7. If the integral in the previous theorem converges, the series is bounded in the following way:

$$\int_n^{+\infty} f(x)\,dx \leq \sum_{i=n}^{+\infty} a_i \leq f(n) + \int_n^{+\infty} f(x)\,dx\,. \tag{16.13}$$

Method 16.4.8 (d'Alembert's ratio test). Consider the quantity

$$R := \lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right|. \tag{16.14}$$

The following cases can be distinguished:

- $R < 1$: the series converges absolutely.
- $R > 1$: the series does not converge.
- $R = 1$: the test is inconclusive.

Method 16.4.9 (Cauchy's root test). Consider the quantity

$$R := \limsup_{n\to\infty} \sqrt[n]{|a_n|}\,. \tag{16.15}$$

The following cases can be distinguished:

- $R < 1$: the series converges absolutely.
- $R > 1$: the series does not converge.
- $R = 1$ and the limit approaches strictly from above: the series diverges.
- $R = 1$: the test is inconclusive.

Definition 16.4.10 (Radius of convergences). The number $1/R$ is called the radius of convergence.

Remark 16.4.11. The root test is stronger than the ratio test. However, if the ratio test can determine the convergence of a series, the radius of convergence of both tests will coincide and, hence, it is a well-defined quantity.

Method 16.4.12 (Gauss's test). If $a_n > 0$ for all $n \in \mathbb{N}$, one can write the ratio of successive terms as follows:

$$\left|\frac{a_n}{a_{n+1}}\right| = 1 + \frac{h}{n} + \frac{B(n)}{n^k}, \tag{16.16}$$

where $k > 1$ and $B(n)$ is a bounded function when $n \longrightarrow \infty$. The series converges if $h > 1$ and diverges otherwise.

Definition 16.4.13 (Asymptotic expansion). Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. A series $\sum_{i=0}^{+\infty} a_n x^n$ is called an asymptotic expansion of f if there exists an $N \in \mathbb{N}$ such that

$$f(x) - \sum_{i=0}^{n} a_i x^i = O(x^{n+1}) \tag{16.17}$$

for all $x \in \mathbb{R}$ and $n \geq N$.

16.5 Differentiation

Formula 16.5.1 (Derivative). Consider a function $f : \mathbb{R} \to \mathbb{R}$. If it exists, the following limit is called the derivative of f at $x \in \mathbb{R}$:

$$\frac{\mathrm{d}f}{\mathrm{d}x} \equiv f'(x) := \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}. \tag{16.18}$$

If the derivative exists at every point of some interval I, then f is said to be differentiable on I. For multivariate functions $f : \mathbb{R}^n \to \mathbb{R}$, one can similarly define the partial derivatives:

$$\frac{\partial f}{\partial x_i} := \frac{f(x + he_i) - f(x)}{h}, \tag{16.19}$$

where e_i is the ith coordinate vector, i.e. the partial derivatives determine the rate of change in the coordinate directions.

Notation 16.5.2. Iterated derivatives are often denoted as follows:

$$f^{(i)}(x) := \frac{\mathrm{d}^i f}{\mathrm{d}x^i}. \tag{16.20}$$

Theorem 16.5.3 (Mean value theorem). *Let f be a continuous function defined on the closed interval $[a, b]$ and differentiable on the open interval $]a, b[$. There exists a point $c \in]a, b[$ such that*

$$f'(c) = \frac{f(b) - f(a)}{b - a}. \tag{16.21}$$

Definition 16.5.4 (Differentiablity class). A function $f : \mathbb{R} \to \mathbb{R}$ is said to be of class C^n if it is $n \in \mathbb{N}$ times continuously differentiable, i.e. $f^{(i)}$ exists and is continuous for $i = 1, \dots, n$. Multivariate functions are said to be of class C^n if all of their partial derivatives are of class C^{n-1} or, by recursion, if all mixed partial derivatives up to order n exists and are continuous.

Definition 16.5.5 (Smooth function). A function f is said to be smooth if it is of class C^∞.

Theorem 16.5.6 (Boman). *Consider a function $f : \mathbb{R}^d \to \mathbb{R}$. If, for every smooth function $g : \mathbb{R} \to \mathbb{R}^d$, the composition $f \circ g$ is smooth, the function f is also smooth.*

Property 16.5.7 (Taylor expansion). Let $f : \mathbb{R} \to \mathbb{R}$ be a smooth function. Around every point $x \in \mathbb{R}$, one can express f as the following series:

$$f(y) = f(x) + f'(x)(y - x) + \frac{f''(x)}{2}(y - x)^2 + \cdots = \sum_{n=0}^{+\infty} \frac{f^{(n)}(x)}{n!}(y - x)^n \,. \tag{16.22}$$

For the special case $x = 0$, the name **Maclaurin series** is sometimes used. A similar expression exists for multivariate functions, where derivatives are replaced by partial derivatives.

Definition 16.5.8 (Analytic function). A function f is said to be analytic if it is smooth and if its Taylor series expansion around any point x converges to f in some neighbourhood of x. The set of analytic functions defined on V is denoted by $C^\omega(V)$.

Theorem 16.5.9 (Hadamard lemma). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth function defined on an open, star-convex set U. One can expand the function as follows:*

$$f(x) = f(x_0) + \sum_{i=1}^{n} (x^i - x_0^i) g_i(x_0) \,, \tag{16.23}$$

where all functions g_i are also smooth on U.

From this expression, one can also see that the functions g_i, evaluated at 0, give the partial derivatives of f. These functions are sometimes called the **Hadamard quotients**.

Remark 16.5.10. This lemma gives a finite-order approximation of the Taylor expansion of f.

Theorem 16.5.11 (Schwarz[1]). *Consider a function $f \in C^2(\mathbb{R}^n, \mathbb{R})$. The mixed partial derivatives of f coincide for all indices $i, j \leq n$:*

$$\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right) = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right). \tag{16.24}$$

Formula 16.5.12 (Derivative of $f(x)^{g(x)}$). Consider a function of the form

$$u(x) = f(x)^{g(x)} \,,$$

[1] Also called **Clairaut's theorem**.

with $f, g : \mathbb{R} \to \mathbb{R}$ differentiable. After taking the logarithm and applying the standard rules of differentiation, one can obtain the following expression:

$$\left(f(x)^{g(x)}\right)' = f(x)^{g(x)}\left(g'(x)\ln\left[f(x)\right] + \frac{g(x)}{f(x)}f'(x)\right). \tag{16.25}$$

Definition 16.5.13 (Euler operator). On the space $C^{n>1}(\mathbb{R}^n, \mathbb{R})$, the Euler operator $\mathbb{E}$ is defined as follows:

$$\mathbb{E} := \sum_{i=1}^{n} x_i \frac{\partial}{\partial x^i}. \tag{16.26}$$

Theorem 16.5.14 (Euler). *Let f be a homogeneous function, i.e.*

$$f(\lambda x_1, \dots, \lambda x_n) = \lambda^n f(x_1, \dots, x_n). \tag{16.27}$$

This function satisfies the following equality:

$$\mathbb{E}(f) = n f(x_1, \dots, x_n). \tag{16.28}$$

16.6 Integration theory

16.6.1 Riemann integral

Definition 16.6.1 (Improper Riemann integral).

$$\int_{-\infty}^{+\infty} f(x)\,dx := \lim_{\substack{a\to-\infty \\ b\to+\infty}} \int_a^b f(x)\,dx \tag{16.29}$$

One-sided improper integrals are defined in a similar fashion.

Theorem 16.6.2 (First fundamental theorem of calculus). *Let f be a continuous function defined on an open interval I and consider any number $c \in I$. Then the integral*

$$F(x) = \int_c^x f(x')\,dx' \tag{16.30}$$

is differentiable (and uniformly continuous) and gives an antiderivative of f:

$$\forall x \in I : F'(x) = f(x). \tag{16.31}$$

Remark 16.6.3. The function F in the previous theorem is called a **primitive (function)** of f. Remark that F is just 'a' primitive function, since adding a constant to F does not change anything because the derivative of a constant is zero (the number $c \in \mathbb{R}$ was arbitrary).

Theorem 16.6.4 (Second fundamental theorem of calculus[2]). *Consider an integrable function $f : [a,b] \to \mathbb{R}$. If $F : [a,b] \to \mathbb{R}$ is an antiderivative of f, then*

$$\int_a^b f(x)\,dx = F(b) - F(a)\,. \tag{16.32}$$

Formula 16.6.5 (Differentiation under the integral sign[3]).

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_{a(x)}^{b(x)} f(x,y)\,dy = f\big(x,b(x)\big)b'(x) - f\big(x,a(x)\big)a'(x) + \int_{a(x)}^{b(x)} \frac{\partial f(x,y)}{\partial x}\,dy \tag{16.33}$$

Definition 16.6.6 (Borel transform). Consider the following function:

$$F(x) := \sum_{n=0}^{+\infty} \frac{a_n}{n!} x^n\,. \tag{16.34}$$

If

$$\int_0^{+\infty} e^{-t} F(xt)\,dt < +\infty \tag{16.35}$$

for all $x \in \mathbb{R}$, then F is called the Borel transform of

$$f(x) = \sum_{n=0}^{+\infty} a_n x^n\,. \tag{16.36}$$

Furthermore, the integral gives a convergent expression for f.

Proof. The function F is defined as follows:

$$F(x) := \sum_{n=0}^{+\infty} \frac{a_n}{n!} x^n\,.$$

The Borel transform gives:

$$\begin{aligned}
\int_0^{+\infty} F(xt)e^{-t}\,dt &= \sum_{n=0}^{+\infty} \int_0^{+\infty} \frac{a_n}{n!} x^n t^n e^{-t}\,dt \\
&= \sum_{n=0}^{+\infty} \frac{a_n}{n!} x^n \int_0^{+\infty} t^n e^{-t}\,dt \\
&= \sum_{n=0}^{+\infty} \frac{a_n}{n!} x^n\,\Gamma(n+1) \\
&= \sum_{n=0}^{+\infty} a_n x^n\,,
\end{aligned}$$

where the definition of the Gamma function (Formula 16.6.9) was used on line 3,

[2]Sometimes called the **Newton–Leibniz** theorem.
[3]This is a more general version of the *Leibniz integral rule*.

and the relation (16.6.12) between the factorial function and the Gamma function was used on line 4. □

Theorem 16.6.7 (Watson). *The Borel transform F is unique if the function f is holomorphic (see Definition 17.2.1) on the domain* $\{z \in \mathbb{C} \mid |\arg(z)| < \frac{\pi}{2} + \varepsilon\}$.

16.6.2 Euler integrals

Formula 16.6.8 (Beta function). The beta function (also known as the **Euler integral of the first kind**) is defined as follows:

$$B(x, y) := \int_0^1 t^{x-1}(1-t)^{y-1}\, dt\,. \tag{16.37}$$

Formula 16.6.9 (Gamma function). The gamma function (also known as the **Euler integral of the second kind**) is defined as follows:

$$\Gamma(x) := \int_0^{+\infty} t^{x-1}e^{-t}\, dt\,. \tag{16.38}$$

Formula 16.6.10. The following formula relates the beta and gamma functions:

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}\,. \tag{16.39}$$

Property 16.6.11 (Recursion). The gamma function satisfies the following recursion relation for all points x in its domain:

$$\Gamma(x+1) = z\Gamma(x)\,. \tag{16.40}$$

Formula 16.6.12 (Factorial). For integers $n \in \mathbb{N}$, the gamma function can be expressed in terms of the factorial (Definition 16.1.2):

$$\Gamma(n) = (n-1)!\,. \tag{16.41}$$

Formula 16.6.13 (Stirling). This formula (originally stated for the factorial of natural numbers) gives an asymptotic expansion of the gamma function:

$$\ln\Gamma(z) \approx z\ln z - z + \frac{1}{2}\ln\left(\frac{2\pi}{z}\right)\,. \tag{16.42}$$

Property 16.6.14 (Euler's reflection formula). When $x \notin \mathbb{Z}$, the following formula holds:

$$\Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin(\pi x)}\,. \tag{16.43}$$

16.6.3 Gaussian integrals

Formula 16.6.15 (n-dimensional Gaussian integral). An integral of the form

$$I\left(A,\vec{\mathbf{b}}\right) := \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\vec{\mathbf{x}}\cdot A\vec{\mathbf{x}} + \vec{\mathbf{b}}\cdot\vec{\mathbf{x}}\right) d^n x\,, \tag{16.44}$$

where A is a real symmetric matrix. By performing the transformation $\vec{\mathbf{x}} \to A^{-1}\vec{\mathbf{b}} - \vec{\mathbf{x}}$ and diagonalizing A, one can obtain the following expression:

$$I\left(A,\vec{\mathbf{b}}\right) = \sqrt{\frac{(2\pi)^n}{\det(A)}}\exp\left(\frac{1}{2}\vec{\mathbf{b}}\cdot A^{-1}\vec{\mathbf{b}}\right). \tag{16.45}$$

More generally, one has the following result:

$$\int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\vec{\mathbf{x}}\cdot A\vec{\mathbf{x}}\right) f(\vec{\mathbf{x}})\, d^n x = \sqrt{\frac{(2\pi)^n}{\det(A)}}\exp\left(\frac{1}{2}\sum_{i,j=1}^{n} A^{-1}_{ij}\partial_i\partial_j\right) f(\vec{\mathbf{x}})\Big|_{\vec{\mathbf{x}}=0}. \tag{16.46}$$

This result is sometimes called **Wick's lemma**.

Corollary 16.6.16. A functional generalization is given by:

$$\begin{aligned} I(iA,iJ) &= \int \exp\left(-i\int_{\mathbb{R}^n\times\mathbb{R}^n} \varphi(x)A(x,y)\varphi(y)\, d^n x\, d^n y + i\int_{\mathbb{R}^n}\varphi(x)J(x)\, d^n x\right)[d\varphi] \\ &= C\det(A)^{-1/2}\exp\left(\frac{i}{2}\int_{\mathbb{R}^n\times\mathbb{R}^n} J(x)A^{-1}(x,y)J(y)\, d^n x\, d^n y\right), \end{aligned} \tag{16.47}$$

where the analytic continuation $I(iA,iJ)$ of Eq. (16.45) was used. One should pay attention to the normalization factor C which is infinite in general.

Method 16.6.17 (Feynman diagrams). The expansion of the exponential in the general expression for Gaussian integrals admits a diagrammatic expression. Let $f(\vec{\mathbf{x}})$ be a polynomial function of the coordinates.

If the number of factors in a monomial is odd, the resulting integral will vanish (since the integral of an odd function over an even domain is zero). For an even number of factors, one gets the following expression:

$$\int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\vec{\mathbf{x}}\cdot A\vec{\mathbf{x}}\right) x^{i_1}\cdots x^{i_k}\, d^n x = \sqrt{\frac{(2\pi)^n}{\det(A)}}\sum_{\sigma\in S_k} A^{-1}_{\sigma(i_1)\sigma(i_2)}\cdots A^{-1}_{\sigma(i_{k-1})\sigma(i_k)}\,. \tag{16.48}$$

To every coordinate dimension, one can assign a vertex in the plane, i.e. the object x_i can be interpreted as a real-valued function on the set of $k \in \mathbb{N}$ elements. The sum on the right-hand side above can then be expressed as a 'sum' over all possible diagrams, where a factor A^{-1}_{ij} is represented by a line connecting the vertices i and j.

Example 16.6.18 (Feynman diagrams). Some simple examples are given:

$A^{-1}_{13}A^{-1}_{12}A^{-1}_{24} =$

Higher powers of a given coordinate would then, for example, give rise to diagrams with loops at a given vertex:

$A_{11}^{-1}A_{12}^{-1}A_{22}^{-1} =$

Remark 16.6.19 (Normalization). In practice, one often divides all Gaussian integrals by the quantity $I(A, 0)$ to cancel the normalization factor. In the functional setting, this is even imperative since, as mentioned above, the normalization factor diverges for infinite-dimensional spaces.

16.6.4 Generalizations

Definition 16.6.20 (Henstock–Kurzweil integral[4]). Consider the usual definition of the (proper) Riemann integral, where tagged partitions P of $[a, b]$ are chosen and the integral is obtained as the limit of the Riemann sums

$$I = \sum_P f(x_i)(t_i - t_{i-1}) \tag{16.49}$$

as the mesh size of the partitions goes to zero.

Now, to obtain the generalized integral, consider a strictly positive function $\delta : [a, b]\mathbb{R}^{>0}$, the **gauge function**. Given such a gauge, a tagged partition P is said to be **δ-fine** if

$$[t_{i-1}, t_i] \subset [x_i - \delta(x_i), x_i + \delta(x_i)] \tag{16.50}$$

for subintervals in the partition.[5]

If the integral exists, it is given by the number $I \in \mathbb{R}$ such that for all $\varepsilon > 0$ there exists a gauge $\delta : [a, b] \to \mathbb{R}^{>0}$ such that, if P is δ-fine, then

$$\left| I - \sum_P f(x_i)(t_i - t_{i-1}) \right| < \varepsilon\,. \tag{16.51}$$

Remark 16.6.21 (Riemann integral). If the gauge functions are chosen to be constant, the classical (ε, δ)-definition of ordinary Riemann integrals is obtained.

The following statement can be seen as a refinement of Theorem 6.5.13. Moreover, it is also sometimes known as the **Borel–Lebesgue theorem**.

Property 16.6.22 (Cousin). For every gauge $\delta : [a, b] \to \mathbb{R}^{>0}$, there exists a δ-fine partition.

[4] Also called the **Perron**, **Lusin**, (**narrow**) **Denjoy** or **gauge** integral.

[5] If the condition $x_i \in [t_{i-1}, t_i]$ in the definition of tagged partitions is dropped, the **McShane integral** is obtained. This can be shown to be equivalent to the *Lebesgue integral* (see Chapter 18).

Property 16.6.23 (Integrability). If $f : [a, b] \to \mathbb{R}$ is bounded, then the following are equivalent:

- f is Henstock–Kurzweil integrable, and
- f is *Lebesgue integrable* (see Chapter 18).

More generally, a function $f : [a, b] \to \mathbb{R}$ is Henstock–Kurzweil integrable if and only if both f and $|f|$ are *Lebesgue integrable*.

The following property shows that 'improper' Henstock–Kurzweil integrals are only truly improper for unbounded domains.

Property 16.6.24 (Hake).

$$\int_a^b f\,dx = \lim_{c \nearrow b} \int_a^c f\,dx\,, \tag{16.52}$$

whenever either side exists.

One of the most important arguments for using the Henstock–Kurzweil integral is its refinement of the Second Fundamental Theorem of Calculus 16.6.4. Note that the theorem for the Riemann integral required that the derivative was integrable. The gauge integral relaxes this condition.

Theorem 16.6.25 (Second fundamental theorem of calculus). *Let $f : [a, b] \to \mathbb{R}$ be differentiable, then*

$$\int_a^x f'(x')\,dx' = f(x) - f(a) \;\; a.e. \tag{16.53}$$

16.7 Convexity

Definition 16.7.1 (Convex set). A subset of X of a vector space V (Definition 22.1.1) is said to be convex if $x, y \in X$ implies that $\{\lambda x + (1-\lambda)y \mid \lambda \in [0,1]\} \subset X$, i.e. if all straight lines connecting elements of the set are completely contained in that set. The **convex hull** of a subset X is defined as the smallest convex subset containing X.

Definition 16.7.2 (Extreme point). Consider a convex set X. The extreme points of X are the points $p \in X$ such that, if

$$p = \lambda p_1 + (1-\lambda)p_2 \tag{16.54}$$

for some $p_1, p_2 \in X$ and $\lambda \in [0,1]$, then $p_1 = p_2 = p$.

Definition 16.7.3 (Convex function). Let X be a convex set. A function $f : X \to \mathbb{R}$ is said to be convex if for all $x, y \in X$ and $\lambda \in [0,1]$:

$$f\big(\lambda x + (1-\lambda)y\big) \leq t\lambda(x) + (1-\lambda)f(y)\,. \tag{16.55}$$

For the definition of a **concave** function, the inequality has to be turned around.

Definition 16.7.4 (Linear map). A function $f : X \to \mathbb{R}$ is linear if and only if it is both convex and concave.

Theorem 16.7.5 (Karamata's inequality). *Consider an interval $I \subset \mathbb{R}$ and let $f : I \to \mathbb{R}$ be a convex function. If $(x_1, \dots, x_n)$ is a tuple that majorizes $(y_1, \dots, y_n)$, i.e.*

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \tag{16.56}$$

and

$$x_{(1)} + \cdots + x_{(k)} \geq y_{(1)} + \cdots + y_{(k)} \tag{16.57}$$

for all $k \leq n$, where $x_{(i)}$ denotes the ith largest element of $(x_1, \dots, x_n)$, then

$$\sum_{i=1}^{n} f(x_i) \geq \sum_{i=1}^{n} f(y_i)\,. \tag{16.58}$$

The following inequality can be derived directly from the definition of convexity by induction.

Theorem 16.7.6 (Jensen's inequality). *Let f be a convex function and consider a point $\{a_i\}_{i \leq n}$ in the probability simplex Δ^n (Example 7.2.3).*

$$f\left(\sum_{i=1}^{n} a_i x_i\right) \leq \sum_{i=1}^{n} a_i f(x_i)\,. \tag{16.59}$$

Definition 16.7.7 (Legendre transformation). Consider a function $f : \mathbb{R} \to \mathbb{R}$. In certain cases (especially in physics) it is sometimes useful to replace the argument x by the slope of f at x, i.e. to perform the transformation

$$x \longrightarrow f'(x)\,. \tag{16.60}$$

However, it should be clear that this transformation is not always well-defined and, even if it is, it does not always preserve all the information contained in f.

These conditions are satisfied exactly if f is convex (or concave). In this case, the Legendre transform of f is defined as

$$f^*(x^*) := \sup_x \big(x^* x - f(x)\big)\,. \tag{16.61}$$

Now, consider the case where f is differentiable. The above supremum can then be obtained by differentiating the right-hand side and equating it to zero. This results in $x^* = f'(x)$, which is exactly the transformation that was required. By expressing everything in terms of the Legendre tranformed quantity x^*, one can also find the derivative of f^*:

$$\frac{\mathrm{d}f^*}{\mathrm{d}x^*}(x^*) = x(x^*)\,. \tag{16.62}$$

Property 16.7.8 (Alternative characterization). In fact, up to an additive constant, the condition

$$(f^*)' = (f')^{-1} \tag{16.63}$$

uniquely determines the Legendre transformation.

Remark 16.7.9. These definitions can easily be extended to higher dimensions ($n \geq 2$).

16.8 Trigonometry

Definition 16.8.1 (Trigonometric functions). Consider Fig. 16.1. In a right(-angled) triangle, the trigonometric functions of the angle θ are defined as follows:

$$\sin(\theta) := \frac{y}{x} \qquad \cos\theta := \frac{z}{x} \qquad \tan(\theta) := \frac{y}{z} \tag{16.64}$$

Corollary 16.8.2.

$$\tan(\theta) = \frac{\sin(\theta)}{\cos(\theta)} \tag{16.65}$$

Figure 16.1: Right(-angled) triangle.

Chapter 17

Complex Analysis

Chapter Contents

17.1 Complex algebra **365**
17.2 Holomorphic functions **366**
17.3 Contour integrals **367**
17.4 Laurent series **368**
17.5 Singularities **369**
17.5.1 Poles 369
17.5.2 Branch cuts 370
17.5.3 Residue theorem 371
17.6 Limit theorems **372**

17.1 Complex algebra

The set of complex numbers $\mathbb{C}$ forms a 2-dimensional vector space over the field of real numbers (Chapter 22). At the same time, the operations of complex addition and complex multiplication also turn the complex numbers into a field.

Definition 17.1.1 (Complex conjugate). Complex conjugation

$$\overline{a + bi} := a - bi \tag{17.1}$$

is an involution (Definition 1.5.16). It is sometimes denoted by z^* instead of $\bar{z}$, but, unless this would cause confusion, the former notation will be used throughout this compendium.

Formula 17.1.2 (Real/imaginary part). A complex number can also be written as

$$\Re(z) + i\Im(z)\,, \tag{17.2}$$

where

$$\Re(z) := \frac{z+\bar{z}}{2}, \tag{17.3}$$

$$\Im(z) := \frac{z-\bar{z}}{2i}. \tag{17.4}$$

Note that the notations $\mathsf{Re}(z)$ and $\mathsf{Im}(z)$ are also frequently used.

Definition 17.1.3 (Argument and modulus). Consider a complex number expressed in *polar form*:

$$z = re^{i\theta} \tag{17.5}$$

The number $\theta \in [0, 2\pi[$ is called the argument of z and it is denoted by $\arg(z)$. The number $r \in \mathbb{R}^+$ is called the modulus of Z and is denoted by $|z|$.

@@ DERIVE POLAR FORM @@

Definition 17.1.4 (Riemann sphere). Consider the one-point compactification $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ (Definition 6.5.30). This set is called the Riemann sphere or **extended complex plane**. The standard operations on $\mathbb{C}$ can be generalized to $\overline{\mathbb{C}}$ for all nonzero $z \in \mathbb{C}$ in the following way:

$$\begin{aligned} z + \infty &:= \infty \\ z * \infty &:= \infty \\ \frac{z}{\infty} &:= 0. \end{aligned} \tag{17.6}$$

Since there exists no multiplicative inverse for ∞, the Riemann sphere is not a field.

17.2 Holomorphic functions

Definition 17.2.1 (Holomorphic function). A function f on an open set $U \subseteq \mathbb{C}$ that is complex differentiable at every point $z_0 \in U$, i.e. for every point $z_0 \in U$ the following limit exists:

$$f'(z_0) := \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}. \tag{17.7}$$

Definition 17.2.2 (Biholomorphic function). A holomorphic function admitting a holomorphic inverse.

Definition 17.2.3 (Entire). A function that is holomorphic on all of $\mathbb{C}$.

Property 17.2.4 (Cauchy–Riemann conditions). A holomorphic function f satisfies the following conditions:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad \text{and} \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}. \tag{17.8}$$

These conditions can be combined into one equation using the so-called **Wirtinger derivative**:

$$\frac{\partial f}{\partial \bar{z}} = 0\,. \tag{17.9}$$

Theorem 17.2.5 (Looman–Menchoff[1]). *Let f be a continuous function defined on a subset $U \in \mathbb{C}$. If the partial derivatives of the real and imaginary part exist and if f satisfies the Cauchy–Riemann conditions, then f is holomorphic on U.*

Property 17.2.6 (Laplace equation). The functions u, v satisfying the Cauchy–Riemann conditions are harmonic functions, i.e. they satisfy the *Laplace equation* (see Example 21.1.1).

Property 17.2.7 (Level sets). The functions u, v satisfying the Cauchy–Riemann conditions have orthogonal level curves (Definition 1.5.8).

Property 17.2.8 (Real functions). Consider a real-valued function f defined on the complex plane. If it is holomorphic, the Cauchy–Riemann conditions imply that f is a constant.

Theorem 17.2.9 (Identity theorem). *If two holomorphic functions on a domain D coincide on a set containing an accumulation point of D, they coincide on all of D.*

17.3 Contour integrals

Remark. Whenever contours are considered for integration purposes, they have been chosen to be evaluated counter-clockwise (by convention). To obtain results concerning clockwise evaluation, most of the time adding a minus sign is sufficient.

Definition 17.3.1 (Contour integral). The contour integral of a complex function

$$f(z) = u(z) + iv(z) \tag{17.10}$$

is defined as the following line integral:

$$\int_{z_1}^{z_2} f(z)\,dz = \int_{(x_1,y_1)}^{(x_2,y_2)} \big(u(x,y) + iv(x,y)\big)(dx + idy)\,. \tag{17.11}$$

Theorem 17.3.2 (Cauchy's integral theorem[2]). *Let Ω be a simply connected subset of $\mathbb{C}$ and let $f : \mathbb{C} \to \mathbb{C}$ be a holomorphic function on Ω. The contour integral around every closed, rectifiable (i.e. of finite length) contour C in Ω vanishes:*

$$\oint_C f(z)\,dz = 0\,. \tag{17.12}$$

[1] This is the most general theorem on the holomorphy of continuous functions. It generalizes the original results by *Riemann* and *Cauchy–Goursat*.

[2] Also called the **Cauchy–Goursat theorem**.

Corollary 17.3.3 (Freedom of contour). The contour integral of a holomorphic function depends only on the limits of integration and not on the contour connecting them.

Formula 17.3.4 (Cauchy's integral formula). Let Ω be a connected subset of $\mathbb{C}$ and let C be a closed contour in Ω. If $f : \mathbb{C} \to \mathbb{C}$ is holomorphic on Ω, then, at every point z_0 inside C, one can express f as follows:

$$f(z_0) = \frac{1}{2\pi i} \oint_C \frac{f(z)}{z - z_0} \, dz \,. \tag{17.13}$$

Corollary 17.3.5 (Analytic function). Let Ω be a connected subset of $\mathbb{C}$ and let C be a closed contour in Ω. If $f : \mathbb{C} \to \mathbb{C}$ is holomorphic on Ω, then f is analytic (Definition 16.5.8) on Ω and

$$f^{(n)}(z_0) = \frac{1}{2\pi i} \oint_C f(z) \frac{n!}{(z - z_0)^{n+1}} \, dz \,. \tag{17.14}$$

Furthermore, the derivatives are also holomorphic on Ω.

Theorem 17.3.6 (Morera). *If $f : \mathbb{C} \to \mathbb{C}$ is continuous on a connected open set Ω and*

$$\oint_C f(z) \, dz = 0 \tag{17.15}$$

for every closed contour C in Ω, then f is holomorphic on Ω.

Theorem 17.3.7 (Liouville). *Every bounded, entire function is constant.*

Theorem 17.3.8 (Sokhotski–Plemelj). *Let $f : \mathbb{R} \to \mathbb{C}$ be continuous and consider $a < 0 < b$, then*

$$\lim_{\varepsilon \to 0^+} \int_a^b \frac{f(x)}{x \pm i\varepsilon} \, dx = \mp i\pi f(0) + \mathcal{P} \int_a^b \frac{f(x)}{x} \, dx \,, \tag{17.16}$$

where $\mathcal{P}$ denotes the Cauchy principal value.

@@ ADD (Cauchy principal value) @@

17.4 Laurent series

Definition 17.4.1 (Laurent series). If $f : \mathbb{C} \to \mathbb{C}$ is an analytic function defined on an **annulus**, i.e. a ring-shaped region, then f can be expanded as the following series:

$$f(z) = \sum_{n=-N}^{+\infty} a_n (z - z_0)^n \qquad \text{with} \qquad a_n = \frac{1}{2\pi i} \oint \frac{f(z)}{(z - z_0)^{n+1}} \, dz \,, \tag{17.17}$$

The subseries containing all terms of negative degree is called the **principal part** of the Laurent series.

Notation 17.4.2. The ring of Laurent series in the indeterminate z is given by the ring $\mathbb{C}[[z, z^{-1}]]$.

Remark 17.4.3 (Multiplication). This definition can easily be generalized from $\mathbb{C}$ to general rings R. However, to be able to multiply two Laurent series, one has to restrict to series with a finite number of negative powers (unless R is topological such that convergence can be defined). This gives rise to the ring of **formal Laurent series** $R((z))$. This ring can be obtained as the fraction field of $R[[z]]$ (Definition 2.4.6).

Property 17.4.4 (Convergence). The Laurent series of an analytic function $f : \mathbb{C} \to \mathbb{C}$ converges uniformly to f on the annulus $R_1 < |z - z_0| < R_2$, with R_1 and R_2 the distances from z_0 to the two closest *poles* (see Definition 17.5.1 further below).

Definition 17.4.5 (Analytic continuation). Consider an analytic function $f : \mathbb{C} \to \mathbb{C}$ defined on an open subset $U \subset \mathbb{C}$. If $V \subset \mathbb{C}$ is an open subset containing U and if there exists an analytic function F on V such that $F(z) = f(z)$ for all $z \in U$, then F is called the analytic continuation of f to V. Using the identity theorem for holomorphic functions, one can prove that analytic continuations are unique (on connected domains).

Theorem 17.4.6 (Schwarz's reflection principle). *Let $f : \mathbb{C} \to \mathbb{C}$ be analytic on the upper half plane. If $z \in \mathbb{R} \implies f(z) \in \mathbb{R}$, then*

$$f(\overline{z}) = \overline{f(z)}. \tag{17.18}$$

17.5 Singularities

17.5.1 Poles

Definition 17.5.1 (Pole). A function $f : \mathbb{C} \to \mathbb{C}$ has a pole of order $m \in \mathbb{N}_0$ at a point $z_0 \in \mathbb{C}$ if its Laurent series at z_0 satisfies $\forall n < -m : a_n = 0$ and $a_{-m} \neq 0$.

Definition 17.5.2 (Removable singularity). Let $f : U \backslash \{a\} \to \mathbb{C}$ be a holomorphic function on a punctured open set. The point a is called a removable singularity of f if there exists a holomorphic function $g : U \to \mathbb{C}$ such that $f(z) = g(z)$ for all $z \in U \backslash \{a\}$.

Theorem 17.5.3 (Riemann). *A point $a \in \mathbb{C}$ is a removable singularity for a holomorphic function $f : U \backslash \{a\} \to \mathbb{C}$, with $U \subseteq \mathbb{C}$ an open set, if and only if there exists a neighbourhood of a on which f is bounded.*

Definition 17.5.4 (Meromorphic). A function $f : \mathbb{C} \to \mathbb{C}$ is said to be meromorphic if it is analytic on the whole complex plane with exception of isolated poles and removable singularities. Every meromorphic function can be written as a fraction of two holomorphic functions, where the poles coincide with the zeros of the denominator.

Definition 17.5.5 (Essential singularity). A function f has an essential singularity at a point z_0 if its Laurent series at z_0 satisfies $\forall n \in \mathbb{N} : a_{-n} \neq 0$, i.e. if its Laurent series has infinitely many negative degree terms.

Method 17.5.6 (Frobenius transformation). To study the behaviour of a function f at $z \longrightarrow \infty$, one can apply the Frobenius transformation $h = 1/z$ and study the limit $\lim_{h\to 0} f(h)$. For example, a singularity at ∞ is defined as a singularity of $f(1/z)$ at 0.

Property 17.5.7 (Polynomials). An entire function $f : \mathbb{C} \to \mathbb{C}$ is polynomial if and only if it has a pole at ∞.

Theorem 17.5.8 (Casorati–Weierstrass). *Let $f : \mathbb{C} \to \mathbb{C}$ be holomorphic on the punctured open set $U\backslash\{z_0\}$ with an essential singularity at z_0. For every neighbourhood V of z_0 contained in U, the image $f(V\backslash\{z_0\})$ is dense in $\mathbb{C}$.*

Corollary 17.5.9. If $f : \mathbb{C} \to \mathbb{C}$ is a nonpolynomial, entire function, then, for every $c \in \mathbb{C}$, there exists a sequence $z_n \longrightarrow \infty$ such that $f(z_n) \longrightarrow c$.[3]

Theorem 17.5.10 (Picard's little theorem). *The range of a nonconstant, entire function is the complex plane with at most a single exception.*

Theorem 17.5.11 (Picard's great theorem). *Let $f : \mathbb{C} \to \mathbb{C}$ be an analytic function with an essential singularity at z_0. On every punctured neighbourhood of z_0, f takes on all possible values, with at most a single exception, infinitely many times.*

17.5.2 Branch cuts

Formula 17.5.12 (Roots). Let $z \in \mathbb{C}$. The nth roots[4] of $z = re^{i\theta}$ are given by

$$\left\{ \sqrt[n]{r} \exp\left(\frac{\theta + 2\pi k}{n} i \right) \middle| k \in \{0, 1, \dots, n\} \right\}. \tag{17.19}$$

Formula 17.5.13 (Complex logarithm). The natural logarithm can be continued to the complex plane (as a multivalued function) as follows:

$$\mathsf{LN}(z) := \left\{ \ln(r) + i(\theta + 2\pi k) \mid k \in \mathbb{Z} \right\}. \tag{17.20}$$

Definition 17.5.14 (Branch). The problem with the previous two formulas is that they represent multivalued functions. To get an unambiguous image it is necessary to fix a value of the parameter k. By doing so there will arise curves, called **branch cuts**, in the complex plane where the function becomes discontinuous. A **branch** is defined as a particular choice of the parameter k.

For the logarithm, the choice for $\arg(\mathsf{LN}) \in]\alpha, \alpha + 2\pi]$ is often denoted by LN_α or $\log_\alpha$.

Definition 17.5.15 (Principal value). The principal value[5] of a multivalued complex function is defined as the value associated with a choice of branch for which $\arg(f) \in]-\pi, \pi]$.

[3] Polynomials are excluded due to the property above.
[4] Also see the fundamental theorem of algebra (Theorem 2.5.7).
[5] Not to be confused with the Cauchy principal value.

Definition 17.5.16 (Branch point). Let f be a complex-valued function. A point z_0 for which there exists no neighbourhood on which f is single-valued is called a branch point.

Definition 17.5.17 (Branch cut). A line connecting exactly two branch points, one possibly being ∞, is called a branch cut. In case there exist multiple branch cuts, they are required to never cross.

Example 17.5.18. Consider the complex function

$$f(z) = \frac{1}{\sqrt{(z - z_1) \cdots (z - z_n)}} \,. \tag{17.21}$$

This function has singularities at $z_1, \dots, z_n$. If $n \in \mathbb{N}$ is even, this function will have n (finite) branch points. This implies that the points can be grouped in pairs connected by non-intersecting branch cuts. If n is odd, this function will have n (finite) branch points and one branch point at infinity. The finite branch points will be grouped in pairs connected by non-intersecting branch cuts and the remaining branch point will be joined to infinity by a branch cut that does not intersect the others.

17.5.3 Residue theorem

Definition 17.5.19 (Residue). By applying Definition 17.3.1 to a polynomial function, one finds

$$\oint_C (z - z_0)^n \, dz = 2\pi i \delta_{n,-1} \,, \tag{17.22}$$

where C is a contour around the pole $z = z_0$. This means that integrating a Laurent series around a pole isolates the coefficient a_{-1}. This coefficient is, therefore, called the residue of the function at the given pole.

Notation 17.5.20. The residue of a complex function $f : \mathbb{C} \to \mathbb{C}$ at a pole z_0 is denoted by

$$\mathrm{Res}[f(z)]_{z=z_0} \,. \tag{17.23}$$

Formula 17.5.21. For a pole of order $m \in \mathbb{N}_0$, the residue is calculated as follows:

$$\mathrm{Res}\big[f(z)\big]_{z=z_j} = \lim_{z \to z_0} \frac{1}{(m-1)!} \left(\frac{\partial}{\partial z} \right)^{m-1} \big(f(z)(z - z_0) \big) \,. \tag{17.24}$$

For essential singularities, the residue can be found by writing out the Laurent series explicitly.

Theorem 17.5.22 (Residue theorem). *If $f : \mathbb{C} \to \mathbb{C}$ is meromorphic on $\Omega \subseteq \mathbb{C}$ and if C is a closed contour in Ω that contains the poles z_j of f, then*

$$\oint_C f(z) \, dz = 2\pi i \sum_j \mathrm{Res}\big[f(z)\big]_{z=z_j} \,. \tag{17.25}$$

For poles on the contour C, only half of the residue contributes to the integral.

Formula 17.5.23 (Argument principle). Let $f : \mathbb{C} \to \mathbb{C}$ be meromorphic and denote the number of zeros and poles of f inside the contour C by $Z_f(C)$ and $P_f(C)$, respectively. From the residue theorem, one can derive the following formula:

$$\frac{1}{2\pi i}\oint_C \frac{f(z)}{f'(z)}\,dz = Z_f(C) - P_f(C)\,. \tag{17.26}$$

Definition 17.5.24 (Winding number). Let $f : \mathbb{C} \to \mathbb{C}$ be meromorphic and let C be a simple closed contour. For all $a \notin f(C)$, the winding number, also called the **index**, of a with respect to the function f is defined as follows:

$$\mathsf{Ind}_f(a) := \frac{1}{2\pi i}\oint_C \frac{f'(z)}{f(z)-a}\,dz\,. \tag{17.27}$$

This number is always an integer.

17.6 Limit theorems

Theorem 17.6.1 (Small limit theorem). *Let $f : \mathbb{C} \to \mathbb{C}$ be a function that is holomorphic almost everywhere and let the contour C be a circular segment with radius ε and central angle α. If z is parametrized as $z = \varepsilon e^{i\theta}$, then*

$$\oint_C f(z)\,dz = i\alpha A \tag{17.28}$$

with

$$A = \lim_{\varepsilon\to 0} f(z)\ . \tag{17.29}$$

Theorem 17.6.2 (Great limit theorem). *Let $f : \mathbb{C} \to \mathbb{C}$ be a function that is holomorphic almost everywhere and let the contour C be a circular segment with radius R and central angle α. If z is parametrized as $z = Re^{i\theta}$, then*

$$\oint_C f(z)\,dz = i\alpha B \tag{17.30}$$

with

$$B = \lim_{R\to\infty} f(z)\,. \tag{17.31}$$

Theorem 17.6.3 (Jordan's lemma). *Let $g : \mathbb{C} \to \mathbb{C}$ be a continuous function that can be written as $g(z) = f(z)e^{bz}$ and let the contour C be a semicircle lying in the half-plane bounded by the real axis and oriented away of the point $i\bar{b}$. If z is parametrized as $z = Re^{i\theta}$ and*

$$\lim_{R\to\infty} f(z) = 0\,, \tag{17.32}$$

then

$$\oint_C g(z)\,dz = 0\,. \tag{17.33}$$

Chapter 18

Measure & Integration Theory

The main references for this chapter are Capinski and Kopp (2013); Choquet-Bruhat, DeWitt-Morette, and Dillard-Bleick (1991).

Chapter Contents

18.1 Measure theory **374**

18.1.1 General definitions 374

18.1.2 Lebesgue measure 376

18.1.3 Measurable functions 377

18.1.4 Limit operations 379

18.1.5 Polish spaces 381

18.2 Lebesgue integral **381**

18.2.1 Simple functions 381

18.2.2 Measurable functions 382

18.2.3 Integrable functions 382

18.2.4 Convergence theorems 383

18.2.5 Relation to the Riemann integral 385

18.3 Space of integrable functions **385**

18.3.1 Distance 385

18.3.2 Hilbert space L^2 386

18.3.3 L^p-spaces 386

18.4 Product measures **388**

18.4.1 Construction 388

18.4.2 Fubini's theorem 389

18.5 Radon–Nikodym theorem **389**

18.6 Lebesgue–Stieltjes integral 391
18.7 Signed measures 393
18.8 Bochner integral 395

18.1 Measure theory

18.1.1 General definitions

Definition 18.1.1 (Measure). Recall Definition 1.6.12 and consider a measurable space (X, Σ). A function $\mu : \Sigma \to \overline{\mathbb{R}}$ is called a measure if it satisfies the following conditions:

1. **Nonnegativity**: $\forall E \in \Sigma : \mu(E) \geq 0$,
2. **Empty set is null**: $\mu(\emptyset) = 0$, and
3. **σ-additivity**: $\forall i \neq j : E_i \cap E_j = \emptyset \implies \mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{i=n}^{+\infty} \mu(E_n)$.

When μ is defined on all of $P(X)$ and only satisfies countable subadditivity, i.e. the equality in the last condition becomes an inequality $\leq$, it is called an **outer measure**.

Remark 18.1.2. To show that two measures coincide on a σ-algebra, it suffices to show that they coincide on the generating sets and apply the Monotone Class Theorem 1.6.19.

Definition 18.1.3 (Measure space). A triple (X, Σ, μ), where the pair (X, Σ) is a measurable space and μ is a measure thereon, is called a measure space.

Definition 18.1.4 (Null set). A set $A \subset \mathbb{R}$ is said to be null if $\mu(A) = 0$.

Definition 18.1.5 (Almost everywhere[1]). Let (X, Σ, μ) be a measure space. A property P is said to hold on X almost everywhere (abbreviated as **a.e.**) if it satisfies the following equation:

$$\mu\big(\{x \in X \mid \neg P(x)\}\big) = 0\,, \tag{18.1}$$

i.e. it holds everywhere except for a null set.

Definition 18.1.6 (Complete measure space). A measure space (X, Σ, μ) is said to be complete if for every $E \in \Sigma$ with $\mu(E) = 0$ the implication $A \subset E \implies A \in \Sigma$ holds. Additivity then necessarily implies that $\mu(A) = 0$.

Definition 18.1.7 (Completion). Let $\Sigma \subseteq \overline{\Sigma}$ be σ-algebras on a set X. $(X, \overline{\Sigma}, \overline{\mu})$ is called the completion of (X, Σ, μ) if:

1. $\forall A \in \Sigma : \overline{\mu}(A) = \mu(A)$,
2. $(X, \overline{\Sigma}, \overline{\mu})$ is complete, and

[1] In probability theory this is often called **almost surely**.

3. $\overline{\Sigma}$ is the smallest σ-algebra with these properties.

Definition 18.1.8 (σ-finite measure). Let (X, Σ, μ) be a measure space. The measure μ is said to be σ-finite if there exists a sequence $(A_n)_{n\in\mathbb{N}}$ of measurable sets such that $\bigcup_{n=1}^{+\infty} A_n = X$ with $\forall n \in \mathbb{N} : \mu(A_n) < +\infty$.

Definition 18.1.9 (Borel measure). Consider a topological space together with its Borel σ-algebra (Definition 6.1.10). Any measure defined on this measurable space is called a Borel measure.

Definition 18.1.10 (Locally finite measure). A measure on a Hausdorff space whose measurable sets contain the Borel sets such that every point has an open neighbourhood of finite measure. On locally compact spaces, this is equivalent to requiring that every compact subset has finite measure.

Given a Borel measure on a topological space, there are different ways of how one can approximate the measure of Borel sets by those of open or closed sets. Sadly, different distinct definitions can be found in the literature all under the name of 'regular' measure. The next definition gives the most widely used ones.

Definition 18.1.11 (Regular measure). An outer measure μ on a topological space X whose completion contains the Borel sets is said to be **Borel regular** if for every set $A \subset X$, there exists a Borel set $B \subseteq A$ such that $\mu(A) = \mu(B)$.

Let μ be a measure on (X, Σ), where X is a topological space. It is said to be **regular** if if it satisfies the following conditions for every measurable set B:

1. **Outer regularity**: $\mu(B) = \inf\{\mu(O) \mid O \text{ open and measurable}, O \supset B\}$, and
2. **Inner regularity**: $\mu(B) = \sup\{\mu(K) \mid K \text{ closed and measurable}, K \subset B\}$.

If X is Hausdorff, the measure is said to be **regular** if it satisfies the following conditions for every measurable set B:

1. **Outer regularity**: $\mu(B) = \inf\{\mu(O) \mid O \text{ open and measurable}, O \supset B\}$, and
2. **Inner regularity** or **tightness**: $\mu(B) = \sup\{\mu(K) \mid K \text{ compact and measurable}, K \subset B\}$.

Note that since compact subsets of Hausdorff spaces are closed, tight measures are in particular regular.

By slightly modifying the definition of tight Borel measures, one can obtain yet another type of regularity.

Definition 18.1.12 (Radon measure). A Borel measure on a Hausdorff space that is outer regular on Borel sets, inner regular on open sets and locally finite. (For locally compact spaces, outer regularity is superfluous.)

Definition 18.1.13 (Radon space). A topological space on which every finite Borel measure is Radon. It is sometimes called **strongly Radon** if every locally finite Borel measure is Radon.

Example 18.1.14. Locally compact, separable metric space, Polish spaces and even Suslin spaces (Definition 18.1.60) are Radon. Slightly weaker, every finite Borel measure on a metric space is regular.

18.1.2 Lebesgue measure

Definition 18.1.15 (Lebesgue outer measure). Let $X \subseteq \mathbb{R}$ be a set. The (Lebesgue) outer measure of X is defined as follows:

$$\lambda^*(X) := \inf\left\{ \sum_{n=1}^{+\infty} l(I_n) \,\middle|\, (I_n)_{n\in\mathbb{N}} \text{ a sequence of open intervals that covers } X \right\}. \tag{18.2}$$

Property 18.1.16 (Intervals). The outer measure of an interval I equals its length: $\lambda^*(I) = l(I)$.

Property 18.1.17 (Translation-invariance). The outer measure is translation-invariant:

$$\lambda^*(A + t) = \lambda^*(A) \tag{18.3}$$

for all $A \subset \mathbb{R}$ and $t \in \mathbb{R}$.

Theorem 18.1.18 (Carathéodory's criterion). *Let X be a subset of $\mathbb{R}$. If X satisfies the following equation, it is said to be **Lebesgue measurable**:*

$$\forall A \subseteq \mathbb{R} : \lambda^*(A) = \lambda^*(A \cap X) + \lambda^*(A \cap X^c)\,. \tag{18.4}$$

*The collection of all Lebesgue-measurable sets is denoted by $\mathcal{M}$ and the outer measure $\lambda^*(X)$, now denoted by λ, is called the **Lebesgue measure** of X.*

Corollary 18.1.19. The Lebesgue σ-algebra $\mathcal{M}$ is the completion of the Borel σ-algebra $\mathcal{B}$. (This is how the Lebesgue σ-algebra was introduced historically.)

Property 18.1.20. Any countable set is null with respect to the Lebesgue outer measure.

Property 18.1.21 (Regularity). The Lebesgue measure is a regular Borel measure. For every $A \subseteq \mathbb{R}$, there exists a sequence $(O_n)_{n\in\mathbb{N}}$ of open sets such that

$$A \subset \bigcap_{n=1}^{+\infty} O_n \qquad \text{and} \qquad \lambda\left(\bigcap_{n=1}^{+\infty} O_n\right) = \lambda^*(A)\,, \tag{18.5}$$

and for every $E \in \mathcal{M}$, there exists a sequence $(F_n)_{n\in\mathbb{N}}$ of closed sets such that

$$\bigcup_{n=1}^{+\infty} F_n \subset E \qquad \text{and} \qquad \lambda\left(\bigcup_{n=1}^{+\infty} F_n\right) = \lambda(E)\,. \tag{18.6}$$

Property 18.1.22. Consider a set $A \subset \mathbb{R}$. $A \in \mathcal{M}$ if and only if for every $\varepsilon > 0$ there exist an open set $O \supset A$ and a closed set $F \subset A$ such that $\lambda^*(O \backslash A) < \varepsilon$ and $\lambda^*(A \backslash F) < \varepsilon$.

Property 18.1.23. Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of sets in $\mathcal{M}$. The following two properties apply:

$$\forall i \in \mathbb{N} : A_i \subseteq A_{i+1} \implies \lambda\left(\bigcup_{n=1}^{+\infty} A_n\right) = \lim_{n\to\infty} \lambda(A_n)\,, \tag{18.7}$$

$$\forall i \in \mathbb{N} : A_i \supseteq A_{i+1} \wedge \lambda(A_1) < +\infty \implies \lambda\left(\bigcap_{i=n}^{+\infty} A_n\right) = \lim_{n\to\infty} \lambda(A_n)\,. \tag{18.8}$$

Remark 18.1.24. This property is valid for every σ-additive set function.

Construction 18.1.25 (Restriction). Let $A \in \mathcal{M}$ have nonzero measure. The restriction of the Lebesgue measure to the set B is defined as follows:

$$\mathcal{M}_A := \{A \cap B \mid B \in \mathcal{M}\} \qquad \text{and} \qquad \forall E \in \mathcal{M}_A : \lambda_A(E) := \lambda(E)\,. \tag{18.9}$$

It can be shown that the measure space $(A, \mathcal{M}_A, \lambda_A)$ is complete.

The construction of Lebesgue measurable sets from the Lebesgue outer measure can be generalized to arbitrary sets and outer measures.

Construction 18.1.26 (Carathéodory's extension theorem[2]). Every outer measure μ^* gives rise to a σ-algebra consisting of those sets that satisfy Carathéodory's criterion (Theorem 18.1.18) with respect to μ^*. Furthermore, consider a **premeasure** μ_0, i.e. a σ-additive function defined on an algebra of sets (Definition 1.6.11) such that $\mu_0(\emptyset) = 0$. Definition 18.1.15 can be used to define an outer measure μ^* in terms of the premeasure μ_0 by replacing intervals with elements from the given algebra of sets. The σ-algebra generated by this outer measure contains the given algebra of sets and μ^* restricts to μ_0. This shows that any premeasure can be extended to a genuine measure, uniquely if μ_0 is σ-finite. Moreover, it can be shown that this measure is complete.

18.1.3 Measurable functions

Definition 18.1.27 (Measurable function). Consider two measurable spaces (X, Σ_X) and (Y, Σ_Y). A function $f : X \to Y$ is said to be measurable if for every measurable set $A \in \Sigma_Y$ the preimage $f^{-1}(A)$ is also measurable. Equivalently, the σ-algebra generated by the preimages of measurable sets in Σ_Y should be a sub-σ-algebra of Σ_X.

Property 18.1.28. Measurability is closed under composition. This turns the collection of measurable spaces and measurable functions into a category **Meas**. (This notation is also used for the subcategory on *measure spaces* (Definition 18.1.39).)

Definition 18.1.29 (Measurable). Sometimes an equivalence class of real- or complex-valued measurable functions up to functions that are zero a.e. is called a measurable.

Two important examples are given below:

[2] Also called the **Hahn–Kolmogorov theorem**.

Example 18.1.30 (Borel-measurable function). A continuous function $f : X \to Y$ such that for every open set $O \in \mathcal{B}_Y : f^{-1}(O) \in \mathcal{B}_X$.

Example 18.1.31 (Lebesgue-measurable function). A function $f : \mathbb{R} \to \mathbb{R}$ such that for every interval $I \subset \mathbb{R} : f^{-1}(I) \in \mathcal{M}$.

Remark 18.1.32. The inclusion $\mathcal{B} \subset \mathcal{M}$ implies that every Borel-measurable function is also Lebesgue-measurable.

Property 18.1.33. The class of Borel/Lebesgue-measurable functions defined on $E \in \mathcal{M}$ forms an algebra.

Example 18.1.34. The following types of functions are Lebesgue-measurable:

- monotonic functions,
- continuous functions, and
- indicator functions.

Corollary 18.1.35. Let f, g be Lebesgue-measurable functions and let $F : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a continuous function. The composition $F\big(f(x), g(x)\big)$ is also measurable.

Property 18.1.36. Let f be a Lebesgue-measurable function. The level set $\{x \mid f(x) = a\}$ is measurable for all $a \in \mathbb{R}$.

Property 18.1.37. Define the following functions (which are measurable if f is measurable as a result of the previous properties):

$$f^+(x) := \mathsf{max}(f, 0) = \begin{cases} f(x) & \text{if } f(x) > 0 \\ 0 & \text{if } f(x) \leq 0, \end{cases} \tag{18.10}$$

$$f^-(x) := \mathsf{max}(-f, 0) = \begin{cases} 0 & \text{if } f(x) > 0 \\ -f(x) & \text{if } f(x) \leq 0\,. \end{cases} \tag{18.11}$$

The function $f : \mathbb{R} \to \mathbb{R}$ is measurable if and only if both f^+ and f^- are measurable. Furthermore, f is measurable if $|f|$ is measurable (the converse is false in general).

Definition 18.1.38 (Pushforward). Consider two measurable spaces (X_1, Σ_1) and (X_2, Σ_2) together with a measurable function $f : X_1 \to X_2$. For every measure μ on X_1 one can define the pushforward measure $f_*\mu$ on X_2 as follows:

$$f_*\mu(A) := \mu\big(f^{-1}(A)\big)\,. \tag{18.12}$$

Definition 18.1.39 (Measure-preserving function). Let (X, Σ, μ) be a measure space and consider a measurable function $T : X \to X$. T is said to be measure-preserving if

$$T_*\mu = \mu\,. \tag{18.13}$$

These functions form the morphisms in the category **Meas** of measure spaces.

Definition 18.1.40 (Ergodic function). Let (X, Σ, μ) be a measure space and consider a measure-preserving function $T : X \to X$. It is said to be ergodic if the following condition is satisfied:

$$T(A) = A \implies \mu(A) = 0 \vee \mu(X \backslash A) = 0\,. \tag{18.14}$$

This is equivalent to stating that for every set $A \in \Sigma$ with positive measure the following condition holds:

$$\mu\left(\bigcup_{n=1}^{+\infty} T^{-n}(A)\right) = 1\,. \tag{18.15}$$

Property 18.1.41. Consider a topological space X with Borel σ-algebra $\mathcal{B}$ and let T be an ergodic function. Almost every T-orbit is dense in the support of μ.

Definition 18.1.42 (Mixing). An endomorphism of a measure spaces (X, Σ, μ) is said to be mixing if for all measurable spaces A, B the following equality holds:

$$\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B) \tag{18.16}$$

Property 18.1.43. All mixing transformations are ergodic.

Property 18.1.44 (Additivity). Every measurable, additive function $f : \mathbb{R} \to \mathbb{R}$ is linear.

Corollary 18.1.45. From the basic properties of exponential and logarithmic functions, the following results can be obtained:

- Let $f : \mathbb{R} \to \mathbb{R}$ be a measurable function. If $f(x + y) = f(x)f(y)$, then $f(x) = e^{\lambda x}$ for some $\lambda \in \mathbb{R}$.
- Let $f : [0, +\infty] \to \mathbb{R}$ be a measurable function. If $f(xy) = f(x) + f(y)$, then $f(x) = \lambda \log(x)$ for some $\lambda \in \mathbb{R}$.
- Let $f : [0, +\infty] \to [0, +\infty]$ be a measurable function. If $f(xy) = f(x)f(y)$, then $f(x) = x^{\lambda}$ for some $\lambda \in \mathbb{R}$.

18.1.4 Limit operations

Property 18.1.46. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of measurable functions. The following functions are also measurable:

- $\min\limits_{i\leq k}(f_i)$ and $\max\limits_{i\leq k}(f_i)$,
- $\inf\limits_{n\in\mathbb{N}}(f_n)$ and $\sup\limits_{n\in\mathbb{N}}(f_n)$, and
- $\liminf\limits_{n\to\infty}(f_n)$ and $\limsup\limits_{n\to\infty}(f_n)$.

Property 18.1.47. If f is a measurable function and g is a function such that $f = g$ almost everywhere, then g is measurable as well.

Corollary 18.1.48. As a result of the previous two properties, if a sequence of measurable functions converges pointwise a.e., the limit is also a measurable function.

Definition 18.1.49 (Essential supremum).

$$\operatorname{ess\,sup}(f) := \inf\{z \in \mathbb{R} \mid f \leq z \text{ a.e.}\} \tag{18.17}$$

Definition 18.1.50 (Essential infimum).

$$\operatorname{ess\,inf}(f) := \sup\{z \in \mathbb{R} \mid f \geq z \text{ a.e.}\} \tag{18.18}$$

Property 18.1.51. Every measurable function f satisfies the following inequalities:

- $f \leq \operatorname{ess\,sup}(f)$ a.e. and $f \geq \operatorname{ess\,inf}(f)$ a.e., and
- $\operatorname{ess\,sup}(f) \leq \sup(f)$ and $\operatorname{ess\,inf}(f) \geq \inf(f)$.

The latter pair of inequalities becomes a pair of equalities if f is continuous.

Property 18.1.52. If f, g are measurable functions, then $\operatorname{ess\,sup}(f + g) \leq \operatorname{ess\,sup}(f) + \operatorname{ess\,sup}(g)$. An analogous inequality holds for the essential infimum.

Definition 18.1.53 (Weak convergence). A sequence of measures $(\mu_n)_{n\in\mathbb{N}}$ is said to converge weakly to a measure μ on a metrizable space X if any of the following conditions is satisfied:

1. $\int_X f\, d\mu_n \longrightarrow \int_X f\, d\mu$ for all bounded, continuous functions f.
2. $\mu_n(A) \longrightarrow \mu(A)$ for all **continuity sets** A of μ, i.e. for all Borel sets A such that $\mu(\partial A) = 0$.
3. $\liminf \mu_n(U) \geq \mu(U)$ for all open sets U.
4. $\limsup \mu_n(V) \leq \mu(V)$ for all closed sets V.

If $X = \mathbb{R}$ with its standard topology, the sequence $(\mu_n)_{n\in\mathbb{N}}$ converges weakly to μ if and only if $\mu_n(\{x \in \mathbb{R} : x \leq y\}) \longrightarrow \mu(\{x \in \mathbb{R} : x \leq y\})$ for all points $y \in \mathbb{R}$ where these functions are continuous.

Remark 18.1.54 (Relation to functional weak convergence ♣). Consider Definition 25.1.2 of weak convergence in functional analysis. By the Riesz–Markov theorem 19.1.6 one can identify measures with functionals on functions vanishing at infinity. By noting that bounded continuous functions on X are equivalent to functions of compact support on the *Stone–Čech compactification* βX, one obtains that a sequence of measures converges weakly if and only if the associated sequence of functionals converges weakly.

@@ CHECK THIS STATEMENT @@

18.1.5 Polish spaces

The content of this section heavily relies on that of Chapter 6 and Chapter 9.

Definition 18.1.55 (Polish space). A separable, completely metrizable space.

Definition 18.1.56 (Standard Borel space). A Borel space associated to a Polish space.

Theorem 18.1.57 (Kuratowski). *Every standard Borel space is (Borel) isomorphic to either $\mathbb{R}$, $\mathbb{Z}$ or a finite set. Hence, it is completely characterized by its cardinality.*

Property 18.1.58. All open and closed subsets of a Polish space are Polish. So are G_δ-subsets (Definition 6.1.13).

Definition 18.1.59 (Lusin space). A topological space that is homeomorphic to a Borel subset of a compact metric space. Equivalently, a topological space that admits a stronger topology that is Polish or, in other words, it is the image of a Polish space under a continuous bijection. In particular, every Polish space is Lusin.

Definition 18.1.60 (Suslin space[3]). The image of a Polish space under a continuous function. In particular, every Lusin space is Suslin.

Theorem 18.1.61 (Lusin–Suslin). *A subset of a Polish space is Lusin if and only if it is Borel.*

18.2 Lebesgue integral

18.2.1 Simple functions

Definition 18.2.1 (Indicator function).

$$\mathbb{1}_A(x) := \begin{cases} 1 & \text{if } x \in A\,, \\ 0 & \text{if } x \notin A\,. \end{cases} \tag{18.19}$$

Definition 18.2.2 (Simple function). A function $f : X \to \mathbb{R}$ on a measurable space (X, Σ) that can be expressed as

$$f(x) = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}(x) \tag{18.20}$$

for some $\{a_i \geq 0\}_{i \leq n}$, $\{A_i\}_{i \leq n} \subset \Sigma$ and $n \in \mathbb{N}$.

Definition 18.2.3 (Step function). If $(X, \Sigma) = (\mathbb{R}, \mathcal{M})$ and the sets A_i are intervals, the above function is often called a step function.

Definition 18.2.4 (Lebesgue integral of simple functions). Consider a simple function φ on a measure space (X, Σ, μ). The Lebesgue integral of φ over a measurable set $A \in \Sigma$

[3]Sometimes written as 'Souslin'.

with respect to μ is given by

$$\int_A \varphi \, d\mu := \sum_{i=1}^{n} a_i \mu(A \cap A_i) \,. \tag{18.21}$$

As usual, if the domain of integration is not mentioned explicitly, an integral over the whole space X is implied.

Example 18.2.5. Let $\mathbb{1}_{\mathbb{Q}}$ be the indicator function of the rational numbers. Contrary to the case of Riemann integrals, the above definition makes it possible to integrate the rational indicator function over the real line:

$$\int_{\mathbb{R}} \mathbb{1}_{\mathbb{Q}} \, d\lambda = 1 \times \lambda(\mathbb{Q}) + 0 \times \lambda(\mathbb{R} \backslash \mathbb{Q}) = 0 \,, \tag{18.22}$$

where the measure of the rational numbers is 0 because it is a countable set (Property 18.1.20).

18.2.2 Measurable functions

Definition 18.2.6 (Integral for nonnegative functions). The definition for simple functions can be generalized to nonnegative measurable functions f as follows:

$$\int_A f \, d\mu := \sup \left\{ \int_A \varphi \, d\mu \,\middle|\, \varphi \text{ a simple function such that } \varphi \leq f \right\} . \tag{18.23}$$

This integral is always nonnegative.

Formula 18.2.7. The following equality allows to change the domain of integrals:

$$\int_A f \, d\mu = \int_X f \mathbb{1}_A \, d\mu \,. \tag{18.24}$$

Property 18.2.8. The Lebesgue integral over a null set is 0.

Theorem 18.2.9 (Mean value theorem). *If $a \leq f(x) \leq b$, then $a\lambda(A) \leq \int_A f \, d\lambda \leq b\lambda(A)$.*

Property 18.2.10 (Simple approximation). Let f be a nonnegative measurable function. There exists an increasing sequence $(\varphi_n)_{n \in \mathbb{N}}$ of simple functions such that $\varphi_n \nearrow f$. Moreover, if f is bounded on $A \in \Sigma$, the sequence can be chosen to be uniformly convergent on A.

18.2.3 Integrable functions

Definition 18.2.11 (Integrable function). Let A be a measurable subset of a measure space (X, Σ, μ). A measurable function f is said to be integrable over A if both $\int_A f^+ \, d\mu$ and $\int_A f^- \, d\mu$ are finite. The Lebesgue integral of f over A is then defined as

$$\int_A f \, d\mu := \int_A f^+ \, d\mu - \int_A f^- \, d\mu \,. \tag{18.25}$$

If only one of the functions f^+, f^- is finite, f is said to be **quasi-integrable**.

Property 18.2.12 (Absolute integrability). f is integrable if and only if $|f|$ is integrable. Furthermore,

$$\int_A |f|\, d\mu = \int_A f^+\, d\mu + \int_A f^-\, d\mu\,. \tag{18.26}$$

Property 18.2.13. Let f, g be integrable functions on a measure space (X, Σ, μ). The following important properties hold:

- **Linearity**: $\int_A (f + \lambda g) d\mu = \int_A f\, d\mu + \lambda \int_A g\, d\mu$ for all $\lambda \in \mathbb{R}$
- **Monotonicity**: $f \leq g$ a.e. implies $\int_A f\, d\mu \leq \int_A g\, d\mu$ and $\forall A \in \Sigma : \int_A f\, d\mu \leq \int_A g\, d\mu \implies f \leq g$ a.e.
- **Finiteness**: f is finite a.e.
- $|\int_A f\, d\mu| \leq \int_A |f|\, d\mu$.
- $\int_A f\, d\mu = 0, \forall A \in \Sigma \implies f = 0$ a.e.

Definition 18.2.14 (Integrable functions). The set of integrable functions over a set $A \in \mathcal{M}$ forms the vector space $\mathcal{L}^1(A)$.

Property 18.2.15 (Continuous approximation). Let $f \in \mathcal{L}^1$ and $\varepsilon > 0$. There exists a continuous (or step or even simple) function g, vanishing outside a finite (or even compact) set, such that $\int |f - g|\, d\mu < \varepsilon$.

Definition 18.2.16 (Locally integrable function). A measurable function is said to be locally integrable if it is integrable on every compact subset of its domain. The space of locally integrable functions is denoted by $\mathcal{L}^1_{\text{loc}}$.

Example 18.2.17. All continuous functions are locally integrable.

Property 18.2.18 (Absolute continuity). Let $f \geq 0$ be a measurable function. The mapping $A \mapsto \int_A f\, d\mu$ defines a measure that is σ-finite if f is locally integrable and finite if f is integrable. Furthermore, this measure is said to be absolutely continuous (with respect to μ). See Section 18.5 for a generalization to arbitrary measures.

18.2.4 Convergence theorems

Theorem 18.2.19 (Fatou's lemma). *Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of nonnegative measurable functions.*

$$\int_A \left(\liminf_{n\to\infty} f_n\right) d\mu \leq \liminf_{n\to\infty} \int_A f_n\, d\mu \tag{18.27}$$

Theorem 18.2.20 (Monotone convergence). *Let A be measurable and let $(f_n)_{n\in\mathbb{N}}$ be an increasing sequence of nonnegative measurable functions such that $f_n \nearrow f$ pointwise a.e.*

$$\int_A f\, d\mu = \lim_{n\to\infty} \int_A f_n\, d\mu\,. \tag{18.28}$$

Method 18.2.21. To prove results concerning integrable functions in spaces such as $\mathcal{L}^1$ it is often useful to proceed as follows:

1. Verify that the property holds for indicator functions. (This often follows by definition.)
2. Use linearity to extend the property to simple functions.
3. Apply the monotone convergence theorem to show that the property holds for all nonnegative measurable functions.
4. Extend the property to all integrable functions by decomposing $f = f^+ - f^-$ and applying linearity again.

Theorem 18.2.22 (Dominated convergence). *Let A be measurable set and consider a sequence of measurable functions $(f_n)_{n\in\mathbb{N}}$ such that $\forall n \in \mathbb{N} : |f_n| \leq g$ a.e. for some function $g \in \mathcal{L}^1(A)$. If $f_n \longrightarrow f$ pointwise a.e., then f is integrable over A and*

$$\int_A f\,d\mu = \lim_{n\to\infty} \int_A f_n\,d\mu\,. \tag{18.29}$$

Property 18.2.23. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of nonnegative measurable functions

$$\int_A \sum_{n=1}^{+\infty} f_n\,d\mu = \sum_{n=1}^{+\infty} \int_A f_n\,d\mu\,. \tag{18.30}$$

One cannot conclude that the right-hand side is finite a.e., so the series on the left-hand side need not be integrable.

Theorem 18.2.24 (Beppo Levi[4]). *Suppose that*

$$\sum_{i=1}^{+\infty} \int_A |f_n|\,d\mu$$

is finite. The series $\sum_{i=1}^{+\infty} f_n(x)$ converges a.e. Furthermore, the series is integrable and

$$\int_A \sum_{i=1}^{+\infty} f_n\,d\mu = \sum_{i=1}^{+\infty} \int_A f_n\,d\mu\,. \tag{18.31}$$

Theorem 18.2.25 (Riemann–Lebesgue lemma). *Let $f \in \mathcal{L}^1(\mathbb{R})$. The sequences*

$$s_k = \int_{\mathbb{R}} f(x)\sin(kx)\,dx \tag{18.32}$$

and

$$c_k = \int_{\mathbb{R}} f(x)\cos(kx)\,dx \tag{18.33}$$

both converge to 0.

[4] Various other theorems and variants of this theorem can be found in the literature under the same name.

Theorem 18.2.26 (Birkhoff ergodicity). *Let (X, Σ, μ) be a measure space and let T be a μ-ergodic map. For every measurable function f and for μ-almost every element $x \in X$ the integral of f can be computed as an average over the orbit of x:*

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{t=0}^{n} f\big(T^n(x)\big) = \int_X f \, d\mu \,. \tag{18.34}$$

18.2.5 Relation to the Riemann integral

Property 18.2.27. Let $f : [a, b] \to \mathbb{R}$ be a bounded function.

- f is Riemann-integrable if and only if f is continuous a.e. with respect to the Lebesgue measure on $[a, b]$, i.e. the set of discontinuities of f has measure zero.
- Riemann-integrable functions on $[a, b]$ are integrable with respect to the Lebesgue measure on $[a, b]$ and the integrals coincide.

Property 18.2.28. If $f \geq 0$ and the improper Riemann integral (Definition 16.6.1) exists, the Lebesgue integral exists and the two integrals coincide. Note that positivity is necessary here. Because the Lebesgue integral is absolute (Property 18.2.12), positive and negative parts cannot cancel (Lebesgue integrals can never be conditionally convergent).

The following definition should be compared to Definition 18.2.1 and Definition 19.1.20.

Definition 18.2.29 (Dirac measure). Define the Dirac (delta) measure as follows:

$$\delta_a(A) := \begin{cases} 1 & \text{if } a \in A \,, \\ 0 & \text{if } a \notin A \,. \end{cases} \tag{18.35}$$

Integration with respect to the Dirac measure has the following important property:

$$\int_X f \, d\delta_a = f(a) \,. \tag{18.36}$$

18.3 Space of integrable functions

18.3.1 Distance

To define a distance between functions, a notion of the length of a function is introduced first. Normally, this would not be a problem, one could use the integral of a function to define a norm. However, the fact that two functions differing on a null set have the same integral carries problems with it: a nonzero function could have a zero length. To avoid this issue, these degenerate functions are quotiented out.

Definition 18.3.1 (L^1-space). Define the set of equivalence classes $L^1 = \mathcal{L}^1_{/\equiv}$ by introducing the following equivalence relation: $f \equiv g$ if and only if $f = g$ a.e.

Property 18.3.2. L^1 is a *Banach space* (see Definition 25.1.9). The norm on L^1 is given by

$$\|f\|_1 := \int_X |f| \, d\mu \,. \tag{18.37}$$

In particular,

$$\int_X |f| \, d\mu = 0 \implies f = 0 \text{ a.e.} \tag{18.38}$$

18.3.2 Hilbert space L^2

Property 18.3.3. L^2 is a *Hilbert space* (see Definition 25.2.4). The norm on L^2 is given by

$$\|f\|_2 := \left(\int_X |f|^2 \, d\mu \right)^{1/2} . \tag{18.39}$$

This norm is induced by the following inner product:

$$\langle f \,|\, g \rangle := \int_X \overline{f} g \, d\mu \,. \tag{18.40}$$

Formula 18.3.4 (Cauchy–Schwarz inequality). Let $f, g \in L^2(X, \mathbb{C})$. Formula 18.3.7 implies that $fg \in L^1(X, \mathbb{C})$ and

$$\left| \int \overline{f} g \, d\mu \right| \leq \|fg\|_1 \leq \|f\|_2 \|g\|_2 \,. \tag{18.41}$$

18.3.3 L^p-spaces

Generalizing the previous two function classes leads to the notion of L^p-spaces with the following norm.

Formula 18.3.5. For all $1 \leq p \leq \infty$, $L^p(X)$ is a *Banach space* (see Definition 25.1.9) when equipped with the following norm:

$$\|f\|_p := \left(\int_X |f|^p \, d\mu \right)^{1/p} . \tag{18.42}$$

Remark 18.3.6. Note that L^2 is the only L^p-space that is also a *Hilbert space* (see 25.2.4). The other L^p-spaces do not have a norm induced by an inner product.

Formula 18.3.7 (Hölder's inequality). Let $\frac{1}{p} + \frac{1}{q} = 1$ with $p \geq 1$ (numbers satisfying this equality are called **Hölder conjugates**). For every $f \in L^p$ and $g \in L^q$ one has that

$$\|fg\|_1 \leq \|f\|_p \|g\|_q \,. \tag{18.43}$$

This also implies that $fg \in L^1$.

Formula 18.3.8 (Minkowski's inequality). For every $p \geq 1$ and $f, g \in L^p$ one has that

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p \,. \tag{18.44}$$

This also implies that $f + g \in L^p$.

Property 18.3.9 (Inclusions). $L^1(X) \cap L^\infty(X) \subset L^2(X)$. Moreover, if X has finite measure, then $L^q(X) \subset L^p(X)$ whenever $1 \leq p \leq q < +\infty$.

Using the Hölder inequality one can prove the following property.

Property 18.3.10. Let p, q be Hölder conjugates. The spaces L^p and L^q are topological duals, i.e. every function $f \in L^p$ can be identified (one-to-one) with a continuous functional on L^q.

Definition 18.3.11 (Essentially bounded function). Let f be a measurable function satisfying $\operatorname{ess\,sup} |f| < +\infty$. The function f is said to be essentially bounded and the set of all such functions is denoted by L^∞ (again after quotienting out all functions that are equal a.e.).

Formula 18.3.12. A norm on L^∞ is given by

$$\|f\|_\infty := \operatorname{ess\,sup} |f| \,. \tag{18.45}$$

This norm is called the **supremum norm** and it induces the supremum metric (Example 9.3.3).

Property 18.3.13. Equipped with the above norm the space L^∞ becomes a *Banach space* (see Section 25.1).

Definition 18.3.14 (Weak type operator). Consider a function $T : L^p(X, \mu) \to L^q(Y, \nu)$. It is said to be of weak type-(p, q) if it satisfies the following estimate: There exists a constant $C > 0$ such that

$$\nu\big(\|Tf\|_q > \lambda\big) \leq C\left(\frac{\|f\|_p}{\lambda}\right)^q . \tag{18.46}$$

It is said to be of strong type-(p, q) if it is Lipschitz continuous.

Definition 18.3.15 (Hardy–Littlewood maximum function). Consider the following function on (locally) integrable functions $f \in L^1_{\text{loc}}(\mathbb{R}^n)$:

$$Mf(x) := \sup_{r>0} \frac{1}{\lambda^n\big(B(x,r)\big)} \int_{B(x,r)} |f(y)| \, dy \,, \tag{18.47}$$

where λ^n is the Lebesgue measure.

Property 18.3.16 (Hardy–Littlewood maximal inequality). The Hardy–Littlewood function is of weak (and even strong) type-$(1, 1)$. Moreover, the constant in these estimates does not depend on the dimension $d \in \mathbb{N}$.

The following result should be compared to Theorem 16.6.2.

Corollary 18.3.17 (Lebesgue differentiation theorem). Consider $f \in L^1_{\text{loc}}(\mathbb{R}^n)$. For almost all $x \in \mathbb{R}^n$, the following equality holds:

$$f(x) = \lim_{r \to 0} \frac{1}{\lambda^n\big(B(x,r)\big)} \int_{B(x,r)} f(y)\, dy\,. \tag{18.48}$$

18.4 Product measures

18.4.1 Construction

The general condition for product measures is given by the following equation that should hold for all $A_1 \in \Sigma_1$ and $A_2 \in \Sigma_2$:

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)\,. \tag{18.49}$$

Definition 18.4.1 (Section). Let $A = A_1 \times A_2$. The following two sets are called sections:

$$A_{x_1} := \{x_2 \in X_2 \mid (x_1, x_2) \in A\} \subset \Sigma_2\,,$$
$$A_{x_2} := \{x_1 \in X_1 \mid (x_1, x_2) \in A\} \subset \Sigma_1\,.$$

The following property follows immediately from the definition of product σ-algebras (Definition 1.6.17).

Property 18.4.2. Let $\Sigma := \Sigma_1 \times \Sigma_2$ be the product σ-algebra. If $A \in \Sigma$, then $A_{x_1} \in \Sigma_2$ for each $x_1 \in X_1$ and $A_{x_2} \in \Sigma_1$ for each $x_2 \in X_2$. Equivalently, the sets $\mathcal{G}_1 = \{A \in \Sigma \mid \forall x_1 \in X_1 : A_{x_1} \in \Sigma_2\}$ and $\mathcal{G}_2 = \{A \in \Sigma \mid \forall x_2 \in X_2 : A_{x_2} \in \Sigma_1\}$ coincide with Σ.

Property 18.4.3. The function $A_{x_2} \mapsto \mu(A_{x_2})$ is a step function:

$$\mu(A_{x_2}) = \begin{cases} \mu_1(A_1) & \text{if } x_2 \in A_2\,, \\ 0 & \text{if } x_2 \notin A_2\,. \end{cases} \tag{18.50}$$

Formula 18.4.4 (Product measure). From the previous property it follows that the product measure $\mu(A)$ can be written in the following way:

$$\mu(A) = \int_{X_2} \mu_1(A_{x_2})\, d\mu_2(x_2)\,. \tag{18.51}$$

Property 18.4.5. Let μ_1, μ_2 be σ-finite measures. If $A \in \Sigma$, the functions

$$x_1 \mapsto \mu_2(A_{x_1}) \qquad \text{and} \qquad x_2 \mapsto \mu_1(A_{x_2})$$

are measurable with respect to Σ_1 and Σ_2, respectively. Moreover,

$$\int_{X_2} \mu_1(A_{x_2})\, d\mu_2(x_2) = \int_{X_1} \mu_2(A_{x_1})\, d\mu_1(x_1)\,. \tag{18.52}$$

Furthermore, the set function μ is countably additive and if any other product measure coincides with μ on all product sets, it coincides with μ on the whole product σ-algebra.

18.4.2 Fubini's theorem

Property 18.4.6. Let $f : X_1 \times X_2 \to \mathbb{R}$ be a nonnegative function. If f is measurable with respect to $\Sigma_1 \times \Sigma_2$, then for each $x_1 \in X$ the function $x_2 \mapsto f(x_1, x_2)$ is measurable with respect to Σ_2 (and vice versa). Their integrals with respect to μ_1 and μ_2 respectively are also measurable.

Definition 18.4.7 (Section). The functions $x_1 \mapsto f(x_1, x_2)$ and $x_2 \mapsto f(x_1, x_2)$ are called sections of f.

Theorem 18.4.8 (Tonelli). *Let $f : X_1 \times X_2 \to \mathbb{R}$ be a nonnegative function. The following equalities hold:*

$$\begin{aligned}\int_{X_1 \times X_2} f \, d\mu &= \int_{X_1} \left(\int_{X_2} f(x_1, x_2) d\mu_2(x_2) \right) d\mu_1(x_1) \\ &= \int_{X_2} \left(\int_{X_1} f(x_1, x_2) d\mu_1(x_1) \right) d\mu_2(x_2) \,. \end{aligned} \tag{18.53}$$

Corollary 18.4.9 (Fubini). Let $f \in L^1(X_1 \times X_2)$. The sections of f are integrable in the appropriate spaces. Furthermore, the functions

$$x_1 \mapsto \int_{X_2} f(x_1, x_2) \, d\mu_2(x_2) \tag{18.54}$$

and

$$x_2 \mapsto \int_{X_1} f(x_1, x_2) \, d\mu_1(x_1) \tag{18.55}$$

are in $L^1(X_1)$ and $L^1(X_2)$, respectively, and Tonelli's theorem holds.

Remark 18.4.10. The previous construction and theorems also apply to higher-dimensional product spaces. These theorems provide a way to construct higher-dimensional measures by defining them (as the completion of) the product of measures.

18.5 Radon–Nikodym theorem

Definition 18.5.1 (Absolute continuity). Let (X, Σ) be a measurable space and let μ, ν be two measures defined on this space. Then ν is said to be absolutely continuous with respect to μ if

$$\forall A \in \Sigma : \mu(A) = 0 \implies \nu(A) = 0 \,. \tag{18.56}$$

This relation is often denoted by $\nu \ll \mu$. If both $\mu \ll \nu$ and $\nu \ll \mu$, the measures are said to be equivalent. This is denoted by $\mu \sim \nu$.

The following property relates the notion of absolute continuity above with that of Definition 16.2.5.

Property 18.5.2 (Absolute continuity). Let μ, ν be finite measures on a measurable space (X, Σ). Then $\nu \ll \mu$ if and only if

$$\forall \varepsilon > 0 : \exists \delta > 0 : \forall A \in \Sigma : \mu(A) < \delta \implies \nu(A) < \varepsilon \,. \tag{18.57}$$

Definition 18.5.3 (Singular measures). Consider two measures μ, ν. If there exists a set A such that $\mu(A) = 0 = \nu(A^c)$, they are said to be singular (or **orthogonal**). This is denoted by $\mu \perp \nu$.

Theorem 18.5.4 (Lebesgue's decomposition theorem). *Let μ, ν be two σ-finite measures. There exist two other σ-finite measures ν_a, ν_s such that $\nu = \nu_a + \nu_s$, where $\nu_a \ll \mu$ and $\nu_s \perp \mu$.*

Definition 18.5.5 (Dominated measure). Let μ, ν be two measures defined on a measurable space (X, Σ). Then μ is said to **dominate** ν if $0 \leq \nu(F) \leq \mu(F)$ for every $F \in \Sigma$.

Theorem 18.5.6 (Radon–Nikodym theorem). *Let (X, Σ) be a measurable space and let μ, ν be two σ-finite measures defined on Σ such that $\nu \ll \mu$. There exists a nonnegative, measurable function $f : X \to \mathbb{R}$ such that*

$$\nu(A) = \int_A f \, d\mu \tag{18.58}$$

for all $A \in \Sigma$.

Definition 18.5.7 (Radon–Nikodym derivative). The function f in the previous theorem is called the Radon-Nikodym derivative of ν with respect to μ. It is generally denoted by $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$.

Remark 18.5.8. The function f in this theorem is unique up to a μ-null (and thus ν-null) set.

Remark 18.5.9 (Freudenthal). ALthough not immediately obvious, the Radon–Nikodym theorem is a direct consequence of the Freudenthal spectral theorem 25.4.8. In this setting, the e-simple functions are exactly the simple functions and the convergent sequence admits an application of the monotone convergence theorem. For any (signed) measure μ on a measurable space (X, Σ), the spectral theorem establishes a lattice isomorphism between the band generated by μ and the space $L^1(X, \Sigma, \mu)$.

Property 18.5.10 (Integrability). In general, the Radon–Nikodym derivative is not integrable (unless the measures are finite). However, it is always locally integrable (Definition 18.2.16). Together with Property 18.2.18, this implies that (densities of) absolutely continuous measures are in bijection with locally integrable functions.

Property 18.5.11 (Change of variables). Let μ, ν be finite measures such that $\nu \ll \mu$ and let $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ be the associated Radon–Nikodym derivative. For every ν-integrable func-

tion f the following equality holds

$$\int_A f\,d\nu = \int_A f\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\,d\mu \tag{18.59}$$

for all $A \in \Sigma$.

Property 18.5.12. Let λ, ν and μ be σ-finite measures. If $\lambda \ll \mu$ and $\nu \ll \mu$, the following two properties hold:

- **Linearity**: $\dfrac{\mathrm{d}(\lambda+\nu)}{\mathrm{d}\mu} = \dfrac{\mathrm{d}\lambda}{\mathrm{d}\mu} + \dfrac{\mathrm{d}\lambda}{\mathrm{d}\mu}$.
- **Chain rule**: If $\lambda \ll \nu$, then $\dfrac{\mathrm{d}\lambda}{\mathrm{d}\mu} = \dfrac{\mathrm{d}\lambda}{\mathrm{d}\nu}\dfrac{\mathrm{d}\nu}{\mathrm{d}\mu}$ a.e.

18.6 Lebesgue–Stieltjes integral

Aside from the Lebesgue measure, one can construct some other important measures (and their associated integrals) on the Borel σ-algebra of the real line $\mathbb{R}$. These constructions will be important in the study of density functions in probability theory and the extension of calculus to stochastic processes. To this end, consider a function F that is right-continuous, i.e. $F(x^+) = F(x)$, and increasing. The length of an interval can be generalized in the following way.

Definition 18.6.1 (F-length). The F-length of an interval $]a, b]$ is defined as follows:

$$l_F\big(]a,b]\big) := F(b) - F(a)\,. \tag{18.60}$$

The restriction to half-open intervals assures that this function is additive when taking unions of intervals. The footnote in Definition 6.1.10 also assures that the σ-algebra generated by these intervals is the Borel σ-algebra on $\mathbb{R}$.

An immediate extension of Definition 18.1.15 gives the outer measure associated to F.

Definition 18.6.2 (F-outer measure). Let $X \subseteq \mathbb{R}$ be a set. The (Lebesgue–Stieltjes) F-outer measure of X is defined as follows:

$$\mu_F^*(X) := \inf\left\{\sum_{n=1}^{+\infty} l_F(I_n)\,\middle|\,(I_n)_{n\in\mathbb{N}} \text{ a sequence of half-open intervals that cover } X\right\}. \tag{18.61}$$

Using this outer measure, one can define the μ_F-measurable sets as those sets satisfying Carathéodory's criterion (with respect to μ_F^*). The main difference with the Lebesgue measure is that μ_F is not necessarily translation invariant and that singletons are not necessarily null.

Property 18.6.3 (Singletons). The F-measure of a singleton $\{x\}$ is equal to the jump of F at x:

$$\mu_F\big(\{x\}\big) = F(x) - F(x^-)\,. \tag{18.62}$$

Such elements are examples of **atoms**, sets of positive measure for which every proper measurable subset is null. Note that by right-continuity of F, the number of discontinuities is countable and the only atoms of μ_F are exactly these discontinuities.

Corollary 18.6.4. It follows that the Lebesgue–Stieltjes measures having null singletons are exactly those for which F is continuous.

Property 18.6.5 (Regularity). The Lebesgue–Stieltjes measure is a Radon measure. Furthermore, every Radon measure μ on $\mathbb{R}$ is equal to a Lebesgue–Stieltjes measure induced by the *cumulative distribution function* (see Definition 45.3.4)

$$F_\mu(x) = \mu\big(]-\infty, x]\big)\,. \tag{18.63}$$

Example 18.6.6 (Lebesgue measure). The Lebesgue measure is the Lebesgue–Stieltjes measure associated to $F(x) = x$.

Example 18.6.7 (Dirac measure). The Dirac measure at $x \in \mathbb{R}$ can be obtained as the Lebesgue–Stieltjes measure for $F(x) = \mathbb{1}_{[x,\infty[}$.

Property 18.6.8. If the Lebesgue–Stieltjes measure μ_F is nonatomic, then

$$F_*\mu_F = \lambda\,, \tag{18.64}$$

for λ the Lebesgue measure.

Property 18.6.9. Let μ, ν be two absolutely continuous measures[5] on $\mathbb{R}^n$ (with respect to the Lebesgue measure). There exists a unique increasing triangular Borel function $T : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$T_*\mu = \nu\,. \tag{18.65}$$

Triangular means that $Tx \equiv \big(f_1(x_1), f_2(x_1, x_2), \dots, f_n(x_1, \dots, x_n)\big)$ for some Borel functions $f_i : \mathbb{R}^i \to \mathbb{R}$ and increasing means that each f_i is an increasing function of x_i.

Remark 18.6.10. For $\mathbb{R}$, the Borel function T is obtained by first mapping μ to the Lebesgue measure on $[0, 1]$ and then mapping it to ν. Explicitly, let F_μ be the *cumulative distribution function* (see Definition 45.3.4) of μ:

$$F_\mu(x) := \mu\big(]-\infty, x]\big)\,, \tag{18.66}$$

and let G_ν be the **quantile function** of ν:

$$G_\nu(x) := \inf\{t \in \mathbb{R} \mid F_\nu(t) \geq x\}\,. \tag{18.67}$$

[5] In fact, one can relax this to merely being nonatomic.

The required transformation is then given by the composition:

$$T = G_\nu \circ F_\mu \,. \tag{18.68}$$

@@ ADD quantile function @@

Remark 18.6.11 (Bounded variation). Note that the Lebesgue–Stieltjes integral can be generalized to signed measures, through functions of (locally) bounded variation, since these are exactly the functions that can be decomposed as the difference of two increasing functions on any interval.

Formula 18.6.12 (Integration by parts). Let $F, G : \mathbb{R} \to \mathbb{R}$ be two right-continuous functions of lcoally bounded variation. Then

$$\begin{aligned} F(x)G(x) &= F(0)G(0) + \int_0^x F(s)\,dG(s) + \int_0^x G(s^-)\,dF(s) \\ &= F(0)G(0) + \int_0^x F(s^-)\,dG(s) + \int_0^x G(s^-)\,dF(s) + \sum_{s\le x} \Delta F(s)\Delta G(s)\,, \end{aligned} \tag{18.69}$$

where $\Delta F(s) := F(s) - F(s^-)$ denotes the **jump** at F, i.e. the Lebesgue–Stieltjes measure $\mu_F(\{s\})$. Note that for the sum on the second line, there are only countably many terms by *Froda's theorem*.

Formula 18.6.13 (Chain rule). Let $f : \mathbb{R} \to \mathbb{R}$ be continuously differentiable and let $F : \mathbb{R} \to \mathbb{R}$ be right-continuous and of locally bounded variation. Then

$$f(F(x)) = f(F(0)) + \int_0^x f'(F(s^-))\,dF(s) + \sum_{s\le x} \big(f(F(s)) - f(F(s^-)) - f'(F(s^-))\Delta F(s)\big)\,. \tag{18.70}$$

18.7 Signed measures

Definition 18.7.1 (Signed measure). Consider a measurable space (X, Σ). A function $\mu : \Sigma \to \overline{\mathbb{R}}$ is called a signed measure if it satisfies the following conditions:

1. **Measure zero**: $\mu(\emptyset) = 0$, and
2. **σ-additivity**: $\forall i \neq j : E_i \cap E_j = \emptyset \implies \mu\big(\bigcup_{n=1}^{+\infty} E_n\big) = \sum_{i=n}^{+\infty} \mu(E_n)$.

Note that these requirements are the same as for an ordinary measure (Definition 18.1.1), except that now the function is allowed to become negative. The function is, however, not allowed to attain $-\infty$ to exclude undefined expressions such as $\infty - \infty$.

Remark 18.7.2. An important consequence of this generalization is that signed measures are not necessarily monotonic, i.e. $A \subseteq B \not\Rightarrow \mu(A) \le \mu(B)$. In fact, this is a strict relation. A signed measure is monotonic if and only if it is a genuine measure.

Definition 18.7.3 (Total variation). Consider a signed measure μ on a measurable space (X, Σ). The total variation $|\mu|$ is the measure defined as follows:

$$|\mu|(A) := \sup\left\{ \sum_{P \in \mathcal{P}} |\mu(P)| \,\middle|\, \mathcal{P} \subset \Sigma, \mathcal{P} \text{ partitions } A \right\}. \tag{18.71}$$

If one chooses $A = X$, the total variation norm is obtained: $\|\mu\|_{\mathrm{TV}} := |\mu|(X)$.

Using this measure, one can decompose the signed measure μ as a difference of two genuine measures:

$$\begin{aligned} \mu &= \mu^+ - \mu^- \\ &= \frac{1}{2}(|\mu| + \mu) + \frac{1}{2}(|\mu| - \mu). \end{aligned} \tag{18.72}$$

Furthermore, this decomposition is minimal in the sense that if $\mu = \lambda_1 - \lambda_2$ for any two measures, then $\mu^+ \leq \lambda_1$ and $\mu^- \leq \lambda_2$.

The following theorem generalizes both the Radon–Nikodym and Lebesgue decomposition theorems to the case of signed measures.

Theorem 18.7.4. *Consider a σ-finite signed measure μ and a σ-finite measure ν on a measurable space (X, Σ). There exists a ν-a.e. unique integrable function $f \in L^1(\nu)$ and a σ-finite measure $\mu_s \perp \nu$ such that for all $A \in \Sigma$:*

$$\mu(A) = \int_A f \, d\nu + \mu_s(A). \tag{18.73}$$

As before, the function f in the preceding is called the Radon–Nikodym derivative of μ.

Theorem 18.7.5 (Hahn–Jordan). *Consider a signed measure μ on a measurable space (X, Σ). There exists a set $A \in \Sigma$ such that the minimal decomposition $\mu = \mu^+ - \mu^-$ in terms of two measures $\mu^\pm$ is given by*

$$\mu^+(B) = \mu(A \cap B) \qquad \mu^-(B) = \mu(A^c \cap B). \tag{18.74}$$

Definition 18.7.6 (Integral with respect to a signed measure). Let μ be a signed measure on a measurable space (X, Σ) and consider a measurable function f on $A \in \Sigma$. The integral of f with respect to μ is defined as follows:

$$\int_A f \, d\mu := \int_A f \, d\mu^+ - \int_A f \, d\mu^-. \tag{18.75}$$

Definition 18.7.7 (Lebesgue–Stieltjes signed measure). Let F be a function of bounded variation. According to Property 16.2.8, it can be written as $F = F_1 - F_2$, where F_1, F_2 are monotonically increasing, absolutely continuous functions. The **Lebesgue–Stieltjes (signed) measure** associated to F is defined as $\mu_F := \mu_{F_1} - \mu_{F_2}$.

Theorem 18.7.8 (Fundamental theorem of calculus). *Let F be an absolutely continuous function on the closed interval $[a, b]$. Then F is differentiable λ-a.e. (λ being the Lebesgue measure) and its associated Lebesgue–Stieltjes measure μ_F has Radon–Nikodym derivative $\frac{\mathrm{d}\mu_F}{\mathrm{d}\lambda} = F'$ λ-a.e. Furthermore, for all $x \in [a, b]$, one has*

$$F(x) - F(a) = \mu_F([a, x]) = \int_a^x F'(t)\, dt\,. \tag{18.76}$$

Corollary 18.7.9. If F is absolutely continuous and $F' = 0$ λ-a.e., then F is constant.

18.8 Bochner integral

In this section, the Lebesgue integral is generalized to functions taking values in general *Banach spaces* (see Section 25.1). Since addition still makes sense in these spaces, Definition 18.2.4 can still be used.

Definition 18.8.1 (Bochner integral of simple functions). Consider a simple function φ on a measure space (X, Σ, μ) taking values in a Banach space V. The Lebesgue integral of φ over a measurable set $A \in \Sigma$ with respect to μ is given by

$$\int_A \varphi\, d\mu := \sum_{i=1}^n a_i \mu(A \cap A_i)\,. \tag{18.77}$$

As usual, if the domain of integration is not mentioned explicitly, an integral over the whole space X is implied. If the above integral is finite, the simple function is said to be integrable.

Definition 18.8.2 (Bochner integrability). A Banach space-valued measurable function $f : (X, \Sigma) \to (V, \Omega)$ is said to be Bochner integrable if there exists a sequence $(\varphi_n)_{n\in\mathbb{N}}$ of integrable simple functions such that

$$\lim_{n\to\infty} \int_X \|f - \varphi_n\|_V\, d\mu = 0\,. \tag{18.78}$$

The Bochner integral of such a function is given by

$$\int_X f\, d\mu := \lim_{n\to\infty} \int_X \varphi_n\, d\mu\,. \tag{18.79}$$

@@ COMPLETE @@

Chapter 19

Distributions

The main references for this chapter are Choquet-Bruhat and DeWitt-Morette (2000); Choquet-Bruhat et al. (1991); Georgiev (2015). Although this chapter is technically part of functional analysis and, hence, uses the language of normed spaces (see Chapter 25), it is located in the part on calculus due to its strong relation to measure and integration theory. (Terminology from the chapter on functional analysis will not be indicated in italic for simplicity.)

Chapter Contents

19.1 Distributions **397**

19.1.1 Derivatives 399

19.1.2 Support 399

19.1.3 Examples 400

19.1.4 Growth rates 401

19.2 Convolutions and kernels **402**

19.3 Transformations **403**

19.3.1 Fourier series 403

19.3.2 Fourier transform 404

19.3.3 Wave front sets 406

19.3.4 Laplace transform 407

19.3.5 Integral representations 407

19.4 Harmonic analysis ♣ **408**

19.1 Distributions

Definition 19.1.1 (Distribution). The space of distributions or **generalized functions** on an open set $U \subset \mathbb{R}^n$ is defined as the set of continuous linear functionals on $\mathcal{D}(U) := C_c^\infty(U)$, the space of smooth functions with compact support (also called **bump functions** or **test functions**).

First, $\mathcal{D}(U)$ has to be endowed with a topology. For every compact set $K \subset U$ and every $m \in \mathbb{N}$, a locally convex topology (25.3.9) on $\mathcal{D}_K^m(U) := C_K^m(U)$ is constructed using the following family of seminorms:

$$\mathcal{P} = \left\{ \sup_{x \in K} \|f^{(i)}(x)\| \,\middle|\, |i| \leq m \right\}. \tag{19.1}$$

A topology on all of $\mathcal{D}^m(U)$ is then defined as the inductive limit (Definition 6.2.4) over all compact subsets $K \subset U$, i.e. a subset of $\mathcal{D}^m(U)$ is open if and only if its intersection with all $\mathcal{D}_K^m(U)$ is open. All of these topologies are Fréchet (see Definition 25.3.10). A topology on $\mathcal{D}(U)$ is obtained by taking a further inductive limit of the $\mathcal{D}^m(U)$ over $m \in \mathbb{N}$.

The dual space $\mathcal{D}'(U)$ is equipped with the weak-* topology (see Definition 25.1.3) and, accordingly, a sequence of distributions $(\phi_n)_{n\in\mathbb{N}}$ converges to a distribution ϕ if and only if $\langle \phi_n, f \rangle \longrightarrow \langle \phi, f \rangle$ for all $f \in \mathcal{D}(U)$. This definition immediately implies that two distributions ϕ, ψ are equal if and only if $\langle \phi, f \rangle = \langle \psi, f \rangle$ for all $f \in \mathcal{D}(U)$. Note that $\mathcal{D}(U)$ is not Fréchet in contrast to $C^\infty(U)$, where a countable family of seminorms is obtained by taking K to be closed balls of radius k.

Property 19.1.2 (Equivalent seminorms). The seminorms used in the definition of the locally convex topology on $\mathcal{D}(U)$ can be replaced by the following equivalent ones:

$$\begin{cases} \sup_{|i|\leq m} \sup_{x\in K} \|f^{(i)}(x)\|, \\ \sup_{x\in K} \sum_{|i|\leq m} \|f^{(i)}(x)\|, \\ \sum_{|i|\leq m} \sup_{x\in K} \|f^{(i)}(x)\|. \end{cases} \tag{19.2}$$

Property 19.1.3. A linear functional ϕ on $\mathcal{D}(U)$ is a distribution if and only if it satisfies one of the following equivalent statements:

- It is continuous when restricted to every $\mathcal{D}_K(U)$ for $K \subset U$ compact.
- If the sequence $(f_n)_{n\in\mathbb{N}}$ converges to 0 in $\mathcal{D}(U)$, then $\langle \phi, f_n \rangle \longrightarrow 0$.
- For every compact subset $K \subset U$, there exist a constant $C_K > 0$ and an integer $m_K \geq 0$ such that

$$|\langle \phi, f \rangle| \leq C_K \, p_{K,m_K}(f) \tag{19.3}$$

  for all $f \in \mathcal{D}_K(U)$.

Definition 19.1.4 (Order). The order of a distribution ϕ is the smallest integer m such that

$$|\langle \phi, f \rangle| \leq C_K \, p_{K,m}(f) \tag{19.4}$$

for all $f \in \mathcal{D}_K(U)$ and all compact subsets $K \subset U$. Note that the integer m is independent of the compact set K.

Property 19.1.5. A distribution is of order k if and only if it can be (uniquely) extended to a continuous linear functional on $\mathcal{D}^k(U) \equiv C_c^k(U)$.

Theorem 19.1.6 (Riesz–Markov–Kakutani). *The space of positive continuous functionals on $C_c(X)$, i.e. the space of continuous functions with compact support on a locally compact Hausdorff space X, is homeomorphic to the space of Radon measures (Definition 18.1.12) on X. Every such functional Λ can be represented as*

$$\Lambda(f) = \int_X f \, d\mu \tag{19.5}$$

for some Radon measure μ.

*The topological dual of $C_0(X) \equiv C(\widehat{X})$, the continuous functions on the one-point compactification (Definition 6.5.30), i.e. those functions that **vanish at infinity**, is isometrically isomorphic to the space of finite signed Radon measures equipped with the total variation norm (Definition 18.7.3).*

Example 19.1.7 (Ordinary function as generalized function). By Property 18.2.18, every locally integrable function $f \in L^1_{\text{loc}}$ gives rise to a distribution:

$$\langle f, g \rangle = \int_{\mathbb{R}} f(x) g(x) \, dx \, . \tag{19.6}$$

Distributions of this form are also said to be **regular**. These distributions are of order 0.

Property 19.1.8. The space $\mathcal{D}$ is dense in $\mathcal{D}'$.

The density of regular distributions can be used to generalize operations to all of $\mathcal{D}$.

Property 19.1.9 (Product with smooth functions). For every smooth function f and every distribution ϕ, the product $f\phi$ is defined as

$$\langle f\phi, g \rangle := \langle \phi, fg \rangle \, . \tag{19.7}$$

This turns $\mathcal{D}'$ into a C^∞-module.

19.1.1 Derivatives

Definition 19.1.10 (Derivative of a distribution). The derivative of a distribution ϕ is defined by duality:

$$\left\langle \frac{\partial \phi}{\partial x}, f \right\rangle := -\left\langle \phi, \frac{\partial f}{\partial x} \right\rangle . \tag{19.8}$$

This formula is a reasonable definition, since if ϕ is regular, the above formula is the one obtained through integration by parts.

In general, a function $g \in L^1_{\text{loc}}$ is said to be a **weak derivative** of a function $f \in L^1_{\text{loc}}$ if it satisfies the following equation for all $h \in \mathcal{D}$:

$$\langle f, h' \rangle = -\langle g, h \rangle . \tag{19.9}$$

Property 19.1.11 (Smoothness). Every distribution is smooth, i.e. it is infinitely differentiable. Furthermore, it satisfies the conclusion of Schwarz's theorem 16.5.11.

Property 19.1.12 (Constant distributions). If a distribution T satisfies $T' = 0$, it is a regular distribution induced by a constant function.

Definition 19.1.13 (Fundamental solution). Let D be a differential operator. A fundamental solution for D is a distribution ϕ such that

$$D\phi = \delta . \tag{19.10}$$

-

@@ DEFINE differential operator IN CHAPTER ON PDEs @@

19.1.2 Support

Definition 19.1.14 (Support). The support of a distribution is defined as the smallest closed set on which it does not vanish.

Property 19.1.15. A distribution has compact support if and only if it can be extended to a continuous linear functional on $C^\infty(U)$. This gives a nice duality: distributions act on compactly supported functions and compactly supported distributions act on functions.

Property 19.1.16 (Order). Distributions with compact support have finite order.

Property 19.1.17. A distribution that is supported only at 0 can be written as a linear combination of derivatives of the Dirac distribution. More generally, a distribution with support at a finite set of points can be written as a linear combination of (shifted) Dirac distributions.

Definition 19.1.18 (Singular support). The complement of the largest open set on which a distribution is regular. The singular support of ϕ is denoted by $\operatorname{sing\,supp}(\phi)$.

19.1.3 Examples

Definition 19.1.19 (Heaviside distribution). The Heaviside function is defined as follows:[1]

$$H(x) := \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x > 0 \end{cases} \tag{19.11}$$

From this definition, it follows that for every $f \in \mathcal{D}(U)$:

$$\langle H, f \rangle = \int_0^{+\infty} f(x)\, dx\,. \tag{19.12}$$

Definition 19.1.20 (Dirac delta distribution). The Dirac delta distribution is defined as the weak derivative of the Heaviside function:

$$\begin{aligned} \langle \delta, f \rangle &:= \langle H', f \rangle \\ &= -\langle H, f' \rangle \\ &= -\int_0^{+\infty} f'(x)\, dx \\ &= f(0)\,. \end{aligned}$$

Property 19.1.21 (Sampling property). The previous definition can be generalized in the following way (whenever $x_0 \in U$):

$$f(x_0) = \int_U f(x)\delta(x - x_0)\, dx = \int_U f\, d\delta_{x_0}\,, \tag{19.13}$$

where the suggestive notation[2] $\delta(x - x_0)$ was used to denote the Dirac delta distribution with support at x_0.

Definition 19.1.22 (Dirac comb).

$$\text{Ш}_b(x) := \sum_{n=-\infty}^{+\infty} \delta(x - nb) \tag{19.14}$$

Property 19.1.23 (Transformation). Let $f(x) \in C^1(\mathbb{R})$ be a function with $n \in \mathbb{N}$ roots x_i such that $f'(x_i) \neq 0$. The Dirac delta distribution has the following property:

$$\delta\big(f(x)\big) = \sum_{i=1}^{n} \frac{\delta(x - x_i)}{|f'(x_i)|}\,. \tag{19.15}$$

[1] The case $x = 0$ is often left undefined, but since this function will always enter formulas inside an integral this does not matter.

[2] See the section on kernels further on.

Formula 19.1.24 (Differentiation across discontinuities). Let f be a piecewise continuous function with discontinuities at $x_1, \dots, x_n$. Define the **jumps** of f at its discontinuities by $\sigma_i := f^+(x_i) - f^-(x_i)$. Next, define the (continuous) function

$$f_c(x) := f(x) - \sum_{i=1}^{n} \sigma_i H(x - x_i)\,. \tag{19.16}$$

Differentiation of this formula gives

$$f'(x) = f_c'(x) + \sum_{i=1}^{n} \sigma_i \delta(x - x_i)\,. \tag{19.17}$$

It follows that the derivative, in the generalized sense, of a piecewise continuous function equals the derivative in the classical sense plus a sum of delta functions at the jump discontinuities.

Example 19.1.25 (Principal value). The function $\frac{1}{x}$ is clearly not integrable on $\mathbb{R}$. However, its Cauchy principal value exists. This procedure also defines a distribution:

$$\left\langle \mathcal{P}\frac{1}{x}, f \right\rangle := \lim_{\varepsilon \downarrow 0} \int_{\varepsilon}^{+\infty} \frac{f(x) - f(x^-)}{x}\, dx. \tag{19.18}$$

Moreover, this is the distributional derivative of $\ln|x|$.

Formula 19.1.26. Consider a polynomial $P(x) \in \mathbb{C}[x]$. Over L^1 (or any ordinary function space), the following equation would not have any solutions:

$$P(x)f(x) = 0\,, \tag{19.19}$$

i.e. the kernel of the multiplication operator by $P(x)$ is trivial. However, when passing to distributions $\phi \in \mathcal{D}'$, the situation changes. Over $\mathbb{C}$, P splits and, hence, one obtains the following equation:

$$(x - \lambda_i)^{k_i} \cdots (x - \lambda_n)^{k_n} \phi = 0\,. \tag{19.20}$$

for some $n, k_1, \dots, k_n \in \mathbb{N}$ and $\lambda_1, \dots, \lambda_n \in \mathbb{C}$. The solution space is spanned by the following distributions:

$$\phi \in \mathsf{span}_{\mathbb{C}}\left\{\delta(x - \lambda_1), \dots, \delta^{(k_1)}(x - \lambda_1), \dots, \delta(x - \lambda_n), \dots, \delta^{(k_n)}(x - \lambda_n)\right\}, \tag{19.21}$$

i.e. it is spanned by delta distributions at the roots of the polynomial (and their derivatives up to order one less than the mulitplicity of the roots).

19.1.4 Growth rates

Definition 19.1.27 (Schwartz space). The Schwartz space of **rapidly decreasing functions** $\mathcal{S}(\mathbb{R}^n)$ is defined as follows:

$$\mathcal{S}(\mathbb{R}^n) := \left\{ f \in C^\infty(\mathbb{R}^n) \mid \forall i, j \in \mathbb{N}^n, \forall x \in \mathbb{R}^n : |x^i f^{(j)}(x)| < +\infty \right\}, \tag{19.22}$$

where, for every multi-index i, the symbol x^i denotes the monomial $x_1^{i_1} x_2^{i_2} \cdots$. An equivalent condition is the following. For every $p \in \mathbb{N}$ and $j \in \mathbb{N}^n$, there exists a constant $M_{p,j}(f)$ such that

$$\sup_{x \in \mathbb{R}^n} \big(1 + \|x\|\big)^p |f^{(j)}(x)| \leq M_{p,j}(f) \,. \tag{19.23}$$

This space has the structure of a Fréchet space under the family of seminorms

$$s_{p,N}(f) := \sup_{x \in \mathbb{R}^n} \sup_{|j| \leq N} \big(1 + \|x\|\big)^p |f^{(j)}(x)| \,. \tag{19.24}$$

Remark 19.1.28. These functions are said to be rapidly decreasing because every derivative $f^{(j)}(x)$ decays faster than any inverse power x^i for $\|x\| \longrightarrow +\infty$.

Definition 19.1.29 (Functions of slow growth). The space of functions of slow growth $\mathcal{N}(\mathbb{R}^n)$ is defined as follows:

$$\mathcal{N}(\mathbb{R}^n) := \big\{ f \in C^\infty(\mathbb{R}^n) \mid \forall i \in \mathbb{N}, \exists M_i > 0 : |f^{(i)}(x)| = O(\|x\|^i) \text{ for } \|x\| \longrightarrow +\infty \big\} \,. \tag{19.25}$$

Property 19.1.30. If $f \in \mathcal{S}(\mathbb{R})$ and $f \in \mathcal{N}(\mathbb{R})$, then $fg \in \mathcal{S}(\mathbb{R})$.

19.2 Convolutions and kernels

Definition 19.2.1 (Direct product). Consider two distributions $\phi \in \mathcal{D}'(U)$ and $\psi \in \mathcal{D}'(V)$. The direct product distribution $\phi \times \psi \in \mathcal{D}'(U \times V)$ is defined by one of the following two equivalent formulas:

$$\langle \phi \times \psi, f \rangle := \langle \phi, \langle \psi, f \rangle \rangle \tag{19.26}$$

or

$$\langle \phi \times \psi, f \rangle := \langle \psi, \langle \phi, f \rangle \rangle \,. \tag{19.27}$$

Definition 19.2.2 (Convolution). The convolution of two distributions is defined as follows (if it exists):

$$\langle \phi * \psi, f \rangle := \langle \phi \times \psi, g \rangle \,, \tag{19.28}$$

where $g(x, y) := f(x + y)$. It should be noted that the convolution is commutative.

Example 19.2.3 (Convolution with delta distribution). For every distribution ϕ, one has the following property:

$$\delta * \phi = \phi \,. \tag{19.29}$$

Formula 19.2.4 (Convolution of functions). The convolution $f * g$ of two (locally integrable) functions can be defined through Example 19.1.7:

$$(f * g)(x) := \int_{\mathbb{R}} f(y) g(x - y) \, dy \,. \tag{19.30}$$

Property 19.2.5 (Young inequality). If $f, g \in L^1$, then $f * g$ exists a.e. and

$$\|f * g\|_1 \leq \|f\|_1 \|g\|_1 . \tag{19.31}$$

This also implies that $f * g \in L^1$. Furthermore, consider p, q and $r \in]0, \infty]$ such that

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{r} + 1 . \tag{19.32}$$

If $f \in L^p$ and $g \in L^q$, then

$$\|f * g\|_r \leq \|f\|_p \|g\|_q . \tag{19.33}$$

This also implies that $f * g \in L^r$. A result similar to Formula 18.3.7 holds for Hölder conjugates ($r = +\infty$). Their convolution is an element of L^∞. Furthermore, the convolution is uniformly continuous on all of $\mathbb{R}^n$ and, if either $p > 1$ or $q > 1$, the convolution vanishes at infinity.

Theorem 19.2.6 (Schwartz's kernel theorem). *There exists an isomorphism*

$$D'(U \times V) \to D'(V, D'(U)) \tag{19.34}$$

given by

$$\langle \phi, f \times g \rangle = \langle T_\phi g, f \rangle . \tag{19.35}$$

@@ COMPLETE (kernels, ...) @@

19.3 Transformations

19.3.1 Fourier series

Definition 19.3.1 (Dirichlet kernel). The Dirichlet kernel is the collection of functions of the form:

$$D_n(x) := \frac{1}{2\pi} \sum_{k=-n}^{n} e^{ikx} \tag{19.36}$$

for $n \in \mathbb{N}$.

Formula 19.3.2 (Sieve property). If $f \in C^1([-\pi, \pi])$, then

$$\lim_{n \to \infty} \int_{-\pi}^{\pi} f(x) D_n(x) \, dx = 0 . \tag{19.37}$$

Formula 19.3.3 (Generalized Fourier series). Let $f \in L^2([-l, l])$ be a $2l$-periodic function. This function can be approximated by the following series:

$$f(x) = \sum_{n=-\infty}^{+\infty} \left(\frac{1}{2l} \int_{-l}^{l} e^{-i\frac{n\pi x'}{l}} f(x') \, dx' \right) e^{i\frac{n\pi x}{l}} . \tag{19.38}$$

Formula 19.3.4 (Fourier coefficients). As seen in the above formula, the Fourier coefficients can be calculated through the inner product (25.12) on L^2:

$$\tilde{f}(k) = \int_{-l}^{l} e_k^*(x) f(x)\, dx, \qquad \text{where} \qquad e_k := \sqrt{\frac{1}{2l}} e^{i\frac{k\pi x}{l}}. \tag{19.39}$$

Formula 19.3.5. For 2π-periodic functions, the order-n Fourier approximation is given by the following convolution:

$$s_n(x) = \sum_{k=-n}^{n} \tilde{f}(k) e^{ikx} = (D_n * f)(x). \tag{19.40}$$

Property 19.3.6 (Convergence of the Fourier series). Let $f : \mathbb{R} \to \mathbb{R}$ be a 2π-periodic function. If f is piecewise C^1 on $[-\pi, \pi]$, then

$$(D_n * f)(x) \xrightarrow{n\to\infty} \frac{f(x+) + f(x-)}{2}. \tag{19.41}$$

Definition 19.3.7 (Periodic extension). Let f be piecewise C^1 on $[-L, L]$. The periodic extension f^L is defined by gluing 'copies' of f together. The **normalized periodic extension** is defined as follows:

$$f^{L,\nu}(x) := \frac{f^L(x^+) + f^L(x^-)}{2}. \tag{19.42}$$

Property 19.3.8. If f is piecewise C^1 on $[-L, L]$, the Fourier series approximation of f converges to $f^{L,\nu}$ on all of $\mathbb{R}$.

19.3.2 Fourier transform

The Fourier series can be used to expand a $2l$-periodic function as an infinite series of exponentials. However, to expand a nonperiodic function $f \in L^1(\mathbb{R})$, one needs the integral Fourier transform.[3]

$$\mathcal{F}f(k) := \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-ikx}\, dx. \tag{19.43}$$

The inverse transform, if it exists, is given by

$$f(x) = \mathcal{F}^{-1}(\mathcal{F}f)(x) = \frac{1}{\sqrt{2\pi}} \mathcal{P} \int_{\mathbb{R}} \mathcal{F}f(k) e^{ikx}\, dk. \tag{19.44}$$

Equation (19.43) is called the (forward) Fourier transform of f and Eq. (19.44) is called the inverse Fourier transform. The pair $(f, \mathcal{F}f)$ is called a **Fourier transform pair**.

Notation 19.3.9. The Fourier transform of a function f is often denoted by $\tilde{f}$ or $\hat{f}$.

[3] All functions are required to be Lebesgue-integrable to make the integrals converge. Weaker conditions are possible (see the literature).

Property 19.3.10. From the Riemann–Lebesgue lemma 18.2.25, it follows that

$$\mathcal{F}f(\omega) \longrightarrow 0 \qquad \text{if} \qquad |\omega| \longrightarrow 0\,. \tag{19.45}$$

Theorem 19.3.11 (Parceval). *Let $(f, \tilde{f})$ and $(g, \tilde{g})$ be two Fourier transform pairs.*

$$\int_{\mathbb{R}} f(x)g(x)\,dx = \int_{\mathbb{R}} \tilde{f}(k)\tilde{g}(k)\,dk \tag{19.46}$$

Corollary 19.3.12 (Plancherel). The integral of the square (of the modulus) of a Fourier transform is equal to the integral of the square (of the modulus) of the original function:

$$\int_{\mathbb{R}} |f(x)|^2\,dx = \int_{\mathbb{R}} |\tilde{f}(k)|^2\,dk\,. \tag{19.47}$$

This implies that the Fourier transform defines an isometry on L^2. In this case, it is often called the **Fourier–Plancherel transform**.

Now, one can wonder why the Fourier transform is introduced in this chapter. The reason is that this transformation can be generalized to distributions in a convenient way. Naively, one could try to extend the definition through duality, but for an arbitrary $\phi \in \mathcal{D}'$ it is not guaranteed that $\mathcal{F}\phi \in \mathcal{D}'$. This is where the Schwartz spaces come in.

Property 19.3.13. The Fourier transform defines an isomorphism on $\mathcal{S}$.

Definition 19.3.14 (Tempered distribution). The space of tempered distributions is the topological dual of the Schwartz space (with its Fréchet topology). It comes equipped with the weak-* topology.

This space has the following important property.

Property 19.3.15. $\mathcal{D}$ is dense in $\mathcal{S}$ and, hence, tempered distributions are determined by their values on $\mathcal{D}$.

Property 19.3.16. The Fourier transform of tempered distributions has some nice additional properties:

- The Fourier transform defines an isomorphism on $\mathcal{S}^*$.
- The Fourier transform of a compactly supported function is of slow growth.
- The Fourier transform of a convolution is equal to the product of the individual Fourier transforms. (Here, one should restrict to the case of a compactly supported and a tempered distribution such that the convolution is also tempered.) This is the **convolution theorem**:

$$\mathcal{F}(\phi * \psi) = \mathcal{F}\phi * \mathcal{F}\psi\,. \tag{19.48}$$

The Fourier transform also induces the following isomorphisms.

Theorem 19.3.17 (Paley–Wiener[4]). *The space of compactly supported distributions of order N is isomorphic to the space of entire functions satisfying*

$$|F(z)| \leq C(1+|z|)^N e^{b\,|\mathrm{im}(z)|} \tag{19.49}$$

for some constants $b, C \in \mathbb{R}^+$ such that the distributions have support on $\overline{B}(0,b)$.

The space of regular, compactly supported distributions is isomorphic to the space of entire functions satisfying

$$\forall N \in \mathbb{N}, \exists C_N \in \mathbb{R}^+ : |F(z)| \leq C_N(1+|z|)^{-N} e^{b\,|\mathrm{im}(z)|}\,, \tag{19.50}$$

for some constant $b \in \mathbb{R}^+$ such that the distributions have support on $\overline{B}(0,b)$.

The Fourier transform 19.43 can be generalized to finite measures.

Definition 19.3.18 (Fourier–Stieltjes transform). Let μ be a finite Borel measure on $\mathbb{R}^n$. The Fourier(–Stieltjes) transform of μ is defined as follows:

$$\hat{\mu}(\omega) := \int_X e^{-2\pi i x\cdot\omega}\, d\mu(x)\,. \tag{19.51}$$

19.3.3 Wave front sets

By the Paley–Wiener–Schwartz theorem 19.3.17, the regularity of a distribution can be characterized in terms of a growth condition on its Fourier transform. This approach can also be used to characterize in which directions one can move without losing regularity of a distribution.

Definition 19.3.19 (Singular fibre). Consider a distribution ϕ on $\mathbb{R}^n$. Its singular fibre $\Sigma_x(\phi)$ at a point $x \in \mathbb{R}^n$ is the complement of the set of vectors $v \in \mathbb{R}^n$ for which its Fourier transform satisfies the Palais–Wiener estimate

$$|\mathcal{F}(\eta\phi)(w)| \leq C_N\big(1+\|w\|\big)^{-N} \tag{19.52}$$

for all vectors $w \in \Gamma$ in a conical neighbourhood (see Definition 25.3.8) of v and all **cut-off functions** η, i.e. functions that are equal to 1 on a compact neighbourhood of x and equal to 0 outside of some larger compact set.

This definition says that the singular fibre of a distribution at some point contains those directions in which the Fourier distribution (localized around the point) becomes singular and, hence, indicates that the function is not smooth along these directions.

Property 19.3.20. The singular fibre is itself a conical subset.

[4]This version, stated in terms of distributions, is actually due to *Schwartz*.

Definition 19.3.21 (Wave front set). Consider a distribution ϕ on $\mathbb{R}^n$. The wave front set of ϕ is defined as follows:

$$\mathsf{WF}(\phi) := \left\{(x, v) \in \mathbb{R}^n \times \mathbb{R}^n \backslash 0 \mid v \in \Sigma_x(f)\right\}. \tag{19.53}$$

The projection of the wave front set onto its first argument recovers the singular support (Definition 19.1.18).

The following property establishes the result hinted at above.

Property 19.3.22 (Regularity). A compactly supported distribution is regular if and only if its wave front set is empty.

Property 19.3.23 (Derivatives). The wave front set of the derivative of a distribution is contained in the wave front set of the distribution.

19.3.4 Laplace transform

Formula 19.3.24 (Laplace transform).

$$\mathcal{L}\{f\}(s) := \int_0^{+\infty} f(t)e^{-st}\,dt \tag{19.54}$$

Formula 19.3.25 (Bromwich integral).

$$f(t) = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} \mathcal{L}\{f\}(s)e^{st}\,ds \tag{19.55}$$

Just as in the case of the Fourier transform (Definition 19.3.18), the Laplace transform can be generalized to finite measures.

Definition 19.3.26 (Laplace–Stieltjes transform). Let μ be a finite Borel measure on $\mathbb{R}$. The Laplace(–Stieltjes) transform of μ is defined as follows:

$$\mathcal{L}\{\mu\}(s) := \int_{\mathbb{R}} e^{-st}\,d\mu(t). \tag{19.56}$$

19.3.5 Integral representations

@@ ADD LINK WITH Gelfan'd transform @@

Formula 19.3.27 (Mellin transform).

$$\mathcal{M}\{f\}(s) := \int_0^{+\infty} x^{s-1} f(x)\,dx \tag{19.57}$$

Formula 19.3.28 (Inverse Mellin transform).

$$f(x) = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} \mathcal{M}\{f\}_{(s)}x^{-s}\,ds \tag{19.58}$$

Formula 19.3.29 (Heaviside step function).

$$\theta(x) = \frac{1}{2\pi i}\int_{\mathbb{R}} \frac{e^{ikx}}{k - i\varepsilon}\, dk \tag{19.59}$$

Formula 19.3.30 (Dirac delta distribution).

$$\delta(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{ikx}\, dk \tag{19.60}$$

19.4 Harmonic analysis ♣

Definition 19.4.1 (Haar measure). A left (resp. right) Haar measure on a topological group is a regular Borel measure (Definition 18.1.11) that is finite on compact subsets and invariant under the left (resp. right) group action. For locally compact groups, this is a left-invariant (resp. right-invariant) Radon measure (Definition 18.1.12).

Example 19.4.2 (Lebesgue measure). Consider $\mathbb{R}^n$ as an additive group. Property 18.1.17 implies that the Lebesgue measure is a left (and right) Haar measure.

Theorem 19.4.3 (Haar[5]). *If G is locally compact, there exists a left Haar measure that is unique up to a scalar factor. Moreover, if G is compact, this constant can be fixed by requiring the normalization condition $\mu(G) = 1$.*

Definition 19.4.4 (Pontryagin dual). Let G be a locally compact Abelian group. Its (Pontryagin) dual is defined as the group of continuous homomorphisms from G to the circle group:

$$G^\vee := \mathsf{Hom}(G, S^1)\,. \tag{19.61}$$

This group is endowed with the compact-open topology (Definition 6.5.21). Elements of this group are called **group characters** of G.

Theorem 19.4.5 (Pontryagin duality). *There exists a natural isomorphism $G \cong G^{\vee\vee}$.*

Construction 19.4.6 (Fourier transform). Consider a locally compact Abelian group G together with its canonical Haar measure μ. For every $f \in L^1(G, \mu)$, one defines the Fourier transform as follows for all $\chi \in G^\vee$:

$$\hat{f}(\chi) := \int_G f(g)\overline{\chi(g)}\, d\mu(g)\,, \tag{19.62}$$

where the identification $S^1 \cong \mathsf{U}(1)$ is used.

Theorem 19.4.7 (Bochner). *Consider a locally compact Abelian group G. There is a bijective correspondence between normalized, positive-definite, continuous functions on G and probability measures on $G^\vee$, where*

$$f(g) = \int_{G^\vee} \chi\, d\nu\,. \tag{19.63}$$

[5] A similar theorem holds for right Haar measures.

@@ COMPLETE @@

Chapter 20

Ordinary differential equations

This chapter is concerned with the study of equations involving univariate functions $f : \mathbb{R} \to \mathbb{R}$ and their derivatives.

Chapter Contents

20.1 Boundary conditions **410**

20.2 Existence and uniqueness **411**

20.3 First-order ODEs **411**

20.4 Second-order ODEs **412**

20.4.1 General solution 412

20.4.2 Constant coefficients 412

20.4.3 Frobenius' method 413

20.5 Fourier transform **414**

20.6 Sturm–Liouville theory **415**

20.7 Applications **416**

20.7.1 Bessel functions 416

20.1 Boundary conditions

Unique solutions of a differential equation are obtained by supplying additional conditions. These are called **boundary conditions**.

Definition 20.1.1 (Periodic boundary condition). A condition of the following form:

$$f(x) = f(x + a)\,. \tag{20.1}$$

Definition 20.1.2 (Dirichlet boundary condition). A condition of the form

$$f(x) = g(x) \tag{20.2}$$

for all $x \in \partial\Omega$, where Ω is the domain on which the problem is defined.

Definition 20.1.3 (Neumann boundary condition). A condition of the form

$$\frac{\partial f}{\partial \hat{n}}(x) = f(x) \tag{20.3}$$

for all $x \in \partial\Omega$, where Ω is the domain on which the problem is defined.

@@ CHECK (these make more sense for PDEs) @@

20.2 Existence and uniqueness

Theorem 20.2.1 (Picard–Lindelöf). *Consider an ordinary differential equation of the form*

$$\dot{f}(x) = g\big(x, f(x)\big)\,, \tag{20.4}$$

where g is defined on a subset $I \times U \subset \mathbb{R} \times \mathbb{R}^n$.[1] *If g is continuous on I and locally Lipschitz on U, then, for every point $(x_0, y_0) \in I \times U$, there exists a maximal interval $J \supseteq I$ and a unique solution $f : J \to \mathbb{R}^n$ of the differential equation with initial condition (x_0, y_0).*

20.3 First-order ODEs

Definition 20.3.1 (First-order ODE).

$$f'(x) + a(x)f(x) = R(x) \tag{20.5}$$

If the function R is identically zero, the ODE is said to be **homogeneous**.

Formula 20.3.2. Let $U \subseteq \mathbb{R}$ be an open set and let the functions $a, R : U \to \mathbb{R}$ be continuous. The solutions $f : U \to \mathbb{R}$ of Eq. (20.5) are given by[2]

$$f(x) = e^{-\int a(x)\,dx}\left(c + \int R(x)e^{\int a(s)\,ds}\,dx\right), \tag{20.6}$$

where $c \in \mathbb{R}$ is a constant (which is, in general, determined by a choice of boundary condition).

[1] Generalizations to arbitrary *Banach spaces* exist (to be introduced in Section 25.1), see e.g. Choquet-Bruhat et al. (1991).

[2] The integrals represent indefinite integrals or primitives.

20.4 Second-order ODEs

Definition 20.4.1 (Second-order ODE).

$$f''(x) + a(x)f'(x) + b(x)f(x) = R(x) \tag{20.7}$$

If the function R is identically zero, the ODE is again said to be **homogeneous**.

20.4.1 General solution

Formula 20.4.2. Let $\varphi : U \to \mathbb{R}$ be a nowhere zero solution of the homogeneous equation. The general solution of Eq. (20.7) is given by

$$f(x) = c_1\varphi(x) + c_2\varphi(x)\int \frac{e^{-\int a(s)\,ds}}{\varphi^2(x)}\,dx + \psi_0(x)\,, \tag{20.8}$$

where ψ_0 is a particular solution of Eq. (20.7).

Property 20.4.3. Let ψ_0 be a particular solution of Eq. (20.7). The set of all solutions is given by the affine space

$$\left\{\psi_0 + \chi \mid \chi \text{ is a solution of the homogeneous equation}\right\}. \tag{20.9}$$

@@ CHECK HOW THIS MATCHES WITH THE PREVIOUS PROPERTY @@

Property 20.4.4 (Wronskian). Two solutions of the homogeneous equation are independent if the **Wronskian**

$$W\big(\varphi_1(x), \varphi_2(x)\big) := \det\begin{pmatrix} \varphi_1(x) & \varphi_2(x) \\ \varphi_1'(x) & \varphi_2'(x) \end{pmatrix} \tag{20.10}$$

is nonzero.

Formula 20.4.5 (Abel's identity). An explicit formula for the Wronskian is given by

$$W(x) = W(x_0)\exp\left(-\int_{x_0}^{x} a(x')\,dx'\right). \tag{20.11}$$

20.4.2 Constant coefficients

Property 20.4.6. A function $\varphi : U \to \mathbb{C}$ is a complex solution of the homogeneous equation if and only if $\Re(\varphi)$ and $\Im(\varphi)$ are real solutions of the homogeneous equation.

Formula 20.4.7 (Characteristic equation). When studying an ODE of the form[3]

$$f''(x) + pf'(x) + qf(x) = 0\,, \tag{20.12}$$

[3] Any other form of homogeneous, second-order ODEs with constant coefficients can be rewritten like this.

where $p, q \in \mathbb{R}$ are constants, the characteristic equation is defined as follows:

$$\lambda^2 + p\lambda + q = 0\,. \tag{20.13}$$

By the Fundamental Theorem of Algebra 2.5.7, this polynomial equation has two (complex) roots λ_1 and λ_2. From these roots, one can derive the solutions of Eq. (20.12) using the following rules ($c_1, c_2 \in \mathbb{R}$ are constants):

- $\lambda_1 \neq \lambda_2$ with $\lambda_1, \lambda_2 \in \mathbb{R}$: $f(x) = c_1 e^{\lambda_1 x} + c_2 e^{\lambda_2 x}$,
- $\lambda_1 = \lambda_2$: $f(x) = c_1 e^{\lambda x} + c_2 x e^{\lambda x}$, and
- $\lambda_1 = \lambda_2^*$ with $\lambda_1 = a + ib$: $f(x) = c_1 e^{ax} \cos(bx) + c_2 e^{ax} \sin(bx)$.

20.4.3 Frobenius' method

To find a solution of the homogeneous equation, one can assume a solution of the form

$$y(x) = \sum_{i=0}^{+\infty} a_i (x - x_0)^{i+k}\,, \tag{20.14}$$

where $k \in \mathbb{R}$ is a constant.

Definition 20.4.8 (Indicial equation). After inserting the ansatz (20.14) into the homogeneous equation and collecting all terms in x^i, an equation of the form

$$\sum_{i=n}^{+\infty} H_i(k) x^i = 0 \tag{20.15}$$

is obtained, where $n \in \mathbb{N}$ and the $H_i(k)$ are polynomials in k. This means that for every $i \in \mathbb{N}$, one obtains an equation of the form $H_i(k) = 0$, due to the independence of polynomial terms. The equation for the lowest-degree term will be quadratic in k and is called the indicial equation.

Property 20.4.9. The indicial equation generally has two roots k_1, k_2.

- $k_1 = k_2$: Only one solution will be found with Frobenius' method (another one can be found as in the second term of Formula 20.4.2).
- $k_1 - k_2 \in \mathbb{Z}$: A second independent solution might be obtained using this method. If not, a second solution can be found as mentioned in the previous case.
- $k_1 - k_2 \notin \mathbb{Z}$: Two independent solutions can be found using this method.

Theorem 20.4.10 (Fuchs). *If the functions $a, b : U \to \mathbb{R}$ in Eq. (20.7) are analytic at $x = x_0$, the general solution can be expressed as a Frobenius series.*

20.5 Fourier transform

Recall Section 19.3.2. The Fourier transform gives well-defined maps on $L^1, L^2, \mathcal{S}$ and $\mathcal{S}^*$.

Consider an ODE of the form

$$\sum_{l=0}^{n} a_l f^{(l)}(x) = g(x)\,. \tag{20.16}$$

Note that the coefficients of the differential operator are assumed to be constants. Expanding both f and g into Fourier components, gives an algebraic equation:

$$\sum_{l=0}^{n} a_l (ik)^l \tilde{f}(k) = \tilde{g}(k)\,. \tag{20.17}$$

Solving this equation, together with any provided initial value conditions (possibly also Fourier transformed) will give conditions on k and $\tilde{f}(k)$.

Example 20.5.1 (Sines and cosines). Consider the ODE

$$f''(x) = -f(x)\,.$$

It is a standard result that the solution space to this equation is spanned by $\sin(x)$ and $\cos(x)$ or, equivalenty, by $\exp(ix)$ and $\exp(-ix)$.

Now, taking the Fourier transform, one obtains:

$$\int_{-\infty}^{+\infty} (1-k^2)\tilde{f}(k)e^{ikx}\,dk = 0\,.$$

This implies that

$$(1-k^2)\tilde{f}(k) = 0 \qquad (\text{in } \mathcal{S}^*)\,.$$

This implies that $\tilde{f}(k) = \lambda\delta(k-1) + \nu\delta(k+1)$ for some $\lambda, \nu \in \mathbb{C}$. Inserting this into the Fourier transform gives:

$$f(x) = \lambda \exp(ix) + \nu \exp(-ix) \tag{20.18}$$

as expected.

Example 20.5.2 (Affine functions). Now, consider the even simpler equation

$$f''(x) = 0\,.$$

Taking the Fourier transform gives:

$$k^2 \tilde{f}(k) = 0\,. \tag{20.19}$$

For ordinary functions, this would mean either $\tilde{f}(k)$ identically zero or $k = 0$. However, passing to distributions (and Formula 19.1.26), one obtains

$$\tilde{f}(k) = \lambda\delta'(k) + \kappa\delta(k)$$

or, after transforming back to the x-domain:

$$f(x) = \lambda x + \kappa\,.$$

In general, combining the two examples above, starting from a homogeneous ODE (with constant coefficients), the resulting Fourier components will be delta distributions (and their derivatives). Transforming back to the x-domain, this will lead to a combination of polynomials (for $k = 0$) and oscillatory components (for $k \neq 0$). If the ODE is not homogeneous, Formula 19.1.26 is not applicable and a more general (distributional) equation has to be solved.

20.6 Sturm–Liouville theory

Definition 20.6.1 (Sturm–Liouville problem). An ODE of the form

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(p(x)\frac{\mathrm{d}y}{\mathrm{d}x}\right) + \big(g(x) + \lambda r(x)\big)y(x) = 0\,, \tag{20.20}$$

subject to mixed boundary conditions, where:

- $p, q, r : [a, b] \to \mathbb{R}$ are continuous,
- $p \in C^1([a, b])$ with either $p < 0$ or $p > 0$,
- either $r \geq 0$ or $r \leq 0$, and
- r is not identically zero on any subinterval.

The boundary conditions are given by:

$$\begin{aligned} \alpha_1 y(a) + \beta_1 y'(a) &= 0\,, \\ \alpha_2 y(b) + \beta_2 y'(b) &= 0\,, \end{aligned}$$

where at least one of the constants $\alpha_1, \alpha_2, \beta_1$ or β_2 is nonzero.

Formula 20.6.2. The solutions of a Sturm–Liouville problem are of the form

$$y(x) = c_1 u_1(x;\lambda) + c_2 u_2(x;\lambda)\,. \tag{20.21}$$

Only for certain values of λ will these solutions be nontrivial. The values of λ for which the solutions are nontrivial are called **eigenvalues** and the associated solutions are

called **eigenfunctions**. Substituting this form in the boundary conditions gives the following determinant condition for nontrivial solutions:

$$\det\begin{pmatrix} \alpha_1 u_1(a;\lambda) + \beta_1 u_1'(a;\lambda) & \alpha_1 u_2(a;\lambda) + \beta_1 u_2'(a;\lambda) \\ \alpha_1 u_1(b;\lambda) + \beta_1 u_1'(b;\lambda) & \alpha_1 u_2(b;\lambda) + \beta_1 u_2'(b;\lambda) \end{pmatrix} = 0\,. \tag{20.22}$$

The independent eigenfunctions can be found by substituting the solutions λ of this condition in the ODE (20.20).

Property 20.6.3 (Self-adjoint form). A Sturm–Liouville problem can be rewritten as follows:[4]

$$\hat{\mathcal{L}}y(x) = \lambda y(x)\,. \tag{20.23}$$

The operator

$$\hat{\mathcal{L}} = -\frac{1}{r(x)}\left[\frac{\mathrm{d}}{\mathrm{d}x}\left(p(x)\frac{\mathrm{d}}{\mathrm{d}x}\right) + g(x)\right] \tag{20.24}$$

is called the **self-adjoint form**, since $\hat{\mathcal{L}}$ is a self-adjoint operator. Now, consider the following general linear ODE

$$\left(a_2(x)\frac{\mathrm{d}^2}{\mathrm{d}x^2} + a_1(x)\frac{\mathrm{d}}{\mathrm{d}x} + a_0(x)\right)y(x) = 0\,. \tag{20.25}$$

This equation can be rewritten in a self-adjoint form by setting

$$p(x) := \exp\left(\int \frac{a_1(x)}{a_2(x)}\,dx\right) \qquad \text{and} \qquad g(x) := \frac{a_0(x)}{a_2(x)}\exp\left(\int \frac{a_1(x)}{a_2(x)}\,dx\right).$$

Property 20.6.4. The eigenfunctions corresponding to distinct eigenvalues are orthogonal with respect to the weight function $r : [a,b] \to \mathbb{R}$. This can be seen as an instance of Property 22.5.17.

Theorem 20.6.5 (Oscillation theorem). *The nth eigenfunction of a Sturm–Liouville problem has $n-1$ roots.*

20.7 Applications

20.7.1 Bessel functions

A Bessel's differential equation is an ordinary differential equation of the following form:

$$x^2 f''(x) + x f'(x) + (x^2 - n^2) f(x) = 0\,, \tag{20.26}$$

[4]This formulation explains the name 'eigenvalue' for the quantity λ.

where n can actually be any complex number. The solutions of this ODE are the Bessel functions of the first and second kind (also called **Bessel** and **Neumann functions**, respectively):

$$J_n(x) = \sum_{m=0}^{+\infty} \frac{(-1)^m}{m!(m+n)!} \left(\frac{x}{2}\right)^{2m+n}, \tag{20.27}$$

$$N_n(x) = \lim_{\nu \to n} \frac{\cos(\nu\pi) J_n(x) - J_{-n}(x)}{\sin(\nu\pi)}. \tag{20.28}$$

Remark. Solution (20.27) can be found using the Frobenius method.

Property 20.7.1. For $n \notin \mathbb{N}$, the solutions J_n and J_{-n} are independent.

Remark 20.7.2. For $n \notin \mathbb{N}$, the limit in Eq. (20.28) is not necessary since $\sin(n\pi)$ will never become 0 in this case.

Formula 20.7.3 (Generating function). Consider the following function:

$$g(x,t) := \exp\left[\frac{x}{2}\left(t - \frac{1}{t}\right)\right]. \tag{20.29}$$

If this function is expanded as a Laurent series, an expression of the form

$$g(x,t) = \sum_{n=-\infty}^{+\infty} J_n(x) t^n \tag{20.30}$$

is obtained. By applying the Residue Theorem 17.5.22, one can express the functions J_n as follows:

$$J_n(x) = \frac{1}{2\pi i} \oint_C \frac{g(x,t)}{t^{n+1}}\, dt. \tag{20.31}$$

One can show that these functions are exactly the Bessel functions (20.27). For this reason, $g(x,t)$ is called the generating function of the Bessel functions.

@@ SHOW THAT THESE ARE REALLY THE BESSEL FUNCTIONS @@

Chapter 21

Partial differential equations

For a rigorous treatment of partial differential equations, the language of distributions is required. For an introduction, see Chapter 19.

Chapter Contents

21.1 Important examples **418**
21.2 General linear equations **419**
21.3 First-order PDE **420**
21.4 Second-order PDE **421**
21.5 General methods **423**
21.5.1 Separation of variables 423
21.6 Boundary conditions **423**
21.6.1 Dirichlet problem 426
21.7 General **426**
21.7.1 Symbols 426
21.8 Sobolev spaces **427**

21.1 Important examples

Example 21.1.1 (Laplace equation).

$$\Delta f = \sum_{i=1}^{d} \frac{\partial^2 f}{\partial {x_i}^2} = 0 \tag{21.1}$$

Functions satisyfing this equation are said to be **harmonic**.

When solving the Laplace equation in cylindrical coordinates, one obtains Bessel's ODE (20.26) with integer n. The solutions are the Bessel functions (20.27) and (20.28). For this reason, Bessel functions for integer n are also called **cylinder functions** or **cylindrical harmonics**.

Example 21.1.2 (Helmholtz equation).

$$\Delta f = -k^2 f \tag{21.2}$$

When solving the Helmholtz equation in spherical coordinates, one obtains a variant of the Bessel ODE (20.26) for the radial part:

$$r^2 y''(r) + 2ry'(r) + \big(r^2 - n(n+1)\big)y(r) = 0\,, \tag{21.3}$$

where n is an integer. The solutions, called the **spherical Bessel functions**, are related to the cylindrical Bessel functions in the following way (similarly for the Neumann functions):

$$j_n(r) = \sqrt{\frac{\pi}{2x}} J_{n+\frac{1}{2}}(r)\,. \tag{21.4}$$

Formula 21.1.3 (Wave equation). In one spatial dimension, the wave equation reads as follows:

$$\frac{\partial^2 f}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 f}{\partial t^2}\,. \tag{21.5}$$

In higher dimensions, this can be rewritten using the Laplacian (Eq. (23.9)) as

$$\Delta f = \frac{1}{v^2}\frac{\partial^2 f}{\partial t^2}\,. \tag{21.6}$$

or, using the **d'Alembertian**

$$\Box_v := \frac{1}{v^2}\frac{\partial^2}{\partial t^2} - \Delta\,, \tag{21.7}$$

as

$$\Box_v f = 0\,. \tag{21.8}$$

@@ ADD (Poisson, heat, ...) @@

21.2 General linear equations

Definition 21.2.1 (Characteristic curve). A curve along which the highest-order partial derivatives are discontinuous. Equivalently, a curve along which the partial differential equation reduces to an ordinary differential equation (see the next sections).

@@ COMPLETE or RESTRUCTURE @@

21.3 First-order PDE

Formula 21.3.1 (First-order quasilinear PDE).

$$P(x,y,u)\frac{\partial u}{\partial x} + Q(x,y,u)\frac{\partial u}{\partial y} = R(x,y,u) \tag{21.9}$$

Formula 21.3.2 (Characteristic curve). Consider a first-order quasilinear PDE as above. The idea of the method of the characteristics is to obtain a new coordinate system (s,t) such that the PDE becomes an ODE in the coordinate s and, accordingly, that the solution has an arbitrary dependence on t. The curves of constant t are called the **characteristics** of the PDE.

To this end, the coefficients of the PDE are interpreted as the coefficients of a vector field $X \equiv (P,Q,R) : \mathbb{R}^3 \to \mathbb{R}^3$. By Property 23.1.2, it can be shown that this vector field is tangent to the surface $u(x,y)$. The characteristic curves then correspond to the *integral curves* of X (see Definition 34.3.11), which foliate u. Along a general curve, one obtains two linear equations for the partial derivatives $u_{,x}, u_{,y}$, expressed as the following *Pfaffian system*[1]:

$$\begin{cases} P\frac{\partial u}{\partial x} + Q\frac{\partial u}{\partial y} = R\,, \\ \frac{\partial u}{\partial x}\mathrm{d}x + \frac{\partial u}{\partial y}\mathrm{d}y = \mathrm{d}u\,. \end{cases} \tag{21.10}$$

For characteristics, the derivatives with respect to the new coordinate s, should correspond with the coefficients of X, i.e. the coefficients of the PDE. It follows that the above system should degenerate. Algebraically, this gives the following condition:

$$\det\begin{pmatrix} P & Q \\ \mathrm{d}x & \mathrm{d}y \end{pmatrix} = 0\,. \tag{21.11}$$

By Cramer's rule 22.4.15, the existence of a (nonunique) solution also requires that

$$\det\begin{pmatrix} P & R \\ \mathrm{d}x & \mathrm{d}u \end{pmatrix} = 0\,. \tag{21.12}$$

The characteristic curves are thus defined by

$$\frac{\mathrm{d}x}{P} = \frac{\mathrm{d}y}{Q} \tag{21.13}$$

and along these curves the condition

$$\frac{\mathrm{d}x}{P} = \frac{\mathrm{d}u}{R} \tag{21.14}$$

should hold as a consistency condition. The solution of the ODEs for the integral curves are given by the Picard–Lindelöf theorem 20.2.1, provided that initial conditions are known.

[1] Formalized in Property 34.6.8.

Formula 21.3.3 (Lagrange–Charpit equations). The general solution of Formula 21.3.1 is implicitly given by $F(\xi, \eta) = 0$, where $F(\xi, \eta)$ is an arbitrary differentiable function and $\xi(x, y, u) = c_1, \eta(x, y, u) = c_2$ are solutions of the Lagrange–Charpit equations

$$\frac{dx}{P} = \frac{dy}{Q} = \frac{du}{R}, \tag{21.15}$$

where c_1, c_2 are constants fixed by the boundary conditions.

Remark 21.3.4. Looking at the defining equations of the characteristic curve, it is clear that these fix the general solution of the PDE.

21.4 Second-order PDE

Formula 21.4.1 (Second order quasilinear PDE). Consider the following pseudolinear differential equation for a function $u : \mathbb{R}^2 \to \mathbb{R}$:

$$R(x, y)u_{xx} + S(x, y)u_{xy} + T(x, y)u_{yy} = W(x, y, u, p, q), \tag{21.16}$$

where $p := u_x$ and $q := u_y$.

Formula 21.4.2 (Characteristic equation). Similar to the case of first-order PDEs, characteristic curves are characterized by the following condition:

$$\det \begin{pmatrix} R & S & T \\ \mathrm{d}x & \mathrm{d}y & 0 \\ 0 & \mathrm{d}x & \mathrm{d}y \end{pmatrix} = 0. \tag{21.17}$$

This is equivalent to the following equation:

$$R\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2 - S\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right) + T = 0. \tag{21.18}$$

Accordingly, this equation is often called the **characteristic equation** of the PDE in Formula 21.4.1. The PDE has been reduced to an ODE (as before, Cramer's rule gives rise to additional ODEs).

Definition 21.4.3 (Types of characteristics). Equation (21.18) is quadratic in $\frac{\mathrm{d}y}{\mathrm{d}x}$. If this equation has two distinct real roots, the PDE is said to be **hyperbolic**. If the equation has only one root, the PDE is said to be **parabolic**. In the remaining case, where the equation has two distinct complex roots, the PDE is said to be **elliptic**.

Formula 21.4.4 (Canonical form). Consider the general change of variables

$$\xi = \xi(x, y) \qquad \eta = \eta(x, y) \qquad \zeta \equiv u.$$

After this transformation, the PDE in Formula 21.4.1 becomes

$$A(\xi_x, \xi_y)\frac{\partial^2 \zeta}{\partial \xi^2} + B(\xi_x, \xi_y, \eta_x, \eta_y)\frac{\partial^2 \zeta}{\partial \xi \partial \eta} + A(\eta_x, \eta_y)\frac{\partial^2 \zeta}{\partial \eta^2} = F(\xi, \eta, \zeta, \zeta_\xi, \zeta_\eta), \tag{21.19}$$

where:

- $A(a,b) = Ra^2 + Sab + Tb^2$, and
- $B(a,b,c,d) = 2Rac + S(bc + ad) + 2Tbd$.

The discriminant Δ of the quadratic equation (21.18) allows to rephrase the classification of characteristics in terms of canonical forms. The fact that this classification is well-defined follows from the result that the discriminant of Eq. (21.18) is, up to the square of the Jacobian of $(x,y) \longrightarrow (\xi,\eta)$, equal to $B(\xi_x,\xi_y,\eta_x,\eta_y)^2 - 4A(\xi_x,\xi_y)A(\eta_x,\eta_y)$.

- **Hyperbolic PDE** ($\Delta > 0$): The sign of the discriminant implies that the quadratic equation $A = 0$ has two real solutions $f_1(x,y)$ and $f_2(x,y)$. By choosing the transformation $\xi = f_1(x,y)$ and $\eta = f_2(x,y)$, the coefficients $A(a,b)$ are made to vanish and, hence, the canonical hyperbolic form is obtained:

$$\frac{\partial^2 \zeta}{\partial \xi \partial \eta} = H(\xi,\eta,\zeta,\zeta_\xi,\zeta_\eta)\,, \tag{21.20}$$

  where $H := \dfrac{F}{2B(\xi_x,\xi_y\eta_x\eta_y)}$.

- **Parabolic PDE** ($\Delta = 0$): As in the hyperbolic case the change of variables $\xi = f(x,y)$ is performed. However, there is only one root of the defining equation, so the second variable can be chosen freely under the constraint that it should be independent of $f_1(x,y)$. From the condition $\Delta = 0$ it is also possible to derive the conditions that $B(\xi_x,\xi_y\eta_x\eta_y) = 0$ and $A(\eta_x,\eta_y) \neq 0$. This gives the parabolic canonical form:

$$\frac{\partial^2 \zeta}{\partial \eta^2} = G(\xi,\eta,\zeta,\zeta_\xi,\zeta_\eta)\,, \tag{21.21}$$

  where $G := \dfrac{F}{A(\eta_x,\eta_y)}$.

- **Elliptic PDE** ($\Delta < 0$): In this case there are two complex roots. Writing $\xi = \alpha + i\beta$ and $\eta = \alpha - i\beta$ gives the following (real) equation:

$$\frac{\partial^2 \zeta}{\partial \xi \partial \eta} = \frac{1}{4}\left(\frac{\partial^2 \zeta}{\partial \alpha^2} + \frac{\partial^2 \zeta}{\partial \beta^2}\right).$$

  Substituting this in the PDE (together with $A(a,b) = 0$) results in the following elliptic canonical form:

$$\frac{\partial^2 \zeta}{\partial \alpha^2} + \frac{\partial^2 \zeta}{\partial \beta^2} = K(\alpha,\beta,\zeta,\zeta_\alpha,\zeta_\beta)\,. \tag{21.22}$$

Theorem 21.4.5 (Maximum principle). *Consider a PDE of the parabolic or elliptic type. The maximum of the solution is to be found on the boundary of the domain.*

21.5 General methods

21.5.1 Separation of variables

Remark. Solutions obtained by separation of variables are generalized Fourier series, which tend to converge rather slowly. For numerical purposes, other techniques are recommended. However, the series solutions often give a good insight in the properties of the solutions.

Method 21.5.1 (Separation of variables). Let D be the operator associated with a partial differential equation such that $Du(x) = 0$, where $x := (x_1, \dots, x_n)$ denotes the set of variables. A useful method to find solutions is to assume a solution of the form

$$u(x) = \prod_{i=1}^{n} u_i(x_i)\,. \tag{21.23}$$

By substituting this form in the PDE and using some (basic) algebra it is sometimes possible to reduce the partial differential equation to a system of n ordinary differential equations.

Example 21.5.2. Consider the following PDE:

$$\frac{\partial u}{\partial t} - a\frac{\partial^2 u}{\partial x^2} = 0\,. \tag{21.24}$$

Substituting a solution of the form $u(x,t) = X(x)T(t)$ gives

$$X(x)\frac{\mathrm{d}T(t)}{\mathrm{d}t} - aT(t)\frac{\mathrm{d}^2X(x)}{\mathrm{d}x^2} = 0\,,$$

which can be rewritten as (the arguments are dropped for convenience)

$$\frac{1}{aT}\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{1}{X}\frac{\mathrm{d}^2X}{\mathrm{d}x^2}\,.$$

Because both sides are independent, they must be equal to a constant λ. This results in the following system of ordinary differential equations:

$$\begin{cases} X''(x) &= \lambda X(x)\,, \\ T'(t) &= a\lambda T(t)\,. \end{cases} \tag{21.25}$$

21.6 Boundary conditions

Formula 21.6.1 (Inhomogeneous boundary condition).

$$\alpha u(a,t) + \beta\frac{\partial u}{\partial x}(a,t) = h(t) \tag{21.26}$$

When h is identically zero, the boundary condition becomes **homogeneous**.

Method 21.6.2 (Steady-state solution). Assume that the function h is constant. In this case, it is useful to rewrite the solution as

$$u(x,t) = v(x) + w(x,t)\,. \tag{21.27}$$

The 'time-independent' function v is called the steady-state solution and the function w represents the deviation of this steady-state scenario.

Because the PDE is linear, the partial solutions v and w are required to individually satisfy the equation. Furthermore, the function v is also required too satisfy the given inhomogeneous boundary conditions. This results in w being the solution of a homogeneous PDE with homogeneous boundary conditions. This can be seen in the following proof.

Proof. Assume a boundary condition of the form

$$\alpha u(a,t) + \beta \frac{\partial u}{\partial x}(a,t) = u_0\,.$$

Due to the requirements of a steady-state solution, one also has

$$\alpha v(a) + \beta \frac{\partial v}{\partial x}(a) = u_0\,.$$

Combining these two conditions gives

$$\alpha[v(a) + w(a,t)] + \beta\left[\frac{\partial v}{\partial x}(a) + \frac{\partial w}{\partial x}(a,t)\right] = \alpha v(a) + \beta \frac{\partial v}{\partial x}(a)\,.$$

Using the conditions, this can be rewritten as

$$\alpha w(a,t) + \beta \frac{\partial w}{\partial x}(a,t) = 0\,.$$

The steady-state deviation $w(x,t)$ thus satisfies the homogeneous boundary conditions.□

@@ CHECK (is this proof relevant) @@

Method 21.6.3. If the function h is not a constant, a different method can be used. Rewrite the solution as $u(x,t) = v(x,t) + w(x,t)$, where v is only required to be some function that satisfies the boundary conditions (and not the PDE)[2]. This will lead to w satisfying the homogeneous boundary conditions as in the previous method. After substituting the function v in the PDE, a differential equation for $w(x,t)$ is obtained (it can be inhomogeneous).

A third, sometimes useful method is the following.

[2] As there are infinitely many possible functions that satisfy the boundary conditions, the best choice for v is the one that makes the equation for w as simple as possible.

Method 21.6.4. If the problem consists of three homogeneous and one inhomogeneous boundary conditions, the problem can be solved by first using the homogeneous conditions to restrict the values of the separation constant and then obtain a series expansion. Afterwards, the obtained series can be matched to the inhomogeneous condition to obtain the final remaining coefficients.

If there is more than one inhomogeneous boundary condition, the method can be extended. Let there be k boundary conditions. Rewrite the general solution as $u(x,t) = \sum_{i=1}^{k} v_i(x,t)$, where v_i satisfies the ith inhomogeneous condition and the homogeneous versions of the other conditions. This way, the general solution still satisfies all conditions and the first part of the method can be applied to all functions v_i to obtain a series expansion.

Method 21.6.5 (Inhomogeneous PDE). A possible way to solve inhomogeneous second order partial differential equations of the form

$$Du(x,t) = f(x,t)\,, \tag{21.28}$$

given a set of homogeneous boundary conditions and initial value conditions $w(x,0) = \psi(x)$, is the following method (where all involved functions are assumed to admit a generalized Fourier expansion):

1. Solve the homogeneous version of the PDE. This will result in a series expansion

$$\sum_i w_i(t) e_i(x)\,, \tag{21.29}$$

   where the e_i are a complete set of eigenfunctions in the variable x. This solution should satisfy the (homogeneous[3]) boundary conditions.

2. Expand the function f in the same way as u. The coefficients f_n can be found using the orthogonality relations of the functions e_n.

3. Inserting these expansions in the original PDE and rewriting the equation will lead to a summation of the form

$$\sum_i \big(\widetilde{D} w_i(t)\big) e_i(x) = 0\,, \tag{21.30}$$

   where $\widetilde{D}$ is a linear first-order differential operator. Because all terms are independent, this gives n first order ODEs for the functions w_i. These can generally be solved by using Formula 20.3.2.

4. Initial value conditions for the functions w_i are applied by setting $t = 0$ in the series expansion of u and equating it with the series expansion of ψ. This results in conditions $w_i(0) = \Psi_i$.

[3] Inhomogeneous boundary conditions can be turned into homogeneous ones by the previous two methods.

5. The obtained ODEs, together with the found boundary conditions $w_i(0) = \Psi_i$, will give the general solutions for w_i.

6. Inserting these solutions in the series expansion of u will give the general solution of the inhomogeneous PDE.

Remark 21.6.6. The requirement that all involved functions admit a generalized Fourier expansion is restrictive. Not all inhomogeneous PDEs are solvable by this method.

21.6.1 Dirichlet problem

The (interior) Dirichlet problem is the problem of finding a solution to a PDE in a finite region, given the value of the function on the boundary of the region, i.e. given boundary conditions of the form $u|_{\partial\Omega} = 0$. The uniqueness of a solution can be proven with the Maximum Principle 21.4.5 if the PDE is of the elliptic kind (such as the Laplace equation).

Proof. Let ϕ, ψ be two solutions of the interior Dirichlet problem. Due to linearity both $\psi - \phi$ and $\phi - \psi$ are solutions too (without applying the boundary conditions). According to the maximum principle, these solutions achieve their maximum on the boundary of the domain. Furthermore, due to the Dirichlet boundary conditions, $\phi(x) = \psi(x)$ for all $x \in \partial\Omega$. Combining these two facts gives $\max(\psi - \phi) = \max(\phi - \psi) = 0$ or alternatively $\psi \leq \phi$ and $\phi \leq \psi$ on the whole domain. This implies that $\phi = \psi$ on the whole domain. □

Remark. There is also an exterior Dirichlet problem, where one has to find the solution of the PDE, given the boundary conditions, outside of the boundary.

Definition 21.6.7 (Green's function). A fundamental solution (Definition 19.1.13) of a Dirichlet problem.

21.7 General

21.7.1 Symbols

Definition 21.7.1 (Symbol). Consider a general k^{th}-order differential operator (multi-indices α are used)

$$D := \sum_{|\alpha| \leq k} c_\alpha(x) D^\alpha \,. \tag{21.31}$$

The symbol of this operator is defined by replacing the partial derivatives by indeterminates ξ^i:

$$p(D, \xi) := \sum_{|\alpha| \leq k} c_\alpha(x) \xi^\alpha \,. \tag{21.32}$$

Definition 21.7.2 (Principal symbol). The principal symbol of a kth-order differential operator D is defined as the highest-degree component of $p(D, \xi)$:

$$\sigma_D(\xi) := \sum_{|\alpha|=k} c_\alpha(x)\xi^\alpha \,. \tag{21.33}$$

For a system of partial differential equations, the functions c_α are replaced by matrix-valued functions $\left(c_i^j\right)_\alpha$.

Property 21.7.3. The principal symbol of a differential operator transforms as a tensor.

Definition 21.7.4 (Ellipticity). A system of PDEs $Df(x) = 0$ is elliptic if and only if σ_D is invertible. Note that this is only possible if the number of variables is smaller than the number of equations, hence if the system is at most determined.

21.8 Sobolev spaces

Using the theory of L^p-spaces and distributions (Chapter 18 and Chapter 19), one can define an important class of function spaces that are ubiquitous in the field of PDEs (and beyond).

Definition 21.8.1 (Sobolev space). For all nonnegative integers $m, p \in \mathbb{N}$, with $p \geq 1$, one defines the Sobolev space $W^{m,p}(U)$ as the space of functions in $L^p(U)$ for which the weak derivatives (Definition 19.1.10) up to order m are also in $L^p(U)$. When $p = 2$, i.e. when restricted to square-integrable functions, the notation $H^m(U)$ is often used.

This space can be turned into a normed space by equipping it with the following norm:

$$\|f\| := \left(\sum_{|\alpha|\leq m} \|f^{(\alpha)}\|_{L^p}\right)^{1/p} . \tag{21.34}$$

Using the fact that the Fourier transform $\mathcal{F}$ defines an isometry on L^2, one can also define H^m in a different way:

$$H^m(\mathbb{R}^n) := \left\{ f \in L^2(\mathbb{R}^n) \,|\, (1 + \|x\|^2)^{-m/2}\mathcal{F}f \in L^2(\mathbb{R}^2) \right\} . \tag{21.35}$$

Sobolev spaces inherit the following property from the L^p-spaces.

Property 21.8.2 (Completeness). Every Sobolev space is a Banach space. Moreover, one can show that the spaces H^m are Hilbert spaces.

Sobolev spaces also satisfy the following density theorem.

Property 21.8.3. $\mathcal{D}(\mathbb{R}^n)$ is dense in $W^{m,p}(\mathbb{R}^n)$ for all $m \in \mathbb{N}$. However, only for $m = 0$ can this be proven for open subsets of $\mathbb{R}^n$.

Property 21.8.4 (Sobolev embedding). Consider two integers $m, n \in \mathbb{N}$. If $f \in H^m(\mathbb{R}^n)$ and $m > n/2$, then f vanishes at infinity.

The Sobolev norm is not always easy to work with, especially in practical applications. Luckily there exists a lemma showing that one can equivalently restrict to partial derivatives of order m.

Theorem 21.8.5 (Friedrich). *For all bounded U, one can introduce an equivalent norm on $H_0^m(U)$ as follows:*

$$\langle f \mid g \rangle := \sum_{|\alpha|=m} \int_U f^{(\alpha)} \overline{g^{(\alpha)}} \, dx \,. \tag{21.36}$$

Property 18.3.10 allows to define the dual Sobolev spaces.

Definition 21.8.6. The space $W^{-m,p}(U) \subset \mathcal{D}'(U)$ is defined as the dual of $\overline{W^{m,p}(U)}$. For $m = 2$, one can again use a characterization similar to Eq. (21.35). It can be shown that all elements in $W^{-m,p}(U)$ are of the form

$$T = \sum_{|\alpha| \leq m} f_\alpha^{(\alpha)} \,, \tag{21.37}$$

where $f_\alpha \in L^{p'}(U)$ with p' the Hölder conjugate of p.

@@ COMPLETE (continue in AMP1) @@

Part V

Linear Algebra

Table of Contents

22 Linear Algebra **433**

22.1 Vector spaces . . . 434

22.1.1 Linear independence . . . 434

22.1.2 Bases . . . 435

22.1.3 Sum and direct sum . . . 436

22.2 Linear maps . . . 437

22.2.1 Homomorphisms . . . 437

22.2.2 Dual maps . . . 438

22.3 Inner product . . . 439

22.3.1 Inner product space . . . 440

22.3.2 Orthogonality . . . 441

22.4 Matrices . . . 443

22.4.1 System of equations . . . 444

22.4.2 Coordinates and matrix representations . . . 444

22.4.3 Coordinate transformations . . . 446

22.4.4 Determinant . . . 448

22.4.5 Characteristic polynomial . . . 449

22.4.6 Matrix groups . . . 450

22.4.7 Matrix decompositions . . . 451

22.5 Eigenvectors . . . 452

22.5.1 Diagonalization . . . 453

22.5.2 Multiplicity . . . 453

22.6 Euclidean space . . . 454

22.7 Algebras . . . 455

22.8 Grassmanians . . . 458

23 Vector & Tensor Calculus **459**

23.1 Nabla-operator . . . 460

23.2 Integration . . . 462

23.2.1 Line integrals . . . 462

23.2.2 Integral theorems . . . 463

23.3 Tensors . . . 463
23.3.1 Tensor product . . . 463
23.3.2 Transformation rules . . . 465
23.3.3 Tensor operations . . . 466
23.4 Exterior algebra . . . 467
23.4.1 Antisymmetric tensors . . . 467
23.4.2 Determinant . . . 468
23.4.3 Levi-Civita symbol . . . 468
23.4.4 Wedge product . . . 469
23.4.5 Exterior algebra . . . 470
23.4.6 Hodge star . . . 472
23.5 Graded vector spaces . . . 474
23.5.1 Supermatrices . . . 477
23.5.2 Berezin calculus . . . 478

24 Representation Theory **481**
24.1 Group representations . . . 481
24.2 Irreducible representations . . . 482
24.3 Classification . . . 484
24.4 Generalized representations . . . 486
24.4.1 Projective representations . . . 486
24.4.2 Group cohomology . . . 487
24.5 Tensor operators . . . 487

25 Functional Analysis **489**
25.1 Banach spaces . . . 490
25.1.1 Convex duality . . . 491
25.2 Hilbert spaces . . . 492
25.2.1 General . . . 492
25.2.2 Generalized Fourier series . . . 494
25.2.3 Orthogonality and projections . . . 495
25.2.4 Separable Hilbert spaces . . . 495
25.3 Seminorms . . . 496
25.3.1 Topology . . . 496
25.3.2 Locally convex spaces . . . 497
25.3.3 Tensor products . . . 500
25.3.4 Measure theory . . . 502
25.4 Riesz spaces . . . 502

26 Operator Theory **504**
26.1 Operators . . . 505

26.1.1 Operator topologies . . . 505
26.1.2 Bounded operators . . . 505
26.1.3 Self-adjoint operators . . . 507
26.1.4 Compact operators . . . 508
26.1.5 Spectrum . . . 510
26.1.6 Spectral theorem & functional calculus . . . 510
26.1.7 Dixmier trace . . . 513
26.2 C^*-algebras . . . 514
26.2.1 Involutive algebras . . . 514
26.2.2 Positive maps . . . 516
26.2.3 Traces . . . 517
26.2.4 Representations . . . 518
26.2.5 Gel'fand duality . . . 518
26.2.6 Hilbert modules ♣ . . . 521
26.3 von Neumann algebras . . . 522
26.3.1 Factors . . . 523

27 Clifford Algebra **524**
27.1 Clifford algebra . . . 524
27.2 Geometric algebra . . . 526
27.3 Bott periodicity ♣ . . . 527
27.4 Pin group . . . 528
27.4.1 Clifford group . . . 528
27.4.2 Pin and Spin . . . 530

28 Noncommutative Algebra **533**
28.1 Coalgebras . . . 533
28.2 Hopf algebras . . . 535
28.2.1 Drinfel'd double . . . 538
28.3 Quantum groups . . . 539
28.4 Differential structures . . . 541
28.4.1 Differential calculi . . . 541
28.4.2 Covariance . . . 542

Chapter 22

Linear Algebra

Chapter Contents

22.1 Vector spaces **434**
22.1.1 Linear independence 434
22.1.2 Bases 435
22.1.3 Sum and direct sum 436
22.2 Linear maps **437**
22.2.1 Homomorphisms 437
22.2.2 Dual maps 438
22.3 Inner product **439**
22.3.1 Inner product space 440
22.3.2 Orthogonality 441
22.4 Matrices **443**
22.4.1 System of equations 444
22.4.2 Coordinates and matrix representations 444
22.4.3 Coordinate transformations 446
22.4.4 Determinant 448
22.4.5 Characteristic polynomial 449
22.4.6 Matrix groups 450
22.4.7 Matrix decompositions 451
22.5 Eigenvectors **452**
22.5.1 Diagonalization 453
22.5.2 Multiplicity 453
22.6 Euclidean space **454**

22.7 Algebras 455
22.8 Grassmanians 458

22.1 Vector spaces

Definition 22.1.1 ($\mathfrak{K}$-vector space). Let $\mathfrak{K}$ be a field. A $\mathfrak{K}$-vector space V is a set equipped with two operations, **(vector) addition** $V \times V \to V$ and **scalar multiplication** $\mathfrak{K} \times V \to V$, that satisfy the following axioms:

1. V forms an Abelian group under vector addition.
2. Scalar multiplication is associative: $\lambda(\mu v) = (\lambda\mu)v$ for all $\lambda, \mu \in \mathfrak{K}$ and $v \in V$.
3. The identity of the field $\mathfrak{K}$ acts as a neutral element for scalar multiplication: $1_{\mathfrak{K}} v = v$ for all $v \in V$.
4. Scalar multiplication is distributive with respect to vector addition: $\lambda(v + w) = \lambda v + \lambda w$ for all $\lambda \in \mathfrak{K}$ and $v, w \in V$.
5. Vector addition is distributive with respect to scalar multiplication: $(\lambda + \kappa)v = \lambda v + \kappa w$ for all $\lambda, \kappa \in \mathfrak{K}$ and $v \in V$.

From here on, the underlying field $\mathfrak{K}$ will be left implicit unless the results depend on it.

Remark 22.1.2. The above definition can be restated in abstract algebraic terms. A $\mathfrak{K}$-vector space is a module (Definition 2.4.53) over $\mathfrak{K}$.

22.1.1 Linear independence

Definition 22.1.3 (Linear combination). The vector w is a linear combination of elements in the set $\{v_i\}_{i \leq n} \subset V$ if it can be written as

$$w = \sum_{i=1}^{n} \lambda_i v_i \tag{22.1}$$

for some $\{\lambda_i\}_{i \leq n} \subset \mathfrak{K}$. One can generalize this to general subsets $S \subseteq V$, but the number of nonzero elements λ_i is always required to be finite.[1] (See Remark 22.1.9 about *Hamel bases*.)

[1] Generalizations are possible in the context of topological vector spaces (see Chapter 6 and Chapter 25), where one can define the notion of convergence.

Definition 22.1.4 (Linear independence). A finite set $\{v_i\}_{i\leq n}$ is said to be linearly independent if the following relation holds:

$$\sum_{i=1}^{n} \lambda_i v_i = 0 \iff \forall i \leq n : \lambda_i = 0\,. \tag{22.2}$$

A general set $S \subset V$ is said to be linearly independent if every finite subset of it is linearly independent.

Definition 22.1.5 (Span). A set of vectors $S \subseteq V$ is said to span V if every vector $v \in V$ can be written as a linear combination of elements in S.

Definition 22.1.6 (Frame). A k-frame is an ordered set of $k \in \mathbb{N}$ linearly independent vectors.

22.1.2 Bases

Definition 22.1.7 (Basis). A subset $\mathcal{B} \subset V$ that is linearly independent and spans V.

Property 22.1.8. Every spanning set contains a basis.

Remark 22.1.9 (Hamel basis). In the previous definition, the concept of a Hamel basis was implicitly used. This concept is based on two conditions:

1. The basis is linearly independent.
2. Every element in the vector space can be written as a linear combination of a <u>finite</u> subset of the basis.

For bases consisting of a finite number of vectors, one does not have to worry. However, for infinite bases, one has to keep this in mind. An alternative construction that allows for combinations of a countably infinite number of elements is given by that of a *Schauder basis*.

Nonetheless, it can be shown that every vector space admits a Hamel basis.

Construction 22.1.10 (Hamel basis ♣). Let V be a vector space and consider the set of all linearly independent subsets of V. Under the relation of inclusion, this set becomes a partially ordered set (Definition 1.7.1). Zorn's lemma 1.7.14 then says that there exists at least one maximal linearly independent set.

Now, one can show that this maximal subset S is also a spanning set of V. Choose a vector $v \in V$ that is not already in S. From the maximality of S, it follows that $S \cup v$ is linearly dependent and, hence, there exists a finite sequence of scalars $(a^1, \dots, a^n, b)$ and a finite sequence of elements $(e_1, \dots, e_n)$ in S such that:

$$\sum_{i=0}^{n} a^i e_i + bv = 0\,, \tag{22.3}$$

where not all scalars are zero. This implies that $b \neq 0$, because otherwise the set $\{e_i\}_{i \leq n}$ and, hence, also S would be linearly dependent. It follows that v can be written as[2]

$$v = -\frac{1}{b} \sum_{i=0}^{n} a^i e_i \,. \tag{22.4}$$

Because v was chosen randomly, one can conclude that S is a spanning set for V.

Remark. This construction assumes the axiom of choice in set theory, only ZF does not suffice. It can even be shown that the existence of a Hamel basis for every vector space is equivalent to the axiom of choice.

Property 22.1.11. Every basis of a vector space has the same number of elements. For infinite-dimensional spaces, this means that all bases have the same *cardinality*.

Definition 22.1.12 (Dimension). Let V be a finite-dimensional vector space and let $\mathcal{B}$ be a basis for V with n elements. With the previous property in mind, the dimension of V is defined as follows:

$$\dim(V) := n \,. \tag{22.5}$$

Definition 22.1.13 (Subspace). Let V be a vector space. A subset W of V is called a subspace if W is itself a vector space under (the restriction of) the operations of V:

$$W \leq V \iff \forall w_1, w_2 \in W, \forall \lambda \in \mathfrak{K} : \lambda w_1 + w_2 \in W \,. \tag{22.6}$$

22.1.3 Sum and direct sum

Definition 22.1.14 (Sum). Let V be a vector space and consider a finite collection of subspaces $\{W_1, \dots, W_n\}$. The sum of these subspaces is defined as follows:

$$W_1 + \cdots + W_n := \left\{ \sum_{i=1}^{n} w_i \,\middle|\, w_i \in W_i \right\} . \tag{22.7}$$

For an infinite collection of subspaces, the linear combinations have to be finite.

Definition 22.1.15 (Direct sum). If every element v of the sum can be written as a unique linear combination, the sum is called a direct sum.

Notation 22.1.16 (Direct sum). The direct sum of vector spaces is denoted by

$$W_1 \oplus \cdots \oplus W_k \equiv \bigoplus_{i=1}^{k} W_i \,.$$

Formula 22.1.17. Let V be a finite-dimensional vector space and consider two subspaces $W_1, W_2 \leq V$. The dimensions of these spaces can be related in the following way:

$$\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2) \,. \tag{22.8}$$

[2] It is this step that requires R to be a division ring in Property 2.4.63 because, otherwise, one would, in general, not be able to divide by $b \in R$.

Property 22.1.18. Let V be a vector space and assume that V can be decomposed as $W = W_1 \oplus W_2$. If $\mathcal{B}_1$ is a basis of W_1 and if $\mathcal{B}_2$ is a basis of W_2, then $\mathcal{B}_1 \cup \mathcal{B}_2$ is a basis of W.

Definition 22.1.19 (Complement). Let V be a vector space and let W be a subspace of V. A subspace W' of V is called a complement of W if $V = W \oplus W'$.

Property 22.1.20 (Existence of complements). Let V be a vector space and let U, W be two subspaces of V. If $V = U + W$, there exists a subspace $Y \leq U$ such that $V = Y \oplus W$. In particular, every subspace of V has a complement in V.

22.2 Linear maps

Definition 22.2.1 (Linear map). Let V, W be $\mathfrak{K}$-vector spaces. A function $f : V \to W$ is said to be ($\mathfrak{K}$-)linear if $f(\lambda v + w) = \lambda f(v) + f(w)$ for all $\lambda \in \mathfrak{K}$ and $v, w \in V$.

Remark 22.2.2. Linear maps are also called **linear transformations** or **linear mappings**.

22.2.1 Homomorphisms

Definition 22.2.3 (Homomorphism space). Let V, W be two $\mathfrak{K}$-vector spaces. The set of all linear maps between V and W is called the homomorphism space from V to W:

$$\mathsf{Hom}_{\mathfrak{K}}(V, W) := \left\{ f : V \to W \mid f \text{ is } \mathfrak{K}\text{-linear} \right\}. \tag{22.9}$$

The collection of $\mathfrak{K}$-vector spaces and linear maps between them form a category $\mathbf{Vect}_{\mathfrak{K}}$.

Formula 22.2.4. Let V, W be two finite-dimensional $\mathfrak{K}$-vector spaces.

$$\dim\big(\mathsf{Hom}_{\mathfrak{K}}(V, W)\big) = \dim(V)\dim(W) \tag{22.10}$$

Definition 22.2.5 (Endomorphism ring). The space $\mathsf{Hom}_{\mathfrak{K}}(V, V)$ with composition of maps as multiplication forms a ring, the endomorphism ring. It is denoted by $\mathsf{End}_{\mathfrak{K}}(V)$ or $\mathsf{End}(V)$ when the underlying field is clear.

Property 22.2.6 (Commutator). The endomorphism ring $\mathsf{End}(V)$ can also be endowed with the structure of a *Lie algebra* (see Corollary 32.2.18) by equipping it with the commutator

$$[A, B] := A \circ B - B \circ A. \tag{22.11}$$

Property 22.2.7. Let V be finite-dimensional vector space and let $f : V \to V$ be an endomorphism. The following statements are equivalent:

- f is injective,
- f is surjective, and

- f is bijective.

Definition 22.2.8 (Automorphism). The automorphism group (Definition 2.2.28) of a vector space is called the general linear group[3] and is denoted by $\mathsf{GL}_{\mathfrak{K}}(V)$ or $\mathsf{GL}(V)$ when the underlying field is clear.

Remark 22.2.9. Sometimes automorphisms are also called **linear operators**. However, this terminology is also used for a general linear map in operator theory (Chapter 26) and so this terminology is not adopted in this text.

Definition 22.2.10 (Kernel). Consider a linear map $f : V \to W$. The kernel of f is defined as the following subspace of V:

$$\ker(f) := \{v \in V \mid f(v) = 0\}\,. \tag{22.12}$$

Property 22.2.11. A linear map $f : V \to W$ is injective if and only if $\ker(f) = 0$.

Definition 22.2.12 (Rank). The dimension of the image of a linear map.

Definition 22.2.13 (Nullity). The dimension of the kernel of a linear map.

Theorem 22.2.14 (Dimension theorem[4]). *Let $f : V \to W$ be a linear map.*

$$\dim\big(\mathsf{im}(f)\big) + \dim\big(\ker(f)\big) = \dim(V) \tag{22.13}$$

Corollary 22.2.15. Two finite-dimensional vector spaces are isomorphic if and only if they have the same dimension.

Property 22.2.16 (Jordan–Chevalley decomposition). Every endomorphism A can be decomposed as follows:

$$A = A_{ss} + A_n\,, \tag{22.14}$$

where

- A_{ss} is **semisimple**: for every invariant subspace of A_{ss} there exists an invariant complementary subspace.
- A_n is **nilpotent**: $\exists k \in \mathbb{N} : A_n^k = 0$.

Furthermore, this decomposition is unique and the endomorphisms A_{ss}, A_n can be written as polynomials in A.

22.2.2 Dual maps

Definition 22.2.17 (Dual space). Let V be a $\mathfrak{K}$-vector space. The (algebraic) dual V^* of V is defined as the following vector space:

$$V^* := \mathsf{Hom}_{\mathfrak{K}}(V, \mathfrak{K}) = \big\{f : V \to \mathfrak{K} \mid f \text{ is } \mathfrak{K}\text{-linear}\big\}\,. \tag{22.15}$$

[3] It is isomorphic to the general linear group of invertible *matrices* (see Definition 22.4.3), hence the similar name and notation.

[4] Also called the **rank-nullity theorem**.

The elements of V^* are called **linear forms** or (linear) **functionals**.

Property 22.2.18 (Dimension). From Formula 22.2.4, it follows that $\dim(V^*) = \dim(V)$ whenever V is finite dimensional. If V is infinite dimensional, this property is never valid. In the infinite-dimensional case, $\mathrm{card}(V^*) > \mathrm{card}(V)$ always holds.

Definition 22.2.19 (Dual basis). Let $\mathcal{B} = \{e_1, e_2, \dots, e_n\}$ be a basis for a finite-dimensional vector space V. One can construct a basis $\mathcal{B}^* = \{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n\}$ for V^*, called the dual basis of $\mathcal{B}$, as follows:

$$\varepsilon_i : \sum_{j=1}^{n} a_i e_i \mapsto a_i \,. \tag{22.16}$$

The relation between a basis and its associated dual basis can also be expressed as

$$\varepsilon^i(e_j) = \delta^i_j \,. \tag{22.17}$$

Definition 22.2.20 (Natural pairing). The definition of the dual basis extends to a natural pairing of V and its dual V^* in terms of the following bilinear map:

$$\langle v, v^* \rangle := v^*(v) \,. \tag{22.18}$$

(See Definition 4.4.1 for a generalization of this map.)

Definition 22.2.21 (Dual map). Let $f : V \to W$ be a linear map. The linear map

$$f^* : W^* \to V^* : \varphi \to \varphi \circ f \tag{22.19}$$

is called the dual map or **transpose** of f. It is also often denoted by f^T.

Definition 22.2.22 (Dual pair). The concept of a dual space can be generalized. A dual pair (or **dual system**) is a triple $(V, W, \langle \cdot, \cdot \rangle)$, where V, W are two $\mathfrak{K}$-vector spaces and $\langle \cdot, \cdot \rangle : V \times W \to \mathfrak{K}$ is a nondegenerate bilinear map, i.e.

1. If $\langle v, \cdot \rangle = 0$ for some $v \in V$, then $v = 0$.
2. If $\langle \cdot, w \rangle = 0$ for some $w \in W$, then $w = 0$.

Note that, given a dual pair $(V, W, \langle \cdot, \cdot \rangle)$, one obtains canonical maps $V \to W^*$ and $W \to V^*$.

Definition 22.2.23 (Orthogonal complement). Consider a dual pair $(V, W, \langle \cdot, \cdot \rangle)$. The orthogonal complement of a subset $S \subseteq V$ is defined as follows:

$$S^\perp := \{w \in W \mid \forall s \in S : \langle s, w \rangle = 0\} \,.$$

Remark. $S^\perp$ is pronounced as 'S-perp'.

22.3 Inner product

In this section, all vector spaces will be defined over $\mathbb{R}$ or $\mathbb{C}$.

22.3.1 Inner product space

Definition 22.3.1 (Inner product). A function $\langle \cdot \mid \cdot \rangle : V \times V \to \mathbb{C}$ is called an inner product on V if it satisfies the following properties for all $u, v, w \in V$ and $\lambda \in \mathbb{C}$:

1. **Conjugate symmetry**: $\langle v \mid w \rangle = \langle w \mid v \rangle^*$,
2. **Linearity in the second argument**: $\langle u \mid \lambda v + w \rangle = \lambda \langle u \mid v \rangle + \langle u \mid w \rangle$,
3. **Nondegeneracy**: $\langle v \mid v \rangle = 0 \iff v = 0$, and
4. **Positive-definiteness**: $\langle v \mid v \rangle \geq 0$.

Corollary 22.3.2. The first two properties have the result of conjugate linearity in the first argument:

$$\langle \lambda f + \mu g \mid h \rangle = \overline{\lambda} \langle f \mid h \rangle + \overline{\mu} \langle g \mid h \rangle \,. \tag{22.20}$$

Therefore, these two properties together are often combined into a **sesquilinearity** axiom. When the underlying field is restricted to $\mathbb{R}$, such that the conjugate symmetry property is replaced by proper symmetry, the inner product becomes a bilinear form.

Definition 22.3.3 (Inner product space). A vector space equipped with an inner product $\langle \cdot \mid \cdot \rangle$. This is sometimes called a **pre-Hilbert space**.

Definition 22.3.4 (Metric dual). Using the inner product (or any other nondegenerate Hermitian form), one can define the metric dual of a vector as the following map:

$$L : V \to V^* : v \mapsto \langle v \mid \cdot \rangle \,. \tag{22.21}$$

(See Eq. (36.2) for a generalization.) If the sesquilinearity condition would have been stated in the reversed convention, i.e. conjugate linearity in the second argument, metric duals would be conjugate linear and, hence, would not be proper elements of the dual space.

Definition 22.3.5 (Adjoint map). Let A be a linear map on V. The (**Hermitian**) adjoint of A is defined as the linear map $A^\dagger$ that satisfies

$$\langle A^\dagger v \mid w \rangle = \langle v \mid A w \rangle \tag{22.22}$$

for all $v, w \in V$. Alternatively, one can define the adjoint using the transpose and metric dual as follows:

$$A^\dagger = L^{-1} \circ A^* \circ L \,. \tag{22.23}$$

If $A = A^\dagger$, A is said to be **Hermitian** or **self-adjoint**. (In Chapter 25, a distinction will be made between these two notions.)

22.3.2 Orthogonality

Definition 22.3.6 (Orthogonal). Consider two vectors $v, w \in V$ in an inner product space. These vectors are said to be orthogonal, denoted by $v \perp w$, if they obey the following relation:

$$\langle v \mid w \rangle = 0 \,. \tag{22.24}$$

An **orthogonal system** is a collection of vectors, none of them equal to 0, that are mutually orthogonal.

Property 22.3.7. Orthogonal systems are linearly independent.

Definition 22.3.8 (Orthonormal). A set of vectors S is said to be orthonormal if it forms an orthogonal system and if all the elements $v \in S$ obey the following relation:

$$\langle v \mid v \rangle = 1 \,. \tag{22.25}$$

By specializing Definition 22.2.23, one obtains the more classical notion of orthogonal complements.

Definition 22.3.9 (Orthogonal complement). Let W be a subspace of an inner product space V. The orthogonal complement of W is defined as the following subspace:

$$W^\perp := \left\{ v \in V \mid \forall w \in W : \langle v \mid w \rangle = 0 \right\} . \tag{22.26}$$

Property 22.3.10 (Complements). Let V be a finite-dimensional inner product space. The orthogonal complement $W^\perp$ is a complementary subspace to W, i.e. $W \oplus W^\perp = V$.

Corollary 22.3.11. Let $W \leq V$ with V a finite-dimensional inner product vector space. Forming orthogonal complements defines an involution:

$$(W^\perp)^\perp = W \,. \tag{22.27}$$

Definition 22.3.12 (Orthogonal projection). Let V be a finite-dimensional inner product vector space and consider a subspace $W \leq V$. Consider a vector $w \in W$ and let $\{w_1, \dots, w_k\}$ be an orthonormal basis of W. The projections of $v \in V$ on W and $w \in W$ are defined as follows:

$$\begin{aligned} \mathsf{proj}_W(v) &:= \sum_{i=1}^{k} \langle v \mid w_i \rangle w_i \,, \\ \mathsf{proj}_w(v) &:= \frac{\langle v \mid w \rangle}{\langle w \mid w \rangle} w \,. \end{aligned} \tag{22.28}$$

Property 22.3.13. Orthogonal projections satisfy the following conditions:

$$\forall w \in W : \mathsf{proj}_W(w) = w \qquad \text{and} \qquad \forall u \in W^\perp : \mathsf{proj}_W(u) = 0 \,. \tag{22.29}$$

Method 22.3.14 (Gram–Schmidt orthonormalization). Let $\{u_i\}_{i\leq n}$ be a set of linearly independent vectors. An orthonormal set $\{e_i\}_{i\leq n}$ can be constructed out of $\{u_i\}_{i\leq n}$ using the following procedure:

1. Orthogonalization:

$$\begin{aligned} w_1 &= u_1 \\ w_2 &= u_2 - \frac{\langle u_2 \mid w_1\rangle}{\|u_2\|^2} w_1 \\ &\vdots \\ w_n &= u_n - \sum_{i=1}^{n-1} \frac{\langle u_n \mid w_i\rangle}{\|u_n\|^2} w_i \,. \end{aligned} \tag{22.30}$$

2. Normalization:

$$\begin{aligned} e_1 &= \frac{w_1}{\|w_1\|} \\ e_2 &= \frac{w_2}{\|w_2\|} \\ &\vdots \\ e_n &= \frac{w_n}{\|w_n\|} \,. \end{aligned} \tag{22.31}$$

Remark 22.3.15. Inner products can be generalized to **nondegenerate Hermitian forms** which do not satisfy the positive-definiteness property. For any vector space with a nondegenerate Hermitian form, one can find a basis such $\{e_1, \dots, e_{p+q}\}$ such that

$$\langle e_i \mid e_i\rangle = \begin{cases} 1 & \text{if } i \leq p \,, \\ -1 & \text{if } p < i \leq p+q \,. \end{cases} \tag{22.32}$$

These spaces are said to have (metric) **signature** (p, q). A vector space with signature $(1, k)$ or $(k, 1)$ is said to be **Lorentzian**. Note that the signature is basis independent due to *Sylvester's theorem* (see Theorem 22.5.19).

Definition 22.3.16 (Householder transformation). Let v be an element of an inner product space V. The Householder transformation generated by v is defined as the linear map

$$\sigma_v : V \to V : w \mapsto w - 2\frac{\langle w \mid v\rangle}{\langle v \mid v\rangle} v \,. \tag{22.33}$$

This transformation amounts to a reflection in the hyperplane orthogonal to v.

Definition 22.3.17 (Angle). Let v, w be elements of an inner product space V. The angle θ between v and w is defined by the following formula:

$$\cos\theta := \frac{\langle v \mid w\rangle}{\|v\|\,\|w\|} \,. \tag{22.34}$$

The angle between two vectors v, w is sometimes denoted by $\sphericalangle(v, w)$.

22.4 Matrices

Notation 22.4.1. The vector space of all $m \times n$-matrices defined over the field $\mathfrak{K}$ is denoted by $M_{m,n}(\mathfrak{K})$. If $m = n$, the space is denoted by $M_n(\mathfrak{K})$ or $M(n, \mathfrak{K})$.

Property 22.4.2 (Dimension). The dimension of $M_{m,n}(\mathfrak{K})$ is mn.

Definition 22.4.3 (General linear group). The set of invertible matrices is called the general linear group and is denoted by $\mathsf{GL}_n(\mathfrak{K})$ or $\mathsf{GL}(n, \mathfrak{K})$.

Property 22.4.4. For all $A \in \mathsf{GL}_n(\mathfrak{K})$, one has:

- $A^T \in \mathsf{GL}_n(\mathfrak{K})$, and
- $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$.

Definition 22.4.5 (Trace). Let $A \equiv (a_{ij}) \in M_n(\mathfrak{K})$. The trace of A is defined as follows:

$$\mathsf{tr}(A) := \sum_{i=1}^{n} a_{ii} \,. \tag{22.35}$$

Property 22.4.6. Let $A, B \in M_n(\mathfrak{K})$. The trace satisfies the following properties:

- $\mathsf{tr} : M_n(\mathfrak{K}) \to k$ is a linear map,
- $\mathsf{tr}(AB) = \mathsf{tr}(BA)$, and
- $\mathsf{tr}(A^T) = \mathsf{tr}(A)$.

Formula 22.4.7 (Hilbert–Schmidt norm). The Hilbert–Schmidt (or **Frobenius**) norm is defined by the following formula:

$$\|A\|_{HS}^2 := \sum_{i,j} |A_{ij}|^2 = \mathsf{tr}\left(A^\dagger A\right) . \tag{22.36}$$

If one identifies $M_n(\mathbb{C})$ with $\mathbb{C}^{2n}$, this norm equals the standard Hermitian norm.

Formula 22.4.8 (Hadamard product). The Hadamard product of two matrices is defined as the entrywise product:

$$(A \circ B)_{ij} := A_{ij} B_{ij} \,. \tag{22.37}$$

Property 22.4.9. Let $A \in M_{m,n}(\mathfrak{K})$. Denote the set of columns as $\{A_1, A_2, \dots, A_n\}$ and the set of rows as $\{R_1, R_2, \dots, R_m\}$. The set of columns is a subspace of $\mathfrak{K}^m$ and the set of rows is a subspace of $\mathfrak{K}^n$. Their spans satisfy the following property:

$$\mathsf{dim}\big(\mathsf{span}(A_1, \dots, A_n)\big) = \mathsf{dim}\big(\mathsf{span}(R_1, \dots, R_m)\big) \,. \tag{22.38}$$

Definition 22.4.10 (Rank). Using the invariance relation from the previous property, one can define the rank of a matrix $A \in M_{m,n}(\mathfrak{K})$ as follows:

$$\mathsf{rk}(A) := \mathsf{dim}\big(\mathsf{span}(A_1, \dots, A_n)\big) = \mathsf{dim}\big(\mathsf{span}(R_1, \dots, R_m)\big) \,. \tag{22.39}$$

Property 22.4.11. Let $A \in M_{m,n}(\mathfrak{K}), B \in \mathsf{GL}_n(\mathfrak{K}), C \in M_{n,r}(\mathfrak{K})$ and $D \in M_{r,n}(\mathfrak{K})$. The ranks of these matrices satisfy the following properties:

- $\mathsf{rk}(AC) \leq \mathsf{rk}(A)$,
- $\mathsf{rk}(AC) \leq \mathsf{rk}(C)$,
- $\mathsf{rk}(BC) = \mathsf{rk}(C)$, and
- $\mathsf{rk}(DB) = \mathsf{rk}(D)$.

Property 22.4.12. Let $A \in M_{m,n}(\mathfrak{K})$. The linear map

$$L_A : \mathfrak{K}^n \to \mathfrak{K}^m : v \mapsto Av \tag{22.40}$$

satisfies $\mathsf{im}(L_A) = \mathsf{span}(A_1, \dots, A_n)$.

22.4.1 System of equations

Property 22.4.13. Let $Ax = b$ with $A \in M_{m,n}(\mathfrak{K}), x \in \mathfrak{K}^n$ and $b \in \mathfrak{K}^m$ be a system of $m \in \mathbb{N}$ equations in $n \in \mathbb{N}$ variables and let L_A be the linear map as defined in Property 22.4.12. The following properties hold:

- The system is inconsistent if and only if $b \notin \mathsf{im}(L_A)$.
- If the system is not inconsistent, the solution set is an affine space. If $x_0 \in \mathfrak{K}^n$ is a solution, the solution set is given by: $x_0 + \mathsf{ker}(L_A)$.
- If the system is homogeneous, i.e. $b = 0$, the solution set is equal to $\mathsf{ker}(L_A)$.

Property 22.4.14 (Uniqueness). Let $Ax = b$ with $A \in M_n(\mathfrak{K})$ be a system of $n \in \mathbb{N}$ equations in n variables. If $\mathsf{rk}(A) = n$, the system has a unique solution.

Formula 22.4.15 (Cramer's rule). Let $Ax = b$ be a system of linear equations where the matrix A has a nonzero determinant. There exists a unique solution:

$$x_i = \frac{\mathsf{det}(A_i)}{\mathsf{det}(A)}, \tag{22.41}$$

where A_i is the matrix obtained by replacing the ith column of A by the column vector b.

22.4.2 Coordinates and matrix representations

Definition 22.4.16 (Coordinate vector). Let $\mathcal{B} = \{b_1, \dots, b_n\}$ be a basis of V and consider the vector $v = \sum_{i=1}^n \lambda_i b_i$. The coordinate vector of v with respect to $\mathcal{B}$ is defined as the column vector $(\lambda_1, \dots, \lambda_n)^T$. The scalars λ_i are called the **coordinates** of v with respect to $\mathcal{B}$.

Definition 22.4.17 (Coordinate isomorphism). With the previous definition in mind, one can define the coordinate isomorphism induced by $\mathcal{B}$ as follows:

$$\beta : V \to \mathfrak{K}^n : \sum_{i=1}^{n} \lambda_i b_i \mapsto (\lambda_1, \dots, \lambda_n)^T . \tag{22.42}$$

Construction 22.4.18 (Matrix representation). Let V, W be m- and n-dimensional vector spaces with bases $\mathcal{B} = \{b_1, \dots, b_m\}, \mathcal{C} = \{c_1, \dots, c_n\}$ and consider a linear map $f : V \to W$. The matrix representation of f with respect to $\mathcal{B}$ and $\mathcal{C}$ is defined as the matrix $A_{f,\mathcal{B},\mathcal{C}}$ that satisfies the following condition for all vectors $v \in V$. Let $(\lambda_1, \dots, \lambda_n)^T$ be the coordinate vector of v with respect to $\mathcal{B}$ and let $(\mu_1, \dots, \mu_m)^T$ be the coordinate vector of $f(v)$ with respect to $\mathcal{C}$, then

$$\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix} = A_{f,\mathcal{B},\mathcal{C}} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}. \tag{22.43}$$

This matrix can be constructed as follows. For every $j \in \{1, \dots, m\}$, write $f(b_j) = \sum_{i=1}^{n} a_{ij} c_i$. The matrix $A_{f,\mathcal{B},\mathcal{C}} \equiv (a_{ij}) \in M_{n,m}(\mathfrak{K})$ is called the matrix representation of f. The jth column of $A_{f,\mathcal{B},\mathcal{C}}$ coincides with the coordinate vector of $f(b_j)$ with respect to $\mathcal{C}$.

The following property shows that the matrix algebra $M_{m,n}(\mathfrak{K})$ is isomorphic to the algebra[5] of linear maps $\mathcal{L}(\mathfrak{K}^n, \mathfrak{K}^m)$, thereby explaining why the same notation for the space of invertible matrices (Definition 22.4.3) and the space of automorphisms (Definition 22.2.8) was used.

Property 22.4.19 (Matrices and linear maps). For every matrix $A \in M_{m,n}(\mathfrak{K})$, there exists a linear map $f : \mathfrak{K}^n \to \mathfrak{K}^m$ such that $A_{f,\mathcal{B},\mathcal{C}} = A$. Conversely, for every linear map $f : \mathfrak{K}^m \to \mathfrak{K}^n$ there exists a matrix $A \in M_{n,m}(\mathfrak{K})$ such that $f = L_A$ (given by the previous construction).

Corollary 22.4.20. Let $f \in \mathsf{End}(V)$ and let A_f be the corresponding matrix representation. The linear map f is invertible if and only if A_f is invertible. Furthermore, if A_f is invertible,

$$\left(A_f\right)^{-1} = A_{f^{-1}} . \tag{22.44}$$

In other words, the linear isomorphism $\mathsf{End}(V) \to M_n(\mathfrak{K})$ descends to a group isomorphism

$$\mathsf{GL}(V) \to \mathsf{GL}_n(\mathfrak{K}) : f \mapsto A_f , \tag{22.45}$$

where $n = \mathsf{dim}(V)$.

[5] The multiplication is given by the composition of linear maps.

Formula 22.4.21 (Linear forms). Let $V \cong \mathfrak{K}^n$ and consider a linear form $f \in V^*$. Equation (22.43) can be rewritten as

$$f\big((\lambda_1, \dots, \lambda_n)^T\big) = (f(e_1), \dots, f(e_n))(\lambda_1, \dots, \lambda_n)^T = \sum_{i=1}^{n} f(e_i)\lambda_i\,, \tag{22.46}$$

where $\{e_i\}_{i\in I}$ is the standard basis of k^n. In terms of the standard dual basis $\{\varepsilon_1, \dots, \varepsilon_n\}$, this becomes:

$$f = \sum_{i=1}^{n} f(e_i)\varepsilon_i\,. \tag{22.47}$$

Property 22.4.22 (Transpose). Let $f : V \to W$ be a linear map and let $f^* : W^* \to V^*$ be the corresponding dual map. If A_f is the matrix representation of f with respect to $\mathcal{B}$ and $\mathcal{C}$, the transpose A_f^T is the matrix representation of f^* with respect to the dual basis of $\mathcal{C}$ and the dual basis of $\mathcal{B}$.

Corollary 22.4.23. The Hermitian adjoint of a linear map (Definition 22.3.5) induces the (Hermitian) adjoint of matrices $A \in \mathbb{C}^{m\times n}$. It is given by

$$A^\dagger = \overline{A}^T\,, \tag{22.48}$$

where $\overline{A}$ denotes the complex conjugate of A.

22.4.3 Coordinate transformations

Definition 22.4.24 (Transition matrix). Let $\mathcal{B} = \{b_1, \dots, b_n\}$ and $\mathcal{B}' = \{b'_1, \dots, b'_n\}$ be two bases of V. By definition, every element of $\mathcal{B}'$ can be written as a linear combination of elements in $\mathcal{B}$:

$$b'_j = q_{1j}b_1 + \cdots + q_{nj}b_n\,. \tag{22.49}$$

The matrix $Q \equiv (q_{ij}) \in M_n(\mathfrak{K})$ is called the transition matrix from $\mathcal{B}$ to $\mathcal{B}'$.

Property 22.4.25. Let $\mathcal{B}, \mathcal{B}'$ be two bases of V and let Q be the transition matrix from $\mathcal{B}$ to $\mathcal{B}'$. The following statements hold:

- $Q \in \mathsf{GL}_n(\mathfrak{K})$ and Q^{-1} is the transition matrix from $\mathcal{B}'$ to $\mathcal{B}$.
- Let $\mathcal{C}$ be an arbitrary basis of V with γ the corresponding coordinate isomorphism and define the following matrices:

$$B := (\gamma(b_1), \dots, \gamma(b_n)) \quad \text{and} \quad B' := (\gamma(b'_1), \dots, \gamma(b'_n))\,. \tag{22.50}$$

  In terms of these matrices, one finds that $BQ = B'$.
- Consider $v \in V$. Let $(\lambda_1, \dots, \lambda_n)^T$ be the coordinate vector with respect to $\mathcal{B}$ and let $(\lambda'_1, \dots, \lambda'_n)^T$ be the coordinate vector with respect to $\mathcal{B}'$, then

$$Q\begin{pmatrix}\lambda'_1\\ \vdots\\ \lambda'_n\end{pmatrix} = \begin{pmatrix}\lambda_1\\ \vdots\\ \lambda_n\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}\lambda'_1\\ \vdots\\ \lambda'_n\end{pmatrix} = Q^{-1}\begin{pmatrix}\lambda_1\\ \vdots\\ \lambda_n\end{pmatrix}. \tag{22.51}$$

Corollary 22.4.26 (Basis change). Let V, W be two finite-dimensional vector spaces. Consider two bases $\mathcal{B}, \mathcal{B}'$ of V and two bases $\mathcal{C}, \mathcal{C}'$ of W. Let Q, P be the transition matrices from $\mathcal{B}$ to $\mathcal{B}'$ and from $\mathcal{C}$ to $\mathcal{C}'$, respectively. The matrix representations $A = A_{f,\mathcal{B},\mathcal{C}}$ and $A' = A_{f,\mathcal{B}',\mathcal{C}'}$ of a linear map $f : V \to W$ are related in the following way:

$$A' = P^{-1}AQ\,. \tag{22.52}$$

Remark 22.4.27. From the definition of the transition matrix and the above property, it follows that the basis vectors and coordinate representations transform by Q and Q^{-1} respectively. That they transform in an inverse manner makes sense, since a vector should be independent from its coordinate representation:

$$v' = \sum_{i=1}^{n} \lambda'_i e'_i = \sum_{i,j,k=1}^{n} Q^{-1}_{ji}\lambda_j Q_{ik} e_k = \sum_{i,j,k=1}^{n} \delta_{jk}\lambda_j e_k = v\,. \tag{22.53}$$

This remark gives a new way to define a vector $v \in V$.

Alternative Definition 22.4.28 (Vector). Consider an n-dimensional vector space V. One can define an equivalence relation on the set $\mathfrak{K}^n \times FV$, where FV denotes the set of all bases of V, by saying that the pairs $(c, \mathfrak{b})$ and $(c', \mathfrak{b}')$ are equivalent if and only if there exists a matrix $A \in \mathsf{GL}_n(\mathfrak{K})$ such that $c' = Ac$ and $\mathfrak{b} = A\mathfrak{b}'$. A vector $v \in V$ is then defined as an equivalence class of such pairs.

Definition 22.4.29 (Matrix conjugation). Let $A \in M_n(\mathfrak{K})$. The set

$$\left\{Q^{-1}AQ \mid Q \in \mathsf{GL}_n(\mathfrak{K})\right\} \tag{22.54}$$

is called the conjugacy class of A in accordance with group theory (Definition 2.2.17). Another term for conjugation is **similarity transformation**.

Remark 22.4.30. If A is a matrix representation of a linear operator f, the conjugacy class of A consists of all matrix representations of f.

Property 22.4.31 (Trace). Property 22.4.6 implies that the trace of a matrix is invariant under conjugation:

$$\mathrm{tr}(Q^{-1}AQ) = \mathrm{tr}(A)\,. \tag{22.55}$$

Definition 22.4.32 (Matrix congruence). Let $A, B \in M_n(\mathfrak{K})$. The matrices are said to be congruent if there exists a matrix P such that

$$A = P^T BP\,. \tag{22.56}$$

Property 22.4.33. Every matrix congruent to a symmetric matrix is also symmetric.

Property 22.4.34 (Orthogonality of basis changes). Let V be an inner product space and let $\mathcal{B}, \mathcal{B}'$ be two orthonormal bases of V with transition matrix Q. Q is *orthogonal* (see Definition 22.4.56):

$$Q^T Q = \mathbb{1}_n\,. \tag{22.57}$$

22.4.4 Determinant

Definition 22.4.35 (Minor). The (i,j)-th minor of A is defined as $\det(A_{ij})$ where $A_{ij} \in M_{n-1}(\mathfrak{K})$ is the matrix obtained by removing the ith row and the jth column from A.

Definition 22.4.36 (Cofactor). The cofactor α_{ij} of the matrix element a_{ij} is defined as $(-1)^{i+j}\det(A_{ij})$.

Definition 22.4.37 (Adjugate matrix). The adjugate matrix of $A \in M_n(\mathfrak{K})$ is defined as follows:

$$\operatorname{adj}(A) := \begin{pmatrix} \alpha_{11} & \alpha_{21} & \cdots & \alpha_{n1} \\ \alpha_{12} & \alpha_{22} & \cdots & \alpha_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1n} & \alpha_{2n} & \cdots & \alpha_{nn} \end{pmatrix}, \tag{22.58}$$

or, in terms of the cofactors: $\operatorname{adj}(A) = (\alpha_{ij})^T$, where the transpose is taken after the elements have been replaced by their cofactor.

Formula 22.4.38 (Laplace). The determinant of a matrix $A \equiv (a_{ij}) \in M_n(\mathfrak{K})$ can be evaluated as follows for any $l \leq n$:

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+l} a_{il} \det(A_{il}) \,. \tag{22.59}$$

Property 22.4.39. Let $A, B \in M_n(\mathfrak{K})$ and denote the columns of A by $A_1, \dots, A_n$. The determinant has the following properties:

- $\det(AB) = \det(A)\det(B)$,
- $\det(A^T) = \det(A)$,
- $\det(A_1, \dots, A_i + \lambda A_i', \dots, A_n) = \det(A_1, \dots, A_i, \dots, A_n) + \lambda \det(A_1, \dots, A_i', \dots, A_n)$ for all $A_i, A_i' \in M_{n,1}(\mathfrak{K})$, and
- $\det(A_{\sigma(1)}, \dots, A_{\sigma(n)}) = \operatorname{sgn}(\sigma)\det(A_1, \dots, A_n)$.

Items 2, 3 and 4 further imply that a matrix with two identical rows or columns has a vanishing determinant.

Property 22.4.40. Let $A \in M_n(\mathfrak{K})$. The following statements are equivalent:

- $\det(A) \neq 0$,
- $\operatorname{rk}(A) = n$, or
- $A \in \mathrm{GL}_n(\mathfrak{K})$.

Property 22.4.41. For all $A \in M_n(\mathfrak{K})$, one finds that

$$A \operatorname{adj}(A) = \operatorname{adj}(A) A = \det(A) I_n \,. \tag{22.60}$$

Corollary 22.4.42. For all $A \in \mathsf{GL}_n(\mathfrak{K})$, one finds that

$$A^{-1} = \det(A)^{-1} \operatorname{adj}(A)\,. \tag{22.61}$$

Alternative Definition 22.4.43 (Minor). Let $A \in M_{m,n}(\mathfrak{K})$ and choose $l \leq \min(m,n)$. A $l \times l$-minor of A is the determinant of a $l \times l$-partial matrix obtained by removing $m - l$ rows and $n - l$ columns from A.

Property 22.4.44. Let $A \in M_{m,n}(\mathfrak{K})$ and choose $l \leq \min(m,n)$. Then $\operatorname{rk}(A) \geq l$ if and only if A contains a nonzero $l \times l$-minor.

Property 22.4.45 (Invariance of determinant). Let $f \in \operatorname{End}(V)$. The determinant of the matrix representation of f is invariant under basis transformations.

Definition 22.4.46 (Determinant of a linear map). The previous property allows for an unambiguous definition of the determinant of $f \in \operatorname{End}(V)$:

$$\det(f) := \det(A) \tag{22.62}$$

for any matrix representation A of f.

22.4.5 Characteristic polynomial

Definition 22.4.47 (Characteristic polynomial). Consider a linear map $f \in \operatorname{End}(V)$ and denote its matrix representation by A_f. The function

$$\chi_f(x) := \det(x\mathbb{1}_n - A_f) \in \mathfrak{K}[x] \tag{22.63}$$

is a monic polynomial of degree $n \in \mathbb{N}$ in the variable x. The following equation is called the **characteristic equation** or **secular equation** of f:

$$\chi_f(x) = 0\,. \tag{22.64}$$

Formula 22.4.48. Consider a matrix $A \equiv (a_{ij}) \in M_n(\mathfrak{K})$ with characteristic polynomial

$$\chi_A(x) = x^n + c_{n-1}x^{n-1} + \ldots + c_1 x + c_0\,.$$

The first and last of the coefficients c_i have a simple expression:

$$\begin{cases} c_0 = (-1)^n \det(A), \\ c_{n-1} = -\operatorname{tr}(A)\,. \end{cases} \tag{22.65}$$

Theorem 22.4.49 (Cayley–Hamilton). *Consider a linear map $f \in \operatorname{End}(V)$ with characteristic polynomial χ_f.*

$$\chi_f(f) = f^n + \sum_{i=1}^{n-1} c_i f^i = 0\,. \tag{22.66}$$

Corollary 22.4.50. From Property 2.5.31 and the Cayley–Hamilton theorem, it follows that the minimal polynomial μ_f is a divisor of the characteristic polynomial χ_f.

22.4.6 Matrix groups

Definition 22.4.51 (Elementary matrix). An elementary matrix is a matrix of the form

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & a & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & b & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}, \ldots$$

i.e. it is equal to the sum of an identity matrix and a multiple of a matrix unit U_{ij}. The elementary matrix with the scalar c at position (i, j) is denoted by $E_{ij}(c)$.

A second type of elementary matrix is one of the form

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & a & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & a & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

These matrices are sometimes denoted by $T_{i,j}$.

Property 22.4.52 (Invertibility). Elementary matrices have determinant 1 and, accordingly, are elements of $\mathsf{GL}_n(\mathfrak{K})$.

Property 22.4.53. Multiplication by an elementary matrix has the following properties:

- Left multiplication by an elementary matrix $E_{ij}(c)$ comes down to replacing the ith row of the matrix with the ith row plus c times the jth row.
- Right multiplication by an elementary matrix $E_{ij}(c)$ comes down to replacing the jth column of the matrix with the jth column plus c times the ith column.
- Left multiplication by an elementary matrix $T_{i,j}$ interchanges the ith and j^{th} rows.

Property 22.4.54. Every invertible matrix can be written as a product of elementary matrices.

Definition 22.4.55 (Special linear group). The subgroup of $\mathsf{GL}_n(\mathfrak{K})$ consisting of all matrices with determinant 1:

$$\mathsf{SL}_n(\mathfrak{K}) := \{A \in \mathsf{GL}_n(\mathfrak{K}) \mid \det(A) = 1\}. \tag{22.67}$$

Definition 22.4.56 (Orthogonal group). The orthogonal and special orthogonal group are defined as follows:

$$\begin{aligned} \mathsf{O}(n,k) &:= \{A \in \mathsf{GL}_n(\mathfrak{K}) \mid AA^T = A^TA = \mathbb{1}_n\}, \\ \mathsf{SO}(n,k) &:= \mathsf{O}_n(\mathfrak{K}) \cap \mathsf{SL}_n(\mathfrak{K}). \end{aligned} \tag{22.68}$$

Definition 22.4.57 (Unitary group). Consider a field $\mathfrak{K}$ equipped with an involution $\sigma : \lambda \mapsto \overline{\lambda}$. The unitary and special unitary group are defined as follows:

$$\begin{aligned} \mathsf{U}_n(k,\sigma) &:= \{A \in \mathsf{GL}_n(\mathfrak{K}) \mid A\sigma(A)^T = \sigma(A)^T A = \mathbb{1}_n\}\,, \\ \mathsf{SU}_n(k,\sigma) &:= \mathsf{U}_n(\mathfrak{K}) \cap \mathsf{SL}_n(\mathfrak{K})\,. \end{aligned} \tag{22.69}$$

In practice, $\mathfrak{K}$ is often $\mathbb{C}$ with complex conjugation as the involution. For this reason the notation $A^\dagger := \sigma(A)^T$ is common. Moreover, in the case $\mathfrak{K} = \mathbb{C}$, the notation is further simplified to $\mathsf{U}(n)$ and $\mathsf{SU}(n)$.

Definition 22.4.58 (Unitary equivalence). Let $A, B \in M_n(\mathfrak{K})$ over a field $\mathfrak{K}$ with an involution. The matrices are said to be unitarily equivalent if there exists a unitary matrix U such that

$$A = U^\dagger B U\,.$$

Property 22.4.59. For orthogonal matrices, conjugacy (Definition 22.4.29) and congruency (Definition 22.4.32) coincide. More generally, for unitary matrices, conjugacy and unitary equivalence coincide.

Definition 22.4.60 (Symplectic group). Consider a vector space V with an antisymmetric nonsingular matrix Ω. The symplectic group $\mathsf{Sp}(V,\Omega)$ is defined as follows:

$$\mathsf{Sp}(V,\Omega) := \{A \in \mathsf{GL}(V) \mid A^T \Omega A = \Omega\}\,. \tag{22.70}$$

Over the real or complex numbers one can define the canonical **symplectic** matrix

$$\Omega_{st} := \begin{pmatrix} 0 & -\mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix}. \tag{22.71}$$

The groups of matrices that preserve this matrix are denoted by $\mathsf{Sp}(n,\mathbb{R})$ and $\mathsf{Sp}(n,\mathbb{C})$.

Remark 22.4.61. Symplectic groups can only be defined on even-dimensional spaces because antisymmetric matrices can only be nonsingular if the dimension n is even.

Definition 22.4.62 (Compact symplectic group). The compact symplectic group is defined as follows (although the notation is confusing, it is standard):

$$\mathsf{Sp}(n) := \mathsf{Sp}(2n,\mathbb{C}) \cap \mathsf{U}(2n)\,. \tag{22.72}$$

This group is in fact isomorphic to the *quaternionic* unitary group in n quaternionic dimensions.

Property 22.4.63.

$$\mathsf{Sp}(1) \cong \mathsf{SU}(2)$$

22.4.7 Matrix decompositions

Method 22.4.64 (QR Decomposition). Every square complex matrix M can be decomposed as

$$M = QR \tag{22.73}$$

with Q unitary and R upper triangular. The easiest way to achieve this decomposition is by applying the Gram–Schmidt orthonormalization process:

> Let $\{v_i\}_{i\leq n}$ be a basis for the column space of M. By applying the Gram–Schmidt process to this basis one obtains a new orthonormal basis $\{e_i\}_{i\leq n}$. The matrix M can then be written as a product QR of the following matrices:
>
> - an upper-triangular matrix R with entries $R_{ij} = \langle e_i \mid \mathrm{col}_j(M)\rangle$, where $\mathrm{col}_j(M)$ denotes the jth column of M.
> - a unitary matrix $Q = (e_1, \dots, e_n)$ constructed by setting the ith column equal to the ith basis vector e_i.

Property 22.4.65. If M is invertible and if the diagonal elements of R are required to have positive norm, the QR-decomposition is unique.

@@ COMPLETE (Cholesky, polar, ...) @@

22.5 Eigenvectors

Definition 22.5.1 (Eigenvector). A vector $v \in V\backslash\{0\}$ is called an **eigenvector** of the linear map $f : V \to V$ if it satisfies

$$f(v) = \lambda v \tag{22.74}$$

for some $\lambda \in \mathfrak{K}$. The scalar λ is called the **eigenvalue** associated to v.

Definition 22.5.2 (Eigenspace). The subspace of V spanned by the eigenvectors of a linear map is called the eigenspace of that linear map. It is given by

$$\ker(\lambda \mathbb{1}_V - f)\,. \tag{22.75}$$

It follows, that the eigenvalues are exactly those scalars for which the linear map $\lambda\mathbb{1}_V - f$ is not injective. (This is generalized in Section 26.1.5.)

Property 22.5.3 (Characteristic equation). Consider a linear map $f \in \mathsf{End}(V)$. A scalar $\lambda \in \mathfrak{K}$ is an eigenvalue of f if and only if it satisfies the characteristic equation (22.64).

Property 22.5.4. A linear map $f \in \mathsf{End}(V)$ defined over an n-dimensional vector space V has at most $n \in \mathbb{N}$ different eigenvalues.

These property leads to the following method for finding eigenvectors.

Method 22.5.5 (Finding the eigenvectors of a matrix). To calculate the eigenvectors of a matrix, one should perform the following steps:

1. Find the eigenvalues λ_i of A by solving the characteristic equation (22.64).

2. Find the eigenvector v_i associated to the eigenvalue λ_i by solving

$$(A - \lambda_i \mathbb{1}_V) v_i = 0 \,. \tag{22.76}$$

Theorem 22.5.6 (Perron–Frobenius). *Let $A \in M_n(\mathbb{R})$ be a matrix with positive entries. A has a unique eigenvalue of largest magnitude and, moreover, this eigenvalue is real.*

22.5.1 Diagonalization

Definition 22.5.7 (Diagonalizable map). Let V be a finite-dimensional vector space. A linear map $f \in \mathsf{End}(V)$ is said to be diagonalizable if it admits a diagonal matrix representation.

Property 22.5.8. Every diagonalizable map is semisimple (Property 22.2.16). Conversely, in finite dimensions (and over an algebraically closed field), a semisimple map is diagonalizable.

Theorem 22.5.9. *A matrix $A \in M_n(\mathfrak{K})$ is diagonalizable if and only if there exists a matrix $P \in \mathsf{GL}_n(\mathfrak{K})$ such that $P^{-1}AP$ is diagonal.*

Corollary 22.5.10 (Trace). Using Property 22.4.31, which states that the trace of a linear map is invariant under similarity transformations, the following useful formula can be proven:

$$\mathsf{tr}(f) = \sum_{i=0}^{n} \lambda_i \,, \tag{22.77}$$

where $\{\lambda_i\}_{i \leq n}$ are the eigenvalues of f.

Property 22.5.11. Let V be an n-dimensional vector space and let $f \in \mathsf{End}(V)$ be a linear map. The eigenvalues and eigenvectors of f satisfy the following properties:

- The eigenvectors of f belonging to different eigenvalues are linearly independent.
- If f has exactly n eigenvalues, f is diagonalizable.
- If f is diagonalizable, then V is the direct sum of the eigenspaces of f belonging to the different eigenvalues of f.

Theorem 22.5.12. *A linear map defined on a finite-dimensional vector space is diagonalizable if and only if its set of eigenvectors forms a basis of the vector space.*

22.5.2 Multiplicity

Definition 22.5.13 (Multiplicity). Let V be a vector space and let $f \in \mathsf{End}(V)$ have characteristic polynomial

$$\chi_f(x) = \prod_{i=1}^{n} (x - \lambda_i)^{n_i} \,. \tag{22.78}$$

The multiplicities are defined as follows:

- The **algebraic multiplicity** of an eigenvalue λ_i is equal to n_i.
- The **geometric multiplicity** of an eigenvalue λ_i is equal to the dimension of the eigenspace belonging to that eigenvalue.

Remark 22.5.14 (Splitting field). In the previous definition, it was assumed that the characteristic polynomial can be completely factorized. However, this depends on the possibility to completely factorize the polynomial over $\mathfrak{K}$ (i.e. if it has 'enough' roots' in $\mathfrak{K}$). If not, f cannot even be diagonalized. In general, there always exists a field f containing $\mathfrak{K}$, the splitting field (Definition 2.5.27), over which the polynomial can be completely factorized. Note that, in general, this field is strictly smaller than the algebraic closure of $\mathfrak{K}$, which is the splitting field of the collection of all polynomials over $\mathfrak{K}$.

Property 22.5.15. The algebraic multiplicity is always greater than or equal to the geometric multiplicity.

Theorem 22.5.16. *A linear map $f \in \mathsf{End}(V)$ is diagonalizable if and only if for every eigenvalue the algebraic multiplicity is equal to the geometric multiplicity.*

Property 22.5.17. Every Hermitian linear map $f \in \mathsf{End}(\mathbb{C}^n)$ has the following properties:

- All the eigenvalues of f are real.
- Eigenvectors belonging to different eigenvalues are orthogonal.
- f is diagonalizable and there always exists an orthonormal basis of eigenvectors of f, in particular, the diagonalizing matrix P is unitary, i.e. $P^{-1} = P^\dagger$.

Property 22.5.18 (Commutator). Let $f, g \in \mathsf{End}(V)$ be two diagonalizable maps. If the commutator $[f, g]$ is zero, the two maps have a common eigenbasis.

Theorem 22.5.19 (Sylvester's law of inertia). *The number of positive and negative eigenvalues of a Hermitian matrix is invariant with respect to $\dagger$-congruence (or conjugation due to Property 22.4.59).*

22.6 Euclidean space

A finite-dimensional $\mathbb{R}$-vector space is sometimes called a **Euclidean** or **Cartesian space**.

Notation 22.6.1. When working in a Euclidean space, the inner product $\langle v \mid w \rangle$ is often written as $v \cdot w$.

Definition 22.6.2 (Orientation). Let $\mathcal{B}, \mathcal{B}'$ be two ordered bases of $\mathbb{R}^n$ and let Q be the transition matrix from $\mathcal{B}$ to $\mathcal{B}'$. If $\det(Q) > 0$, the bases are said to have the same

orientation (or to be **consistently oriented**). If $\det(Q) < 0$, the bases are said to have an opposite orientation.

Corollary 22.6.3 (Positive orientation). The previous definition imposes an equivalence relation on the set of bases of $\mathbb{R}^n$ with exactly two equivalence classes. The bases in one of these classes are said to be **positively** (or **directly**) oriented. The bases in the other class are then said to be **negatively** (or **indirectly**) oriented.

Remark 22.6.4. It is convenient to take the standard basis $(e_1, \dots, e_n)$ to be positively oriented.

22.7 Algebras

Definition 22.7.1 (Algebra). Let V be a vector space equipped with a binary operation $\star : V \times V \to V$. The pair $(V, \star)$ is called an algebra over k if it satisfies the following conditions:

1. **Right distributivity**: $(x + y) \star z = x \star z + y \star z$,
2. **Left distributivity**: $x \star (y + z) = x \star y + x \star z$, and
3. **Compatibility with scalars**: $(\lambda x) \star (\mu y) = \lambda\mu(x \star y)$.

These conditions say that the binary operation is bilinear. An algebra V is said to be unital if it contains an identity element with respect to the bilinear map $\star$.

Remark 22.7.2 (Over rings). More generally, one can define an algebra over a commutative unital ring R. The defining conditions remain the same, except that one requires V to be an R-module (Definition 2.4.53) instead of a vector space.

Definition 22.7.3 (Division algebra). A unital algebra in which every nonzero element has both a left and right multiplicative inverse. If the algebra is associative, these inverses coincide. A normed division algebra is a division algebra equipped with a multiplicative quadratic form q such that

$$\langle a \mid b \rangle := \frac{1}{2}\big[q(a + b) - q(a) - q(b)\big] \tag{22.79}$$

is a nondegenerate inner product (Definition 22.3.1).

Theorem 22.7.4 (Frobenius). *There exist three inequivalent finite-dimensional real associative division algebras: $\mathbb{R}, \mathbb{C}$ and $\mathbb{H}$.*

Theorem 22.7.5 (Hurwitz). *There exist four inequivalent finite-dimensional real normed division algebras: $\mathbb{R}, \mathbb{C}, \mathbb{H}$ and $\mathbb{O}$.*

Example 22.7.6 (Frobenius algebra). An associative algebra A equipped with a nondegenerate bilinear form $\eta : A \times A \to k$ satisfying the following condition for all

$a, b, c \in A$:

$$\eta(ab, c) = \eta(a, bc)\,. \tag{22.80}$$

Equivalently, an associative algebra (A, μ) equipped with a linear form $\varepsilon : A \to k$ such that $\varepsilon \circ \mu$ is nondegenerate.[6] A Frobenius algebra is said to be symmetric if η is symmetric.

Example 22.7.7 (Temperley–Lieb algebra). Let R be a commutative unital ring and fix an element $\delta \in R$. The Temperley–Lieb algebra $\mathsf{TL}_n(\delta)$ is the unital R-algebra with generators $\{U_i\}_{i<n}$ that satisfy the **Jones relations**:

1. $U_i^2 = \delta U_i$,
2. $U_i U_j = U_j U_i$ if $|i - j| \neq 1$, and
3. $U_i U_j U_i = U_i$ if $|i - j| = 1$.

One can represent the elements of a Temperley–Lieb algebra diagrammatically. All elements of $\mathsf{TL}_n(\delta)$ are represented as diagrams with n inputs and n outputs. The unit is given by the diagram where all inputs are connected to the outputs directly across the diagram. The generators $\{U_i\}_{i<n}$ are constructed by connecting the ith input (resp. output) to the $i+1^{\text{th}}$ input (resp. output) and all other inputs are connected to the output directly across the diagram. Multiplication in $\mathsf{TL}_n(\delta)$ is performed diagrammatically by placing two diagrams side by side. Closed loops are replaced by a factor δ.

(a) Unit in $\mathsf{TL}_4(\delta)$. (b) Generator U_2 in $\mathsf{TL}_4(\delta)$.

Figure 22.1: Temperley–Lieb algebra.

Definition 22.7.8 (Jordan algebra). A nonassociative, commutative algebra A such that

$$(xy)(xx) = x(y(xx)) \tag{22.81}$$

for all $x, y \in A$.

Property 22.7.9 (Power associativity). It can be shown that the Jordan condition implies that powers of elements are well-defined:

$$(xx)x = x(xx) =: x^3 \tag{22.82}$$

for all $x \in A$ and likeiwse for higher-order powers.

[6] A third equivalent definition will be given in Definition 28.1.7.

The original definition of a Jordan algebra does not admit a lot of intuition. However, by the power-associativity property, one also has expressions of the form

$$(x^m y)x^n = x^m(yx^n)\,. \tag{22.83}$$

By commutativity, one obtains that the multiplication maps $L_{x^m} : y \mapsto x^m y$ associated to powers commute:

$$L_{x^m}L_{x^n} = L_{x^n}L_{x^m}\,. \tag{22.84}$$

This leads to the following equivalent definition.

Alternative Definition 22.7.10. A Jordan algebra is a commutative, power-associative algebra A such that Eq. (22.84) holds for all $x \in A$.

Property 22.7.11. Every associative algebra over a field of characteristic not 2 (or over a ring in which 2 is a unit) the multiplication induces a Jordan structure as follows:

$$x \circ y := \frac{1}{2}(xy + yx)\,, \tag{22.85}$$

i.e. the Jordan product is given by the anticommutator. Jordan algebras of this form are said to be **special**, while all other Jordan algebras are said to be **exceptional**.

Definition 22.7.12 (Rota–Baxter algebra). A Rota–Baxter algebra of weight $\lambda \in k$ is a k-algebra with a linear operator $R : A \to A$ such that

$$R(x)R(y) = R\big(R(x)y\big) + R\big(xR(y)\big) + \lambda R(xy) \tag{22.86}$$

for all $x, y \in A$.

Example 22.7.13 (Integration by parts). Consider the integral operator

$$I : C(X) \to C(X) : f \mapsto \int_0^x f(y)\,dy\,. \tag{22.87}$$

The integration by parts formula 18.6.12 can be rewritten as follows:

$$I(f)I(g) = I\big(I(f)g\big) + I\big(fI(g)\big)\,, \tag{22.88}$$

showing that $C(X)$ forms a Rota–Baxter algebra of weight 0.

Property 22.7.14. If (A, R) is a Rota–Baxter algebra of weight $\lambda \in k$, so is $(A, \mathbb{1}_A - R)$. The operator $\mathbb{1}_A - R$ is often denoted by $\widetilde{R}$.

Formula 22.7.15 (Spitzer's identity). Let (A, R) be a Rota–Baxter $\mathbb{Q}$-algebra of weight 1. For all $a \in A$, the following equality holds:

$$\exp\big(R\big(\ln(1-ax)^{-1}\big)\big) = \sum_{n=0}^{+\infty} x^n \underbrace{R(R(R(\cdots R(a)a)a)a)}_{n \text{ times}}\,. \tag{22.89}$$

22.8 Grassmanians

Definition 22.8.1 (Grassmannian). Let V be a vector space. The set of all subspaces of V of dimension k is called the Grassmannian $\mathsf{Gr}(k, V)$.

Property 22.8.2. $\mathsf{GL}(V)$ acts transitively (Definition 2.3.13) on the k-dimensional subspaces of V. Property 2.3.14 implies that the coset space $\mathsf{GL}(V)/H_W$ for the stabilizer H_W of any $W \in \mathsf{Gr}(k, V)$ is isomorphic (as a set) to $\mathsf{Gr}(k, V)$. When V is an n-dimensional real vector space one can show that this quotient is isomorphic to $\mathsf{O}(n)/(\mathsf{O}(k) \times \mathsf{O}(n-k))$. For complex vector spaces the orthogonal groups should be replaced by unitary groups.

Example 22.8.3 (Projective space). Recall Definition 10.1.29. The Grassmannian $\mathsf{Gr}(1, V)$ is given by the projective space $\mathfrak{K}\mathbb{P}^{\dim(V)-1}$.

Definition 22.8.4 (Flag). Let V be a finite-dimensional vector space. A sequence of proper subspaces $V_1 < \cdots < V_n = V$ is called a flag of V. The sequence

$$\big(\dim(V_1), \ldots, \dim(V_n) = \dim(V)\big)$$

is called the **signature** of the flag. If $\forall i \leq \dim(V) : \dim(V_i) = i$, the flag is said to be **complete**.

Grassmannians are a specific instance of the following object.

Definition 22.8.5 (Flag variety). The set of all flags of a given signature is called the (generalized) flag variety (of that signature). If the underlying field is the field of real (or complex) numbers, the flag variety is a *smooth* (*or complex*) *manifold* (see Chapter 31), called the **flag manifold**.

Property 22.8.2 generalizes as follows.

Property 22.8.6 (Parabolic subgroups). Every flag variety has the structure of a homogeneous space: $\mathsf{Fl}_{n,\underline{d}} = \mathsf{GL}(V)/P_{n,\underline{d}}$, where $\underline{d}$ denotes the signature of the flags. The subgroups $P_{n,\underline{d}}$ are called **parabolic subgroups**. The maximal parabolic subgroups are those that define the Grassmannian varieties. The flag variety of all complete flags defines the **Borel subgroup** B_n. It can be shown that every parabolic subgroup contains the Borel subgroup.

Chapter 23

Vector & Tensor Calculus

References for this chapter are Choquet-Bruhat et al. (1991); Jeevanjee (2015). For a more geometric approach to some of the concepts and results in this chapter, see Chapter 30, Chapter 34 and Section 34.8 (e.g. Remark 36.1.9).

Remark. In this chapter, a **vector field** will mean a vector-valued function $\vec{\mathbf{A}} : \mathbb{R}^n \to \mathbb{R}^n$ with smooth projections.

Chapter Contents

23.1 Nabla-operator **460**

23.2 Integration **462**

23.2.1 Line integrals 462

23.2.2 Integral theorems 463

23.3 Tensors **463**

23.3.1 Tensor product 463

23.3.2 Transformation rules 465

23.3.3 Tensor operations 466

23.4 Exterior algebra **467**

23.4.1 Antisymmetric tensors 467

23.4.2 Determinant 468

23.4.3 Levi-Civita symbol 468

23.4.4 Wedge product 469

23.4.5 Exterior algebra 470

23.4.6 Hodge star 472

23.5 Graded vector spaces **474**

23.5.1 Supermatrices 477

23.5.2 Berezin calculus 478

23.1 Nabla-operator

Definition 23.1.1 (Gradient). Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a smooth function.

$$\nabla\varphi := \left(\frac{\partial\varphi}{\partial x^1}, \dots, \frac{\partial\varphi}{\partial x^n}\right) \tag{23.1}$$

Property 23.1.2. The gradient of a smooth real-valued function is perpendicular to its level sets (Definition 1.5.8).

Definition 23.1.3 (Directional derivative). Consider a smooth function $\varphi : \mathbb{R}^n \to \mathbb{R}$ and let $\hat{a}$ be a unit vector. The directional derivative $\nabla_{\hat{a}}\varphi$ is defined as the change of the function φ in the direction of $\hat{a}$:

$$\nabla_{\hat{a}}\varphi := (\hat{a} \cdot \nabla)\varphi\,. \tag{23.2}$$

Example 23.1.4. Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a smooth function and let $\frac{\mathrm{d}\vec{\mathbf{r}}}{\mathrm{d}s}$ denote the tangent vector to a curve $\vec{\mathbf{r}}(s)$ with *natural parameter* (see Definition 30.1.3). The variation of φ along $\vec{\mathbf{r}}(s)$ is given by

$$\frac{\partial\varphi}{\partial s} = \frac{\mathrm{d}\vec{\mathbf{r}}}{\mathrm{d}s} \cdot \nabla\varphi\,. \tag{23.3}$$

Definition 23.1.5 (Conservative vector field). A vector field that can be expressed as the gradient of a scalar function.

Definition 23.1.6 (Gradient of a tensor). Let T be a tensor field on $\mathbb{R}^n$ and let $\vec{\mathbf{e}}_i$ be the coordinate basis. The gradient of T is defined as follows:

$$\nabla T := \sum_{i=1}^{n} \frac{\partial T}{\partial x^i} \otimes \vec{\mathbf{e}}_i\,. \tag{23.4}$$

Definition 23.1.7 (Divergence). Let $\vec{\mathbf{A}}$ be a vector field on $\mathbb{R}^n$.

$$\nabla \cdot \vec{\mathbf{A}} := \sum_{i=1}^{n} \frac{\partial A_i}{\partial x^i} \tag{23.5}$$

Definition 23.1.8 (Solenoidal vector field). A vector field $\vec{\mathbf{A}}$ that satisfies

$$\nabla \cdot \vec{\mathbf{A}} = 0\,. \tag{23.6}$$

Such a vector field is also said to be **divergence-free** because of Eq. (23.13) below.

Definition 23.1.9 (Rotor / curl). Let $\vec{\mathbf{A}}$ be a vector field on $\mathbb{R}^3$.

$$\nabla \times \vec{\mathbf{A}} := \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}, \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}, \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}\right) \tag{23.7}$$

Definition 23.1.10 (Irrotational vector field). A vector field $\vec{\mathbf{A}}$ that satisfies

$$\nabla \times \vec{\mathbf{A}} = 0\,. \tag{23.8}$$

Definition 23.1.11 (Laplacian). Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a smooth function.

$$\Delta\varphi := \nabla^2\varphi = \frac{\partial^2\varphi}{\partial x^2} + \frac{\partial^2\varphi}{\partial y^2} + \frac{\partial^2\varphi}{\partial z^2} \tag{23.9}$$

For a vector field $\vec{\mathbf{A}}$ on $\mathbb{R}^3$ one can define a **vector Laplacian**:

$$\Delta\vec{\mathbf{A}} := \nabla^2\vec{\mathbf{A}} = \nabla\left(\nabla\cdot\vec{\mathbf{A}}\right) - \nabla\times\left(\nabla\times\vec{\mathbf{A}}\right) \tag{23.10}$$

$$= \left(\Delta A_x, \Delta A_y, \Delta A_z\right). \tag{23.11}$$

Property 23.1.12 (Mixed properties). The differential operators introduced above satisfy the following identities:

$$\nabla\times\left(\nabla\varphi\right) = 0\,, \tag{23.12}$$

$$\nabla\cdot\left(\nabla\times\vec{\mathbf{A}}\right) = 0\,. \tag{23.13}$$

Corollary 23.1.13. All conservative vector fields (on $\mathbb{R}^3$) are irrotational. However, the converse is only true if the domain is simply-connected (Definition 7.1.18). All of this is formalized by *Poincaré lemma* 34.9.8.

Formula 23.1.14 (Helmholtz decomposition). If $\vec{\mathbf{A}}$ is a vector field on $\mathbb{R}^3$ that decays faster than $1/r$ when $r \longrightarrow +\infty$, it can be written as

$$\vec{\mathbf{A}} = \nabla\times\vec{\mathbf{B}} + \nabla\varphi \tag{23.14}$$

for some smooth vector field $\vec{\mathbf{B}}$ and smooth function φ.

The differential operators introduced above can also be generalized to curvilinear coordinates. To this end one needs the *scale factors* as formally defined in Definition 30.2.5. For the remainder of this section the Einstein summation convention will not be used to make everything as explicit as possible.

Formula 23.1.15 (Unit vectors).

$$\frac{\partial\vec{\mathbf{r}}}{\partial q^i} = h_i\hat{e}_i \tag{23.15}$$

Formula 23.1.16 (Gradient).

$$\nabla\varphi = \sum_{i=1}^{n}\frac{1}{h_i}\frac{\partial\varphi}{\partial q^i}\hat{e}_i \tag{23.16}$$

Formula 23.1.17 (Divergence).

$$\nabla\cdot\vec{\mathbf{A}} = \frac{1}{\prod_{i=1}^{n}h_i}\sum_{i=1}^{n}\left(\frac{\partial A_i\prod_{j\neq i}h_j}{\partial q^i}\right) \tag{23.17}$$

Formula 23.1.18 (Rotor).

$$\left(\nabla \times \vec{\mathbf{A}}\right)_i = \sum_{j,k=1}^{3} \frac{\varepsilon_{ijk}}{h_j h_k} \left(\frac{\partial}{\partial q^j}(A_k h_k) - \frac{\partial}{\partial q^k}(A_j h_j) \right), \tag{23.18}$$

where ε_{ijk} is the 3-dimensional *Levi-Civita symbol* (see Definition 23.4.8).

Formula 23.1.19 (Laplacian in curvilinear coordinates). In general the Laplace operator is defined as

$$\Delta f := \nabla \cdot \nabla f \,. \tag{23.19}$$

The Laplacian can also be expressed in different coordinate systems:

- Cylindrical coordinates (ρ, ϕ, z):

$$\frac{1}{\rho}\frac{\partial}{\partial \rho}\left(\rho \frac{\partial}{\partial \rho}\right) + \frac{1}{\rho^2}\frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2} \,. \tag{23.20}$$

- Spherical coordinates (r, ϕ, θ):

$$\frac{1}{r^2}\left[\frac{\partial}{\partial r}\left(r^2 \frac{\partial}{\partial r}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial \phi^2} + \frac{1}{\sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial}{\partial \theta}\right)\right] . \tag{23.21}$$

23.2 Integration

23.2.1 Line integrals

Formula 23.2.1 (Line integral of a continuous function). Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function and let Γ be a piecewise smooth curve $\vec{\boldsymbol{\varphi}} : [a, b] \to \mathbb{R}^n$. The line integral of f along Γ is defined as follows:

$$\int_\Gamma f \, ds := \int_a^b f(\vec{\boldsymbol{\varphi}}(t)) \, \|\vec{\boldsymbol{\varphi}}'(t)\| \, dt \,. \tag{23.22}$$

Formula 23.2.2 (Line integral of a continuous vector field). Let $\vec{\mathbf{F}}$ be a continuous vector field on $\mathbb{R}^n$ and let Γ be a piecewise smooth curve $\vec{\boldsymbol{\varphi}} : [a, b] \to \mathbb{R}^n$. The line integral of $\vec{\mathbf{F}}$ along Γ is defined as follows:

$$\int_\Gamma \vec{\mathbf{F}} \cdot d\vec{\mathbf{s}} := \int_a^b \vec{\mathbf{F}}(\vec{\boldsymbol{\varphi}}(t)) \cdot \vec{\boldsymbol{\varphi}}'(t) \, dt \,. \tag{23.23}$$

Property 23.2.3 (Conservative vector fields). A vector field is conservative if and only if its line integral is path-independent, i.e. if it only depends on the values at the end points. This is a corollary of *Stokes's theorem* 34.8.25.

23.2.2 Integral theorems[1]

Theorem 23.2.4 (Fundamental theorem of calculus for line integrals). *Let $\Gamma : \mathbb{R} \to \mathbb{R}^n$ be a piecewise smooth curve defined on the interval $[a, b]$.*

$$\int_\Gamma \nabla f \cdot d\vec{\mathbf{r}} = \varphi(\Gamma(b)) - \varphi(\Gamma(a)) \tag{23.24}$$

Theorem 23.2.5 (Kelvin–Stokes theorem). *Let $\vec{\mathbf{A}}$ be a vector field defined on $\mathbb{R}^3$ and let S be a smooth surface with boundary ∂S.*

$$\oint_{\partial S} \vec{\mathbf{A}} \cdot d\vec{\mathbf{l}} = \iint_S \left(\nabla \times \vec{\mathbf{A}}\right) \cdot d\vec{\mathbf{S}} \tag{23.25}$$

Theorem 23.2.6 (Divergence theorem[2]). *Let $\vec{\mathbf{A}}$ be a vector field defined on $\mathbb{R}^n$.*

$$\oiint_{\partial V} \vec{\mathbf{A}} \cdot d\vec{\mathbf{S}} = \iiint_V \left(\nabla \cdot \vec{\mathbf{A}}\right) dV \tag{23.26}$$

Corollary 23.2.7 (Green's identity). Let ϕ, ψ be smooth real-valued functions defined on $\mathbb{R}^3$.

$$\oiint_{\partial V} \left(\psi \nabla \phi - \phi \nabla \psi\right) \cdot d\vec{\mathbf{S}} = \iiint_V \left(\psi \nabla^2 \phi - \phi \nabla^2 \psi\right) dV \tag{23.27}$$

23.3 Tensors

23.3.1 Tensor product

There are two possible (equivalent) ways to introduce the concept of a 'tensor' on finite-dimensional vector spaces. One is to interpret tensors as multilinear maps, while the other is to work in a local fashion and express work with the expansion coefficients with respect to a chosen basis.

Definition 23.3.1 (Tensor product space). The tensor product of two finite-dimensional vector spaces V and W is defined as[3] the set of bilinear maps on the Cartesian product $V^* \times W^*$. Let v, w be vectors in respectively V and W and let g, h be vectors in the corresponding dual spaces. The tensor product of v and w is then defined as follows:

$$(v \otimes w)(g, h) := v(g)w(h)\,. \tag{23.28}$$

In this incarnation the tensor product is sometimes known as the **outer product**. Outer products are also frequently called **pure** or **simple tensors**.

[1] These theorems follow from a more general theorem by Stokes (see Theorem 34.8.25).

[2] Also known as **Gauss's theorem** or the **Gauss–Ostrogradsky theorem**.

[3] '*isomorphic to*' would be better terminology. See the universal property below (Universal Property 23.3.3).

Definition 23.3.2 (Tensor component). Let $\mathbf{T}$ be a tensor that takes r vectors and s covectors as input and returns a scalar (element of the underlying field). The components of $\mathbf{T}$ with respect to a frame $\{e_i\}_{i\leq n}$ and a coframe $\{e^i\}_{i\leq n}$ are defined as $T_{i\dots j}{}^{k\dots l} := \mathbf{T}(e_i, \dots, e_j, e^k, \dots, e^l)$.

The above definition can be restated as a universal property (this is also the right way to generalize tensors to infinite-dimensional spaces and avoid the awkward definition involving dual spaces).

Universal Property 23.3.3. Let Z be a vector space. For every bilinear map $T : V \times W \to Z$, there exists a unique linear map $f : V \otimes W \to Z$ such that $T = f \circ \otimes$.

Corollary 23.3.4. The tensor product is unique up to linear isomorphisms. This results in the commutativity of the tensor product:

$$V \otimes W \cong W \otimes V\,. \tag{23.29}$$

Notation 23.3.5 (Tensor power).

$$V^{\otimes n} := \underbrace{V \otimes \cdots \otimes V}_{n \text{ copies}} \tag{23.30}$$

More generally, the tensor product of r copies of V and s copies of V^* is denoted by

$$\mathcal{T}^r_s(V) = V^{\otimes r} \otimes V^{*\otimes s}\,. \tag{23.31}$$

Tensors in this space are said to be of **type** (r, s).

Definition 23.3.6 (Scalar). The elements of the underlying field. These are, by definition, the $(0, 0)$-tensors.

Definition 23.3.7 (Tensor algebra). The tensor algebra over a vector space V is defined as follows:

$$T(V) := \bigoplus_{k=0}^{+\infty} V^{\otimes k}\,. \tag{23.32}$$

The following remark is strongly related to Property 22.2.18.

Remark 23.3.8. For finite-dimensional vector spaces, $\mathcal{T}^1_1 V$ is isomorphic to $\mathsf{End}(V)$ and $\mathcal{T}^1_0 V$ is isomorphic to V itself. However, when including infinite-dimensional spaces, $\mathcal{T}^1_1 V$ is only isomorphic to the endomorphism space $\mathsf{End}(V^*)$ of the dual. This isomorphism is given by the map $\hat{T} : V^* \to V^* : \omega \mapsto T(-, \omega)$ for every $T \in \mathcal{T}^1_1 V$. Dually, in this general setting, the spaces $T^0_1 V$ and V^* are also isomorphic.

The tensor product space can also be algebraically defined as follows.

Alternative Definition 23.3.9 (Tensor product). Consider two vector spaces V, W over a field K. First, construct the free vector space $F(V \times W)$ over K. Then, construct the subspace N of $F(V \times W)$ spanned by elements of the form

- $(v + v', w) - (v, w) - (v', w)$,
- $(v, w + w') - (v, w) - (v, w')$,
- $(\lambda v, w) - \lambda(v, w)$, or
- $(v, \mu w) - \mu(v, w)$,

where $v \in V, w \in W$ and $\lambda, \mu \in K$. The tensor product $V \otimes W$ is defined as the quotient $F(V \times W)/N$. It can be shown that this construction is associative, i.e. $U \otimes (V \otimes W) \cong (U \otimes V) \otimes W$, and as such these brackets will be omitted in all expressions.

Now, consider the case where $W = V^*$. In this case, the basis of the tensor product $\mathcal{T}^r_s(V)$ will be denoted by

$$\underbrace{e_i \otimes \cdots \otimes e_j}_{r \text{ basis vector}} \otimes \underbrace{\varepsilon^k \otimes \cdots \otimes \varepsilon^l}_{s \text{ dual basis vectors}}$$

and the expansion coefficients will be denoted by $T_{i\ldots j}{}^{k\ldots l}$.

Property 23.3.10 (Dimension). From the previous construction, it follows that the dimension of $\mathcal{T}^r_s(V)$ is equal to rs.

For completeness the proof that the values of the tensor operating on r basis vectors and s basis covectors are equal to the corresponding expansion coefficients is given.

Proof. Consider a general tensor $\mathbf{T} = T_{i\ldots j}{}^{k\ldots l} e_k \otimes \cdots \otimes e_l \otimes \varepsilon^i \otimes \cdots \otimes \varepsilon^j$. Combining Definition 23.3.1 and the pairing of dual vectors (22.17) gives

$$\begin{aligned}
\mathbf{T}(\varepsilon^m, \ldots, \varepsilon^n, e_a, \ldots, e_b) &= T_{i\ldots j}{}^{k\ldots l} e_k(\varepsilon^m) \ldots e_l(\varepsilon^n) \varepsilon^i(e_a) \ldots \varepsilon^j(e_b) \\
&= T_{i\ldots j}{}^{k\ldots l} \delta^m_k \ldots \delta^n_l \delta^i_a \ldots \delta^j_b \\
&= T_{a\ldots b}{}^{m\ldots n} .
\end{aligned}$$

□

23.3.2 Transformation rules

In this section, the behaviour of tensors under basis transformations of the form

$$e'_i = A^i{}_j e_j \tag{23.33}$$

is considered.

Definition 23.3.11 (Contravariant). A tensor component that transforms by the following rule is said to be contravariant:

$$v^i = A^i{}_j v'^j . \tag{23.34}$$

Definition 23.3.12 (Covariant). A tensor component that transforms by the following rule is said to be covariant:

$$p'_i = A^j{}_i p_j \,. \tag{23.35}$$

Example 23.3.13 (Mixed tensor). As an example the transformation rule of a mixed third-order tensor $T \in \mathcal{T}_2^1$ is given:

$$T^k{}_{ij} = A^k{}_w (A^{-1})^u{}_i (A^{-1})^v{}_j T'^w{}_{uv} \,. \tag{23.36}$$

The following rule is a useful substitute for the 'illegal' division of tensors.

Method 23.3.14 (Quotient rule). Consider an equation such as $Q_i{}^j A_{jl}{}^k = B_{il}{}^k$, where A and B are two known tensors. The quotient rule asserts the following: *"If the equation holds under all transformations, then Q is a tensor."* Note that this rule does not necessarily hold when $B = 0$ because transformation rules are not well-defined for the zero tensor.

23.3.3 Tensor operations

Definition 23.3.15 (Contraction). Let A be a tensor of type (m, n). Taking a subscript and superscript to be equal and summing over all possible values of this index gives a new tensor of type $(m-1, n-1)$. This operation is called the contraction of A. It is induced by the evaluation map/pairing (Definition 22.2.20).

Example 23.3.16. Let A be a tensor of type $(1, 2)$. The contraction over the first and third index can be written using the Einstein conventionas follows:

$$A^j{}_{ij} = \sum_{j=1}^{n} A^j{}_{ij} \,. \tag{23.37}$$

Definition 23.3.17 (Direct product). The tensor constructed by the componentwise multiplication of two tensors.

Example 23.3.18. Let $A^i{}_k$ and $B^j{}_{lm}$ represent two tensors. Their direct product is equal to

$$C^{i\,j}{}_{k\,lm} = A^i{}_k B^j{}_{lm} \,.$$

Example 23.3.19 (Kronecker product). It is also possible to combine operators acting on different vector spaces to make them act on the tensor product space:

$$(A \otimes B)(v \otimes w) := Av \otimes Bw \,. \tag{23.38}$$

This should not be confused with the Hadamard product (Formula 22.4.8).

Notation 23.3.20 (Abuse of notation). Consider an operator A acting on a vector space V_1. This operator can be extended to any tensor product space $V_1 \otimes V_2$ as $A \otimes \mathbb{1}$. However, it is often still denoted by A.

Notation 23.3.21 (Symmetric part). Consider a second-order tensor T (here taken to be of covariant type for notational simplicity). The symmetric and antisymmetric part of T are often denoted by

$$T_{(ij)} := \frac{1}{2}\big(T_{ij} + T_{ji}\big) \tag{23.39}$$

and

$$T_{[ij]} := \frac{1}{2}\big(T_{ij} - T_{ji}\big)\,. \tag{23.40}$$

This notation is easily generalized to other types of tensors.

Formula 23.3.22 (Gradient of tensor products). The gradient of an outer product is defined through the Leibniz rule:

$$\nabla \cdot (v \otimes w) := (\nabla \cdot v)w + (v \cdot \nabla)w\,. \tag{23.41}$$

Definition 23.3.23 (Complexification). Let V be a real vector space. The complexification of V is defined as follows:

$$V^{\mathbb{C}} := V \otimes \mathbb{C}\,. \tag{23.42}$$

This space can still be considered a real vector space, but it can also be turned into a complex vector space by generalizing the scalar product:

$$\alpha(v \otimes \beta) := v \otimes (\alpha\beta) \tag{23.43}$$

for all $\alpha \in \mathbb{C}$.

Property 23.3.24. By (multi)linearity, every element $v_{\mathbb{C}} \in V^{\mathbb{C}}$ can be written as

$$v_{\mathbb{C}} = (v_1 \otimes 1) + i(v_2 \otimes 1)\,.$$

Therefore, the complexification can be (formally) decomposed as

$$V^{\mathbb{C}} \cong V \oplus iV\,. \tag{23.44}$$

23.4 Exterior algebra

23.4.1 Antisymmetric tensors

Definition 23.4.1 (Antisymmetric tensor). A tensor that changes sign under the interchange of any two indices.

Notation 23.4.2 (Symmetric tensors). The space of symmetric $(n, 0)$-tensors is denoted by $S^n(V)$.

Notation 23.4.3 (Antisymmetric tensors). The space of antisymmetric $(n, 0)$-tensors is denoted by $\Lambda^n(V)$.

Property 23.4.4. Let $n = \dim(V)$. The space $\Lambda^r(V)$ equals the zero space for all $r \geq n$.

23.4.2 Determinant

Definition 23.4.5 (Form). An n-form on a vector space V is a totally antisymmetric element of $\mathcal{T}_n^0 V$, where $n \leq \dim(V)$.

Definition 23.4.6 (Volume form). A form of rank $\dim(V)$ is also called a **top form** or **volume form**.

Definition 23.4.7 (Determinant). Consider a finite-dimensional vector space V with basis $\{e_i\}_{i \leq n}$. Let φ be a tensor in $\mathcal{T}_1^1 V \cong \mathrm{End}(V)$ and let ω be a volume form on V. The determinant of φ is defined as follows:

$$\det(\varphi) := \frac{\omega\big(\varphi(e_1), \dots, \varphi(e_n)\big)}{\omega(e_1, \dots, e_n)} . \tag{23.45}$$

This definition is well-defined, i.e. it is independent of the choice of volume form and basis. Furthermore, it coincides with Definition 22.4.46.

One should note that the determinant is only well-defined for $(1,1)$-tensors. Although other types of tensors can also be represented as matrices, for these the above formula would not be independent of a choice of basis anymore. A more general concept can be defined using the language of *fibre bundles* (see Chapter 35).

23.4.3 Levi-Civita symbol

Definition 23.4.8 (Levi-Civita symbol). In $n \in \mathbb{N}$ dimensions, the Levi-Civita symbol is defined as follows:

$$\varepsilon_{i_1 \dots i_n} = \begin{cases} 1 & \text{if } (i_1 \dots i_n) \text{ is an even permutation of } (1 \dots n)\,, \\ -1 & \text{if } (i_1 \dots i_n) \text{ is an odd permutation of } (1 \dots n)\,, \\ 0 & \text{if any of the indices occurs more than once.} \end{cases} \tag{23.46}$$

Remark 23.4.9 (Pseudotensor). The Levi-Civita symbol is not a tensor, it is a *pseudotensor*. This means that the sign changes under reflections or, more generally, any transformation with negative determinant. (Given a metric tensor g, one can turn it into a proper tensor by multiplying it by $\sqrt{g}$.)

Formula 23.4.10 (Contraction).

$$\varepsilon_{i_1,\dots,i_k} \varepsilon^{j_1,\dots,j_k} = \delta^{j_1,\dots,j_k}_{i_1,\dots,i_k} \tag{23.47}$$

and

$$\varepsilon_{i_1,\dots,i_k,i_{k+1},\dots,i_l} \varepsilon^{i_1,\dots,i_k,j_{k+1},\dots,j_l} = k!\,\delta^{j_{k+1},\dots,j_l}_{i_{k+1},\dots,i_l} \tag{23.48}$$

Definition 23.4.11 (Cross product). Using the Levi-Civita symbol, one can define the ith component of the cross product as follows:

$$(v \times w)_i = \sum_{j,k=1}^{3} \varepsilon_{ijk} v_j w_k \,. \tag{23.49}$$

The previous remark implies that the cross product is in fact not a vector, instead it is a pseudovector.

Remark 23.4.12 (Hurwitz theorem). The cross product actually exists in four cases: $\mathbb{R}^0$, $\mathbb{R}^1$, $\mathbb{R}^3$ and $\mathbb{R}^7$. In general, it is characterized by the following conditions:

1. **Bilinearity**: $(\lambda v) \times (\kappa w) = \lambda\kappa(v \times w)$.
2. **Orthogonality**: $v \cdot (v \times w) = 0 = w \cdot (v \times w)$.
3. **Magnitude**: $\|v \times w\|^2 = \|v\|^2\|w\|^2 - (v \cdot w)^2$.

These conditions imply that on $\mathbb{R}^1$ the cross product is identically zero. However, on $\mathbb{R}^3$ and $\mathbb{R}^7$ one obtains an anticommutative bilinear operation. On $\mathbb{R}^3$, it is unique, while on $\mathbb{R}^7$ different choices exist.

This construction is related to the Hurwitz classification theorem 22.7.5, since one can construct the cross products on $\mathbb{R}^n$ by embedding it as the imaginary part of the (real, normed) division algebra of dimension $n+1$. The cross product is then obtained from the algebra product after discarding the real component. For example, for $\mathbb{R}^1$ embedded in $\mathbb{C}$, one obtains a product of two purely imaginary numbers, which is real. Discarding this component gives exactly zero, as mentioned above.

Property 23.4.13 (Exceptional Lie group G_2). Using the vector product, one can define an associative 3-form, the **triple product**:

$$u \otimes v \otimes w \mapsto u \cdot (v \times w) \,. \tag{23.50}$$

The group of linear isomorphisms that preserve the triple product on $\mathbb{R}^7$ is denoted by G_2. By the relation between vector products and division algebras, this group is also the automorphism group of the octonions.

23.4.4 Wedge product

Definition 23.4.14 (Antisymmetrization). Let S_k denote the permutation group on $k \in \mathbb{N}$ elements (Definition 2.2.27). The antisymmetrization operator is defined as follows:

$$\mathsf{Alt}(e_1 \otimes \cdots \otimes e_k) := \sum_{\sigma \in S_k} \mathsf{sgn}(\sigma) e_{\sigma(1)} \otimes \cdots \otimes e_{\sigma(k)} \,. \tag{23.51}$$

Note that many authors introduce a factor $1/k!$. This convention is not adopted here to keep the subsequent constructions clean. If the factor is included, Formula 23.4.17 below should be modified.

Definition 23.4.15 (Wedge product). Let $\{e_i\}_{i\leq \dim(V)}$ be a basis for V. The wedge product of basisvectors is defined as follows:

$$e_1 \wedge \ldots \wedge e_k := \mathsf{Alt}(e_1 \otimes \cdots \otimes e_k)\,. \tag{23.52}$$

From this definition, it immediately follows that the wedge product is (totally) antisymmetric.

Construction 23.4.16. Let $\{e_i\}_{i\leq \dim(V)}$ be a basis for V. The above definition implies that a basis for $\Lambda^r(V)$ is given by

$$\{e_{i_1} \wedge \ldots \wedge e_{i_r} \mid \forall k \leq r : 1 \leq i_k \leq \dim(V)\}\,. \tag{23.53}$$

Accordingly, the dimension of this space is given by

$$\dim \Lambda^r(V) = \binom{n}{r}\,. \tag{23.54}$$

For $r = 0$, this construction would be vacuous, so one just defines $\Lambda^0(V) := \mathbb{R}$.

Formula 23.4.17. Let $v \in \Lambda^r(V)$ and $w \in \Lambda^m(V)$. The wedge product 23.4.15 can be generalized as follows:

$$v \wedge w := \frac{1}{r!m!}\,\mathsf{Alt}(v \otimes w)\,. \tag{23.55}$$

Definition 23.4.18 (Blade). Elements of $\Lambda^k(V)$ that can be written as the wedge product of k vectors are known as k-blades or **pure k-vectors**.

Formula 23.4.19 (Cross product). In dimension 3, there exists an important isomorphism $J : \Lambda^2(\mathbb{R}^3) \to \mathbb{R}^3$:

$$J(\lambda)^i = \frac{1}{2}\varepsilon^i{}_{jk}\lambda^{jk}\,, \tag{23.56}$$

where $\lambda \in \Lambda^2(\mathbb{R}^3)$. See also the Hodge $*$-operator further below (Definition 23.4.27).

Looking at Definition 23.4.11 of the cross product, one can see that $v \times w$ is actually the same as $J(v \wedge w)$. One can thus use the wedge product to generalize the cross product to arbitrary dimensions.

23.4.5 Exterior algebra

Definition 23.4.20 (Exterior power). In the theory of tensor calculus, the space $\Lambda^k(V)$ is often called the kth exterior power of V. As mentioned before, its elements are called (exterior) k-forms.

Definition 23.4.21 (Exterior algebra). One can define a *graded vector space* (see Definition 23.5.1) as follows:

$$\Lambda^\bullet(V) := \bigoplus_{k\geq 0} \Lambda^k(V)\,. \tag{23.57}$$

This graded vector space can be turned into a graded algebra by taking the wedge product as the multiplication:

$$\wedge : \Lambda^k(V) \times \Lambda^l(V) \to \Lambda^{k+l}(V)\,. \tag{23.58}$$

This algebra is called the exterior algebra or **Grassmann algebra** over V. Elements of the space $\otimes_{k\in 2\mathbb{N}}\Lambda^k(V)$ are said to be **Grassmann-even** and elements of $\otimes_{k\in 2\mathbb{N}+1}\Lambda^k(V)$ are said to be **Grassmann-odd**.

Alternative Definition 23.4.22. Let $T(V)$ be the tensor algebra (Definition 23.3.7) over the vector space V, i.e.

$$T(V) = \bigoplus_{k\geq 0} V^{\otimes k}\,. \tag{23.59}$$

The exterior algebra $\Lambda^\bullet(V)$ over V is defined as the quotient of $T(V)$ by the two-sided ideal I generated by the elements $\{v \otimes v \mid v \in V\}$.

Proof (Proof of equivalence). Consider the equality

$$(u+v)\otimes(u+v) - u\otimes u - v\otimes v = u\otimes v + v\otimes u$$

The left-hand side is an element of the ideal I generated by $\{v \otimes v \mid v \in V\}$. Using the ideal generated by elements of the form of the right-hand side gives the usual definition of the exterior algebra based on the wedge product as defined in Definition 23.4.15 because it imposes the relation $u \wedge v = -v \wedge u$.

However, one should pay attention to one little detail. As mentioned in Definition 23.4.22 the general definition uses the ideal I to construct the quotient space. The other construction is only equivalent when working over a field with characteristic different from 2. This follows from the fact that one has to divide by 2 when trying to obtain the ideal I from the right-hand side when setting $u = v$. □

Property 23.4.23 (Graded-commutativity). The exterior algebra is both a unital associative algebra (with identity $1 \in K$) and a coalgebra. Furthermore, it is also commutative in the graded sense (see Definition 23.5.5).

Property 23.4.24 (Nilpotency). Graded-commutativity implies that the wedge product of any odd exterior form with itself is identically 0. The wedge product of an even exterior form with itself vanishes if and only if the form can be decomposed as a product of one-forms, i.e. if it is pure.

23.4.6 Hodge star

Equation (23.54) says that the spaces $\Lambda^k(V)$ and $\Lambda^{n-k}(V)$ have the same dimension and, hence, that there exists a linear isomorphism between them. Such an isomorphism is given by the Hodge star operator if one restricts to vector spaces equipped with a nondegenerate Hermitian form (Remark 22.3.15).

When equipped with an inner product and, hence, an orthonormal basis $\{e_i\}_{i\leq \dim(V)}$, every finite-dimensional vector space admits a canonical volume form given by

$$\mathsf{Vol} = e_1 \wedge \ldots \wedge e_n \,. \tag{23.60}$$

This convention will also be adopted in the remainder of this section.

Definition 23.4.25 (Orientation). Let $\mathsf{Vol}(V)$ be the volume form on the vector space V as defined above. From the definition of a volume form it follows that every other $\dim(V)$-form is a scalar multiple of $\mathsf{Vol}(V)$. Denote this number by r. This also implies that a choice of volume form induces an equivalence relation on top-dimensional forms, where equivalence class under this relation are called orientations, as follows: If $r > 0$, the orientation is said to be **positive** and, if $r < 0$, the orientation is said to be **negative**.

Formula 23.4.26 (Inner product). Let V be equipped with an inner product $\langle \cdot \mid \cdot \rangle$. One can extend this to an inner product on $\Lambda^k(V)$ by first defining it on blades and extending it by linearity:

$$\langle v_1 \wedge \ldots \wedge v_k \mid w_1 \wedge \ldots \wedge w_k \rangle_k := \det(\langle v_i \mid w_j \rangle) \,. \tag{23.61}$$

For an orthogonal basis this formula factorizes as follows:

$$\langle v_1 \wedge \ldots \wedge v_k \mid w_1 \wedge \ldots \wedge w_k \rangle_k = \langle v_1 \mid w_1 \rangle \cdots \langle v_k \mid w_k \rangle \,. \tag{23.62}$$

Definition 23.4.27 (Hodge star). The Hodge star $* : \Lambda^k(V) \to \Lambda^{n-k}(V)$ is defined as the unique isomorphism such that, for all $\omega \in \Lambda^k(V)$ and $\rho \in \Lambda^{n-k}(V)$, the following equality holds:

$$\omega \wedge \rho = \langle *\omega \mid \rho \rangle_{n-k} \, \mathsf{Vol}(V) \,, \tag{23.63}$$

where $\langle \cdot \mid \cdot \rangle_{n-k}$ is the inner product (23.4.26) on $\Lambda^{n-k}(V)$. The element $*\omega$ is often called the **(Hodge) dual** of ω.

Proof. Fix an element $\omega \in \Lambda^k(V)$. For every element $\rho \in \Lambda^{n-k}(V)$ one can see that $\omega \wedge \rho$ is an element of $\Lambda^n(V)$ and as such it is a scalar multiple of $\mathsf{Vol}(V)$. This implies that it can be written as

$$c_\omega(\rho) \, \mathsf{Vol}(V) \,.$$

The map $c_\omega : \Lambda^{n-k}(V) \to \mathbb{R} : \rho \mapsto c_\omega(\rho)$ is a bounded (and thus continuous) linear

map, so *Riesz's representation theorem* 25.2.7 can be applied to identify c_ω with a unique element $*\omega \in \Lambda^{n-k}(V)$ such that

$$c_\omega(\rho) = \langle *\omega \mid \rho \rangle_{n-k} . \qquad \square$$

Formula 23.4.28. Let $\{e_i\}_{i \leq n}$ be a positively oriented orthonormal basis for V. An explicit formula for the Hodge star is given by the following construction. Let $\{i_1, \dots, i_k\}$ and $\{j_1, \dots, j_{n-k}\}$ be two ordered complementary index sets and consider an element $\omega \equiv e_{i_1} \wedge \dots \wedge e_{i_k} \in \Lambda^k(V)$.

$$*\omega = \mathrm{sgn}(\tau) \prod_{m=1}^{n-k} \langle e_{j_m} \mid e_{j_m} \rangle e_{j_1} \wedge \dots \wedge e_{j_{n-k}} , \tag{23.64}$$

where τ is the permutation that maps $e_{i_1} \wedge \dots \wedge e_{i_k} \wedge e_{j_1} \wedge \dots \wedge e_{j_{n-k}}$ to $\mathrm{Vol}(V)$.

Using this formula, one can easily prove the following important property.

Property 23.4.29. Consider an inner product space V. The Hodge dual is involutive up to a factor:

$$* * \omega = (-1)^{k(n-k)} \omega , \tag{23.65}$$

where $\omega \in \Lambda^k(V)$.

Taking the definition of the Hodge star operator together with the above property leads to the following formula (which is often found in the literature as the definition of the Hodge dual).

Formula 23.4.30. For all $\omega, \rho \in \Lambda^k(V)$, the Hodge star operator satisfies the following formula:

$$\omega \wedge *\rho = \langle \omega \mid \rho \rangle \, \mathrm{Vol}(V) . \tag{23.66}$$

Corollary 23.4.31. Consider three vectors $u, v, w \in \mathbb{R}^3$.

$$*(v \wedge w) = v \times w \tag{23.67}$$

$$*(v \times w) = v \wedge w \tag{23.68}$$

$$*(u \wedge v \wedge w) = u \cdot (v \times w) \tag{23.69}$$

Remark 23.4.32. Formula 23.4.19 is an explicit evaluation of the first equation (23.67).

Proof. The signs $\mathrm{sgn}(\sigma)$ in the definition of the wedge product can be written using the Levi-Civita symbol ε_{ijk} as defined in Definition 23.4.8. The factor $1/2$ is introduced to correct for the double counting due to the contraction over both the

indices j and k. □

Definition 23.4.33 (Selfdual form). Let V be a 4-dimensional inner product space and consider an element $\omega \in \Lambda^2(V)$. Since $*^2 = 1$ on $\Lambda^2(V)$, it has two eigenvalues: $+1$ and -1. The associated eigenvectors are said to be selfdual and antiselfdual, respectively.

Every element $\omega \in \Lambda^2(V)$ can be uniquely decomposed as the sum of a selfdual and an antiselfdual two-form:

$$\omega \equiv \omega^+ + \omega^- = \frac{1}{2}\big((\omega + *\omega) + (\omega - *\omega)\big)\,. \tag{23.70}$$

Remark 23.4.34. In Lorentzian signature, the Hodge star squares to -1 and, hence, the eigenvalues are i and $-i$. This again leads to selfdual and antiselfdual forms and the associated projections look as follows:

$$\mathsf{proj}_\pm(\omega) := \frac{1}{2}(\omega \pm i *\omega)\,. \tag{23.71}$$

23.5 Graded vector spaces

Definition 23.5.1 (Graded vector space). A vector space V that can be decomposed as

$$V = \bigoplus_{i\in I} V_i \tag{23.72}$$

for a collection of vector spaces $\{V_i\}_{i\in I}$, where I can be both finite or countable. The index i is often called the **degree** of the subspace V_i in V. One writes $\deg(v) = i$ if $v \in V_i$.

Definition 23.5.2 (Finite type). A graded vector space is said to be of finite type if it is finite-dimensional in each degree.

Example 23.5.3 (Super-vector space). A $\mathbb{Z}_2$-graded vector space.

Definition 23.5.4 (Graded algebra). A $\mathbb{Z}$-graded vector space V with the additional structure of an algebra $(V, \star)$ such that $V_k \star V_l \subseteq V_{k+l}$ for all $k, l \in \mathbb{Z}$.[4]

Definition 23.5.5 (Graded-commutative algebra). A graded algebra $(V, \star)$ such that

$$v \star w = (-1)^{\deg(v)\deg(w)} w \star v \tag{23.73}$$

holds for all homogeneous elements $v, w \in V$.

Example 23.5.6 (Superalgebra). A $\mathbb{Z}_2$-graded algebra

$$A = A_0 \oplus A_1\,, \tag{23.74}$$

such that for all $i, j \in \{0, 1\}$:

$$A_i \star A_j \subseteq A_{i+j \bmod 2}\,. \tag{23.75}$$

[4]The grading can be relaxed to any commutative monoid.

Definition 23.5.7 (Parity and suspension). Consider the category **sVect** of super-vector spaces. One can define the **parity functor** $\Pi : \mathbf{sVect} \to \mathbf{sVect}$ as the functor that interchanges even and odd subspaces:

$$\begin{aligned}(\Pi V)_0 &:= V_1\,,\\ (\Pi V)_1 &:= V_0\,.\end{aligned} \tag{23.76}$$

A more general construction exists in $\mathbb{Z}\mathbf{Vect}$. For every graded vector space V, the k-**shifted** vector space or k-**suspension** $V[k]$ is defined as follows (some authors use the opposite convention):

$$V[k]_i := V_{i-k}\,. \tag{23.77}$$

Example 23.5.8 (Free GCA). Let V be a graded vector space. The free GCA $\mathsf{Sym}^\bullet V$ on V is defined as the quotient of the tensor algebra $T(V)$ by the relations

$$x \otimes y - (-1)^{\deg(x)\deg(y)} y \otimes x = 0 \tag{23.78}$$

ranging over all homogeneous elements $x, y \in V$. This algebra can equivalently be obtained as the tensor product

$$\mathsf{Sym}^\bullet V = \mathsf{Sym}(V_{\text{even}}) \otimes \mathsf{Alt}(V_{\text{odd}})\,, \tag{23.79}$$

where Sym and Λ denote the symmetric and exterior algebras of ordinary vector spaces. It it not hard to see that this definition combines the definitions of Sym and Alt. A similar definition gives a graded alternating algebra:

$$\mathsf{Alt}^\bullet V = T(V)/\big(x \otimes y + (-1)^{\deg(x)\deg(y)} y \otimes x\big)\,. \tag{23.80}$$

Note that both of these algebras actually carry a bigrading, the total degree coming from V and the **word length**:

$$\deg(v_1 \cdots v_n) := \deg(v_1) + \cdots + \deg(v_n)\,, \tag{23.81}$$

$$\mathsf{wl}(v_1 \cdots v_n) := n\,. \tag{23.82}$$

In general, only the word length is made explicit when writing down the space, i.e. $v \in \mathsf{Alt}^{\mathsf{wl}(v)} V$.

Remark 23.5.9 (Different conventions and décalage). Some authors use the notation $\Lambda^\bullet V$ for the free graded-commutative algebra on V. However, this might be confused with the notation for the Grassmann (exterior) algebra of an ordinary vector space.[5] In fact, there is a good reason why these notations are used in a seemingly interchangeable way for graded vector spaces. The suspension functor $V \to V[1]$ gives a way to relate

[5] This inconvenient change of conventions can be found everywhere in the literature, so one should pay close attention to the conventions that are used.

the Grassmann algebra over an ordinary vector space V to the free GCA on the shifted space $V[1]$, i.e. $\mathsf{Alt}^\bullet V \cong \mathsf{Sym}^\bullet V[1]$. However, at this point, the **décalage isomorphism**

$$\mathsf{dec}_k : \Lambda^k V[k] \cong \mathsf{Sym}^k V[1] \tag{23.83}$$

is only a linear isomorphism. There are two ways to see that it can be extended to an algebra isomorphism.

The first one defines the suspension functor as an intertwiner between the symmetrization and antisymmetrization operations in the definition of Sym and Alt. Define the symmetric and antisymmetric Koszul signs of a permutation $\sigma \in S_n$ as follows:

$$\varepsilon(\sigma; v_1, \dots, v_n) := (-1)^{\#\text{ odd-odd neighbour transpositions in }\sigma}\,, \tag{23.84}$$

$$\chi(\sigma; v_1, \dots, v_n) := \mathsf{sgn}(\sigma)\varepsilon(\sigma; v_1, \dots, v_n)\,. \tag{23.85}$$

Décalage then says that the suspension functor should satisfy

$$\varepsilon(\sigma; v_1, \dots, v_n)\sigma \circ [1]^{\otimes n} = [1]^{\otimes n} \circ \chi(\sigma; v_1, \dots, v_n)\sigma\,. \tag{23.86}$$

Since both Sym and Alt can be defined in terms of the projectors

$$\pi_{\mathsf{Sym}} := \sum_{\sigma \in S_n} \varepsilon(\sigma)\sigma \qquad \text{and} \qquad \pi_{\mathsf{Alt}} := \sum_{\sigma \in S_n} \chi(\sigma)\sigma\,, \tag{23.87}$$

décalage interchanges symmetric and antisymmetric tensors. The most common choice is the following one:

$$[1] : V^{\otimes n} \to V[1]^{\otimes n} : v_1 \otimes \cdots \otimes v_n \mapsto (-1)^{\sum_{i=1}^n (n-i)\deg(v_i)} v_1[1] \otimes \cdots \otimes v_n[1]\,. \tag{23.88}$$

This choice is induced by the following definition of the suspension functor (one could also choose the convention where V is tensored on the right):

$$[1] : \mathbb{Z}\mathbf{Vect}_k \to \mathbb{Z}\mathbf{Vect}_k : V \mapsto k[1] \otimes V\,. \tag{23.89}$$

This definition also directly induces an algebra isomorphism in the following way. Consider two homogeneous elements $v, w \in V$. In $\mathsf{Sym}^2 V$, their product satisfies

$$vw = (-1)^{\deg(v)\deg(w)} wv\,. \tag{23.90}$$

After applying the suspension functor, the product on the left-hand side becomes:

$$v[1]w[1] = (\underline{1} \otimes v)(\underline{1} \otimes w) \cong (-1)^{\deg(w)} \underline{\underline{1}} \otimes (vw) = (-1)^{\deg(w)} vw[2]\,. \tag{23.91}$$

To calculate the suspension of the right-hand side, the braiding in $\mathbb{Z}\mathbf{Vect}_k$ is used:

$$\begin{aligned} v[1]w[1] = (v \otimes \underline{1})(w \otimes \underline{1}) \mapsto &(-1)^{\deg(v)+\deg(w)+\deg(v)\deg(w)+1}(w \otimes \underline{1})(v \otimes \underline{1}) \\ =&(-1)^{\deg(w)+\deg(v)\deg(w)+1}(wv) \otimes \underline{\underline{1}} \\ =&(-1)^{\deg(w)+\deg(v)\deg(w)+1} wv[2]\,. \end{aligned}$$

The difference in signs is $(-1)^{\deg(v)\deg(w)+1}$. If either v or w is even, this final sign is -1 or, equivalently, the product is antisymmetric, while if both v and w are odd, the product is symmetric. This is exactly the opposite situation of that in $\mathsf{Sym}^2 V$. The most thorough review of these issues was found in Miti (2021).

23.5.1 Supermatrices

For this section, the requirement that all algebraic structures are defined over a field k is relaxed to working over a supercommutative ring. This means that the objects will be (graded) modules instead of proper vector spaces.

Definition 23.5.10 (Supermatrix). Every linear transformation between super vector spaces (V_0, V_1) and (W_0, W_1) can be decomposed as the sum of 4 linear transformations between the even/odd subspaces:

- $A : V_0 \to W_0$,
- $B : V_1 \to W_0$,
- $C : V_0 \to W_1$, and
- $D : V_1 \to W_1$.

If these components are represented as matrices, the full transformation can be represented as a block matrix

$$X = \begin{pmatrix} A & B \\ C & D \end{pmatrix}.$$

These matrices can be classified according to their **parity**. Not all supermatrices preserve the grading or, equivalently, not all linear transformations of super vector spaces are morphisms of super vector spaces. The ones that are, are said to have even parity and they are of the form

$$X = \begin{pmatrix} \text{even} & \text{odd} \\ \text{odd} & \text{even} \end{pmatrix},$$

where even/odd means that the entries in these blocks have even/odd parity as elements of the underlying (graded) ring. It should be clear that these matrices indeed preserve the grading, since acting with an odd scalar on an odd vector gives an even vector (and similarly for the other combinations). The matrices that do not preserve the grading are said to have odd parity and are of the form

$$X = \begin{pmatrix} \text{odd} & \text{even} \\ \text{even} & \text{odd} \end{pmatrix}.$$

Definition 23.5.11 (Supertrace). The supertrace of a supermatrix generalizes the trace of an ordinary matrix. Given the block matrix form from the previous definition, the supertrace is defined as follows:

$$\mathrm{str}(X) := \mathrm{tr}(A) - \mathrm{tr}(D)\,. \tag{23.92}$$

Property 23.5.12. As was the case for the ordinary trace, the supertrace is invariant under basis transformations. Furthermore, the cyclicity property also still holds after

a slight modification to make it compatible with the grading:

$$\mathsf{str}(XY) = (-1)^{\deg(X)\deg(Y)}\mathsf{str}(YX)\,. \tag{23.93}$$

Definition 23.5.13 (Berezinian). The Berezinian or **superdeterminant** generalizes the determinant of an ordinary matrix. It is (uniquely) defined through the following two conditions:

1. $\mathsf{Ber}(XY) = \mathsf{Ber}(X)\mathsf{Ber}(Y)$, and
2. $\mathsf{Ber}(e^X) = e^{\mathsf{str}(X)}$.

An explicit formula is given by

$$\mathsf{Ber}(X) = \det(A - BD^{-1}C)\det(D)^{-1} = \det(A)\det(D - CA^{-1}B)^{-1}\,, \tag{23.94}$$

where the last expression involves the *Schur complement* of A relative to X. It should be noted that the Berezinian is only well-defined for invertible even matrices.

23.5.2 Berezin calculus

This section is an application of the concept of exterior algebras (Definition 23.4.21). Grassmann numbers/variables are used in quantum field theory when performing calculations in e.g. the fermionic sector or *Faddeev–Popov quantization*.

Definition 23.5.14 (Grassmann numbers). Let V be a vector space spanned by a set of elements θ_i. The Grassmann algebra with Grassmann variables θ_i is the exterior algebra over V. In this setting the wedge symbol of Grassmann variables is often omitted when writing the product:

$$\theta_i \wedge \theta_j \equiv \theta_i\theta_j\,.$$

Remark 23.5.15. From the (anti)commutativity, it follows that one can regard the Grassmann variables as being nonzero square roots of zero.

Notation 23.5.16 (Parity). In the case of superalgebras and, in particular, that of Grassmann numbers, the degree of an element is often called the (Grassmann) parity of the element. It is also often denoted by $\varepsilon(x)$ or ε_x instead of $\deg(x)$. In this text this convention is only adopted for graded algebras where there is both a supergrading and a (co)homological $\mathbb{Z}$-grading.

Property 23.5.17 (Polynomials). Consider a one-dimensional Grassmann algebra (with generator θ). When constructing the polynomial ring $\mathbb{C}[\theta]$ generated by θ, it can be seen that, due to the anticommutativity, $\mathbb{C}[\theta]$ is spanned only by 1 and θ. All higher-degree terms vanish because $\theta^2 = 0$. This implies that the most general polynomial over a one-dimensional Grassmann algebra is of the form

$$p(\theta) = a + b\theta\,. \tag{23.95}$$

Definition 23.5.18 (DeWitt convention). One can equip the exterior algebra Λ with Grassmann variables θ_i with an involution:

$$(\theta_i\theta_j \dots \theta_k)^* := \theta_k \dots \theta_j\theta_i \,. \tag{23.96}$$

Elements $z \in \Lambda$ that satisfy $z^* = z$ are said to be **(super)real** and elements that satisfy $z^* = -z$ are said to be **(super)imaginary**. This convention is called the DeWitt convention.

To keep the discussion about Grassmann variables self-contained, the calculus of Grassmann variables is introduced here.

Definition 23.5.19 (Derivative of Grassmann variables). Consider the polynomial algebra $\mathbb{C}[\theta_1, \dots, \theta_n]$ on n Grassmann variables (more general functions would be defined through a series expansion, but given that $\theta^2 = 0$, these always reduce to a simple polynomial). Differentiation on this ring is defined through the following relations:

$$\frac{\partial}{\partial\theta_j}\theta_i = \delta_i^j \qquad\qquad \theta_i\frac{\partial}{\partial\theta_j} + \frac{\partial}{\partial\theta_j}\theta_i = 0 \,. \tag{23.97}$$

The second relation implies that the partial derivatives are also Grassmann-odd. The odd parity in fact allows to introduce two distinct differentiation operations. One is the left derivative, this is the one that was just introduced. The other is the right derivative that acts as

$$\theta_i\frac{\partial^R}{\partial\theta^j} = \delta_i^j \,. \tag{23.98}$$

The left and right derivatives are also sometimes denoted by

$$\frac{\vec{\partial}}{\partial\theta^i} \qquad \text{and} \qquad \frac{\overleftarrow{\partial}}{\partial\theta^i} \,,$$

respectively.

Next, one also needs some kind of integration theory. Instead of working with a definition à la Riemann, the integral will be defined purely axiomatically.

Definition 23.5.20 (Berezin integral: axiomatic). Consider a function f of n Grassmann variables $\{\theta_i\}_{i\leq n}$. The Berezin integral $\int_B$ is defined by the following conditions:

1. The map $f \mapsto \int_B f(\theta)\,d\theta$ is linear.
2. The result $\int_B f(\theta)\,d\theta$ is independent of the variable(s) θ, i.e. it is a number.
3. The result is invariant under a translation of the integration variable.

Remark 23.5.21. Multiple integrals can be defined by adding the Fubini theorem as an additional axiom.

It can be shown that this definition is equivalent to the following one.

Alternative Definition 23.5.22 (Berezin integral: analytic). First consider functions in one Grassmann variable, i.e. $f(\theta) = a + b\theta$. The Berezin integral is then defined as follows:[6]

$$\int_B (a + b\theta)\, d\theta := b\,. \tag{23.99}$$

This means that the integral is equal to the coefficient of the highest-degree term. As a simple generalization, define

$$\int_B f(\theta_1, \dots, \theta_n)\, d\theta := \text{coefficient of } \theta_1 \cdots \theta_n\,. \tag{23.100}$$

Some authors reverse the order of the variables in the above definition. Depending on the number of variables, this might introduce an additional minus sign.

Remark 23.5.23. It is interesting to see that the (one-dimensional) Berezin integral is equal to the (Grassmann) derivative. This is completely different from the usual integral in calculus. It also gives some intuition for the distinct transformation behaviour of the Berezin integral as explained in the following property.

Formula 23.5.24 (Change of variables). Consider a general Berezin integral $\int_B f(\theta)\, d\theta$. Now, suppose that a transformation $\theta \to \xi(\theta)$ is applied to the Grassmann variables. If J is the Jacobian matrix associated to this transformation, the integral transforms according to the following formula:

$$\int_B f(\xi)\, d\xi = \int_B f(\theta)(\det J)^{-1}\, d\theta\,. \tag{23.101}$$

Berezin calculus can easily be unified with ordinary calculus by using the fact that ordinary coordinates (even parity) commute with Grassmann numbers (odd parity). A mixed derivative (resp. integral) can always be factorized as the composition of a Berezin derivative (resp. integral) and an ordinary one. The transformation behaviour is then generalized to this case by using the Berezinian.

[6] Technically the axioms only imply this formula up to some multiplicative constant. The original convention by *Berezin* will be adopted, i.e. this constant is chosen to be 1.

Chapter 24

Representation Theory

References for this chapter are Fulton and Harris (2004); Jeevanjee (2015). Section 2.2 and Section 2.3 can be visited for an introduction to groups and group actions.

Chapter Contents

24.1 Group representations **481**
24.2 Irreducible representations **482**
24.3 Classification **484**
24.4 Generalized representations **486**
24.4.1 Projective representations 486
24.4.2 Group cohomology 487
24.5 Tensor operators **487**

24.1 Group representations

Group actions on vector spaces are so important that they receive their own name.

Definition 24.1.1 (Representation). A representation of a group G on a vector space V is a group morphism $\rho : G \to \mathsf{GL}(V)$ from G to the automorphism group (Definition 22.2.8) of V:

$$\rho(g) \circ \rho(g') = \rho(gg') \tag{24.1}$$

for all $g, g' \in G$.

Property 24.1.2 (Freeness). Because every linear map takes the zero vector to itself, a representation can never be free.

Definition 24.1.3 (Subrepresentation). A subrepresentation of a representation on V is a subspace of V invariant under the action of the group G (together with the restricted action).

Example 24.1.4 (Alternating representation). The alternating or **sign** representation of S_n on an $\mathbb{R}$-vector space is defined by multiplication by the parity of the permutation.

Example 24.1.5 (Permutation representation). Consider a vector space V with basis $\{e_i\}_{i\leq n}$. Since the symmetric group S_n acts on the index set $\{1,\dots,n\}$, this induces a representation given by

$$\rho(g) : \sum_{i=1}^{n} v_i e_i \mapsto \sum_{i=1}^{n} v_i e_{g\cdot i} \,. \tag{24.2}$$

Example 24.1.6 (Contragredient representation). For every representation ρ on V, there exists a natural representation on the dual space V^*:

$$\rho^*(g) := \rho^T(g^{-1}) : V^* \to V^* \,, \tag{24.3}$$

where ρ^T is the transpose (Definition 22.2.21). It is implicitly defined by requiring

$$\left\langle \rho^*(g)(v^*), \rho(g)(v) \right\rangle = \langle v^*, v \rangle \tag{24.4}$$

for all $v \in V$ and $v^* \in V^*$, where $\langle \cdot, \cdot \rangle$ is the natural pairing.

Example 24.1.7 (Tensor product representation). A group G that acts on vector spaces V, W also has a representation on the tensor product $V \otimes W$ in the following way:

$$\rho_{V\otimes W}(g)(v \otimes w) := \rho_V(g)(v) \otimes \rho_W(g)(w) \,. \tag{24.5}$$

More generally, consider two representations $\phi : G \to \mathsf{GL}(V)$ and $\psi : H \to \mathsf{GL}(W)$. A representation of the direct product $G \times H$ on $V \otimes W$ is given by the tensor product of representations:

$$\rho(g,h)(v \otimes w) := \phi(g)(v) \otimes \psi(h)(w) \,. \tag{24.6}$$

The former case can be obtained as a subrepresentation induced by the diagonal subgroup inclusion $\Delta_G : G \hookrightarrow G \times G$.

Definition 24.1.8 (Intertwiner). If one interprets G-representations as G-modules, the natural morphisms are the intertwiners (Definition 2.3.6).

24.2 Irreducible representations

Definition 24.2.1 (Irreducibility). A representation is said to be irreducible if there exist no proper nonzero subrepresentations. Irreducible representations are also abbreviated as **irreps** (especially in physics).

Example 24.2.2 (Standard representation). Consider the action of S_n on a vector space V with basis $\{e_i\}_{i\leq n}$. The line generated by $e_1+e_2+\cdots+e_n$ is invariant under the permutation action of S_n. It follows that the permutation representation (on finite-dimensional spaces) is never irreducible.

The $(n-1)$-dimensional complementary subspace

$$W = \left\{ \sum_{i=1}^{n} \lambda_i e_i \,\middle|\, \sum_{i=1}^{n} \lambda_i = 0 \right\} \tag{24.7}$$

forms an irreducible representation. It is called the standard representation of S_n on V.

Schur's lemma 4.3.30 and its corollary are usually found in the following form.

Theorem 24.2.3 (Schur's lemma). *Let V, W be two finite-dimensional irreducible representations of a group G and let $\varphi : V \to W$ be an intertwiner.*

- *Either φ is an isomorphism or $\varphi = 0$.*
- *If $V = W$, then φ is constant, i.e. φ is a scalar multiple of the identity map $\mathbb{1}_V$.*

Property 24.2.4 (Complementary representation). Let V be a representation of a finite group G. For every subrepresentation W, there exists an invariant complementary subspace W'. This space can be found as follows. Choose an arbitrary complement U such that $V = W \oplus U$ with the associated projection map $\pi_0 : V \to W$. Averaging over G gives the intertwiner

$$\pi(v) := \sum_{g\in G} g \circ \pi_0(g^{-1}v)\,. \tag{24.8}$$

On W it is given by multiplication by $|G|$. Its kernel is an invariant subspace of V, complementary to W.

More generally, let V be an inner product space with a representation of a compact group G. A G-invariant inproduct on V can be obtained through the Haar measure (Definition 19.4.1):

$$\langle v \mid w\rangle_G := \int_G \langle gv \mid gw\rangle \, d\mu(g)\,. \tag{24.9}$$

The complementary subrepresentation of W is then simply the orthogonal complement $W^\perp$ with respect to this invariant inner product.

Theorem 24.2.5 (Maschke). *Let G be a finite group with a representation space V such that the characteristic of the underlying field does not divide the order of G. This representation can be uniquely (up to isomorphism) decomposed as*

$$V = V_1^{\oplus a_1} \oplus \cdots \oplus V_k^{\oplus a_k}\,, \tag{24.10}$$

where all V_k's are distinct irreducible representations.

Definition 24.2.6 (Matrix coefficient). Consider a compact topological group G. A matrix coefficient is a character $\chi : G \to \mathbb{C}^\times$ (Definition 19.4.4) of the form

$$\chi \equiv T \circ \rho\,, \tag{24.11}$$

where $\rho : G \to \mathsf{GL}(V)$ is a finite-dimensional representation of G and $T : \mathsf{End}(V) \to \mathbb{C}$ is a linear functional.[1]

Theorem 24.2.7 (Peter–Weyl).

I) The space of matrix coefficients is dense in $C(G)$ under the L^∞-norm and, accordingly, also in $L^2(G)$.[2]

II) Let $\rho : G \to \mathsf{U}(\mathcal{H})$ be a unitary representation on a (complex) Hilbert space $\mathcal{H}$ (see Definition 25.2.4). There exists an orthogonal decomposition of $\mathcal{H}$ in finite-dimensional, unitary irreps of G.

III) $L^2(G)$ can be decomposed as

$$L^2(G) = \widehat{\bigoplus_{[V_\rho]}} V_\rho^{\oplus \dim(V_\rho)}\,, \tag{24.12}$$

where $[V_\rho]$ runs over (the underlying vector spaces of) all isomorphism classes of finite-dimensional unitary irreps of G. In particular, an orthonormal basis is given by

$$\left\{ \sqrt{\dim(V_\rho)}\, u^\rho_{ij} \,\middle|\, [V_\rho], 1 \leq i,j \leq \dim(V_\rho) \right\}, \tag{24.13}$$

where the u^ρ_{ij} are matrix coefficients of the representation ρ with respect to a chosen orthonormal basis of V_ρ.

Corollary 24.2.8 (Class functions). When restricted to the set of Ad_G-invariant functions, the **class functions** play a special role, i.e. functions f such that $f(g^{-1}hg) = f(h)$ for all $g, h \in G$. A basis for the subspace of square-integrable class functions is given by the matrix coefficients for which the linear functional is given by the trace:

$$\mathsf{tr}_\rho = \sum_{i=1}^{\dim(V_\rho)} u^\rho_{ii}\,. \tag{24.14}$$

24.3 Classification

For an introduction to Young tableaux, see Section 1.9.1.

Definition 24.3.1 (Permutation module). Let λ be a partition. The permutation module M^λ is defined as the vector space generated by the Young tabloids of shape λ.

[1] Note that G need not be a matrix group, i.e. none of the representations have to be faithful.
[2] Square-integrability is well defined due to Haar's theorem 19.4.3.

Definition 24.3.2 (Specht module). Consider a permutation module M^λ for some λ with $|\lambda| = n$. Since the permutation group S_n acts on Young tableaux by permuting the entries, it also has an induced action[3] on M^λ. For every Young tableau t, it has a subrepresentation induced by the subgroup Q_t of permutations that leave the columns invariant. The Specht module is the representation generated by the following elements (called **polytabloids**):

$$e_t := \sum_{\sigma \in Q_t} \mathrm{sgn}(\sigma)\sigma \cdot \{t\}\,, \tag{24.15}$$

where t ranges over the Young tableaux of shape λ and $\{t\}$ denotes the Young tabloid associated to the tableau t. In fact, one can just take the standard Young tableaux as generators. This is sometimes called the **Specht basis**.

Property 24.3.3. A representation (over $\mathbb{C}$) of S_n is irreducible if and only if it is a Specht module S^λ for some partition λ of n.

One can restate the definition of a Specht module using the following operator.

Definition 24.3.4 (Young symmetrizer). Given a Young tableau t of shape λ, one can decompose the permutation group $S_{|\lambda|}$ as the union of two types of permutations. First, one has the permutations that preserve the rows, denote these by P_λ. Then, one has also the permutations that preserve the columns, denote these by Q_λ. These two subgroups induce elements in the group algebra $\mathbb{C}[S_{|\lambda|}]$ as follows:

$$a_\lambda := \sum_{p \in P_\lambda} p\,, \tag{24.16}$$

$$b_\lambda := \sum_{q \in Q_\lambda} \mathrm{sgn}(q) q\,. \tag{24.17}$$

The product $c_\lambda := b_\lambda a_\lambda$ is called the Young symmetrizer of λ.

Comparing the above definition of the Specht module to the Young symmetrizers leads to the following alternative definition.

Alternative Definition 24.3.5 (Specht module). The space $\mathbb{C}[S_{|\lambda|}]c_\lambda$ is called the Specht module S_λ.

Consider a vector space V together with its general linear group $\mathsf{GL}(V)$. For all $n \in \mathbb{N}$, there is an induced (diagonal) representation on $V^{\otimes n}$ by $\mathsf{GL}(V)$. There is also an action by the permutation group S_n that permutes the elements in a monomial $v_1 \otimes \cdots \otimes v_n \in V^{\otimes n}$.

Theorem 24.3.6 (Schur–Weyl). *This representation of* $\mathsf{GL}(V) \times S_n$ *can be decomposed as*

[3] This can be generalized to an action of the group algebra $\mathfrak{K}[S_n]$.

follows:

$$V^{\otimes n} \cong \bigoplus_{|\lambda|=n} V_\lambda \otimes S_\lambda V\,, \tag{24.18}$$

where

- *the sum ranges over all partitions of n or, equivalently, all Young diagrams with n boxes,*
- *the V_λ are Specht modules and, hence, irreducible representations of S_n, and*
- *the $S_\lambda V$ are (possibly zero) irreducible representations of $\mathsf{GL}(V)$ of the form $S_\lambda V \equiv \mathsf{Hom}_{S_n}(V_\lambda, V^{\otimes n})$.*

The spaces V_λ can be interpreted as multiplicity spaces. The spaces $S_\lambda V$ can be rewritten more explicitly as $V_\lambda \otimes_{S_n} V^{\otimes n}$ by self-duality of S_n-representations or, by using the explicit characterization of Specht modules given above, as $V^{\otimes n} c_\lambda$. Because of the functoriality of the involved operations, one can also see that S_λ is in fact a functor, the **Schur functor**.

Example 24.3.7 (Algebraic curvature tensor). The above definition of the Schur spaces $S_\lambda V$ allows for an a concise description of representations in terms of Young diagrams (and tableaux). Consider for example the *Riemann curvature tensor* R_{ijkl} from Chapter 36. This tensor has the following symmetries:

- $R_{ijkl} = -R_{jikl} = -R_{ijlk}$,
- $R_{ijkl} = R_{klij}$, and
- $R_{ijkl} + R_{iljk} + R_{iklj} = 0$.

By looking at the definition of the Young symmetrizer, one can see that these symmetries are exactly those of the irreducible component in $S_\lambda V$ associated to the partition $(2, 2)$.

@@ COMPLETE @@

24.4 Generalized representations

24.4.1 Projective representations

Definition 24.1.1 represented group elements by operators in such a way that the composition/multiplication is preserved. However, in some applications such as quantum mechanics (see Chapter 57), operations are only determined up to a phase factor and, hence, it makes sense that this composition rule is also relaxed. This is formalized as follows.

Definition 24.4.1 (Projective representation). A projective representation of a group G on a vector field V is a group morphism $\rho : G \to \mathsf{PGL}(V)$ from G to the **projective general linear group** of V:

$$\mathsf{PGL}(V) := \mathsf{GL}(V)/\mathfrak{K}^\times , \tag{24.19}$$

where $\mathfrak{K}$ is the underlying field of V. This means that

$$\rho(g) \circ \rho(g') = \omega(g, g')\rho(gg') \tag{24.20}$$

for all $g, g' \in G$ for some function $\omega : G \times G \to \mathfrak{K}^\times$.

24.4.2 Group cohomology

Example 24.4.2 (Bilinear forms). Consider a vector space V and a bilinear form $\omega : V \times V \to \mathbb{R}$ ($\mathbb{R}$ can be replaced by any other field). Then ω induces a class in $H^2(V; \mathbb{R})$ by linearity.

@@ COMPLETE (classification by cocycles, covering groups, generalized representations) @@

24.5 Tensor operators

Definition 24.5.1 (Representation operators). An intertwiner $\psi : (\rho, V_0) \to \mathsf{End}(V)$ between a G-representation on an auxiliary vector space V_0 to the space of linear maps on a G-vector space V (equipped with the adjoint action).

More explicitly, consider a set of operators $\{\widehat{O}_i\}_{i\in I} \subset \mathsf{End}(V)$ acting on a vector space V equipped with a representation ρ of the group G. This collection defines a representation operator with respect to G if there exists a matrix representation R of G such that the following equation holds:

$$\rho(g)\widehat{O}_i\rho(g)^{-1} = \sum_{j\in I} R(g)_{ij}\widehat{O}_j . \tag{24.21}$$

Example 24.5.2 (Tensor operators). Consider $G = \mathsf{SO}(3)$ and $V_0 = \mathcal{T}^r_s(\mathbb{R}^3)$. This choice gives a set of operators that transform as tensors under rotations. By choosing $V_0 = \mathbb{R}^3$ or $V_0 = \mathcal{H}_l(\mathbb{R}^3)$, the space of spherical harmonics of degree l, one obtains the **vector** and **spherical operators**.

The following property is often used in quantum mechanics to quickly find forbidden transitions in atomic or molecular systems.

Property 24.5.3 (Selection rules). Let G be a semisimple group and let W_1, W_2 be two inequivalent (finite-dimensional), irreducible, unitary subrepresentations of a Hilbert

space $\mathcal{H}$. Let $\widehat{O}$ be a representation operator indexed by a vector space V. For all $v \in V, w_i \in W_i$ one has

$$\langle w_1 \mid \widehat{O}(v) w_2 \rangle = 0\,, \tag{24.22}$$

unless $V \otimes W_2$ contains a subrepresentation equivalent to W_1.

Theorem 24.5.4 (Wigner–Eckart). *Consider two irreducible* $\mathsf{SU}(2)$*-subrepresentations* W_j *and* $W_{j'}$ *of some unitary representation* $\mathcal{H}$*, together with two degree-q spherical tensors* $\widehat{O}, \widehat{O}' : V_0 \to \mathsf{End}(\mathcal{H})$*. If there exists at least one index* $k \leq q$ *and one pair of vectors* $(v, v') \in W_j \times W_{j'}$ *such that*

$$\langle v' \mid \widehat{O}_k v \rangle \neq 0\,,$$

then for all indices $k \geq 0$ *and pairs* $(v, v') \in W_j, W_{j'}$ *the following equality holds*

$$\langle v' \mid \widehat{O}'_k v \rangle = C \langle v' \mid \widehat{O}_k v \rangle \tag{24.23}$$

for some constant C *that only depends on* q, j *and* j'*.*

By noting that the *Clebsch–Gordan coefficients* (see Formula 57.6.11) are the components of the projection $W_q \otimes W_j \to W_{j'}$, which is itself an intertwiner, one can recast the Wigner–Eckart theorem as a statement about matrix elements.

Corollary 24.5.5. Consider an irreducible tensor operator T_j^m (with respect to the rotation group). The matrix elements of this operator with respect to a symmetry-adapted basis ('angular momentum' basis) decompose as a product of a Clebsch–Gordan coefficient and a factor that only depends on the eigenvalues of the Casimir operator:

$$\langle j', m' \mid R^{(q)} \mid j, m \rangle = \langle j' \| R_k^{(q)} \| j \rangle \langle j', m' \mid q, j; k, m \rangle\,. \tag{24.24}$$

The factor $\langle j' \| R^{(q)} \| j \rangle$ is sometimes called the **reduced matrix element**.

Chapter 25

Functional Analysis

The main references for this chapter are Choquet-Bruhat and DeWitt-Morette (2000); Choquet-Bruhat et al. (1991). For a revision of topological spaces and inner product spaces, see Chapter 6 and Section 22.3, respectively.

In this chapter, the term 'linear operator', which was previously reserved for vector space automorphisms, is now used synonymously with 'linear map'. This is to keep the terminology in line with the standard literature on Banach spaces and operator spaces. Moreover, in this chapter, 'dual space' will indicate the topological/continuous dual and not just the algebraic/linear dual (unless stated otherwise).

Chapter Contents

25.1 Banach spaces **490**

25.1.1 Convex duality 491

25.2 Hilbert spaces **492**

25.2.1 General 492

25.2.2 Generalized Fourier series 494

25.2.3 Orthogonality and projections 495

25.2.4 Separable Hilbert spaces 495

25.3 Seminorms **496**

25.3.1 Topology 496

25.3.2 Locally convex spaces 497

25.3.3 Tensor products 500

25.3.4 Measure theory 502

25.4 Riesz spaces **502**

25.1 Banach spaces

Definition 25.1.1 (Topological vector space). A vector space for which addition and scalar multiplication are continuous. This is abbreviated as TVS.

Definition 25.1.2 (Weak topology). The initial topology (Definition 6.2.3) on a TVS with respect to its dual, i.e. a net $(v_\alpha)_{\alpha\in I}$ in V **converges weakly** to v if and only if $\lambda(v_\alpha) \longrightarrow \lambda(v)$ for all $\lambda \in V^*$.

Definition 25.1.3 (Weak-* topology). Every TVS admits a canonical embedding into its double dual:

$$\iota : V \to V^{**} : v \mapsto \mathsf{ev}_v\,, \tag{25.1}$$

where the evaluation map ev_v is defined as

$$\mathsf{ev}_v : V^* \to K : \lambda \mapsto \lambda(v)\,. \tag{25.2}$$

The weak-* topology on the dual space V^* is defined as the weak topology with respect to the image $\iota(V) \subseteq V^{**}$. Equivalently, it is the topology defined by pointwise convergence of nets.

Definition 25.1.4 (Norm). Let V be a TVS over a field K. A function $\|\cdot\| : V \to [0, +\infty[$ is called a norm if it satisfies following conditions for all $v, w \in V$:

1. **Nondegeneracy**: $\|v\| = 0 \iff v = 0$,
2. **Homogeneity**: for all scalars $\lambda \in K : \|\lambda v\| = |\lambda|\,\|v\|$, and
3. **Triangle equality (subadditivity)**: $\|v + w\| \leq \|v\| + \|w\|$.

Definition 25.1.5 (Norm topology). A norm $\|\cdot\|$ induces a metric (Definition 9.1.1):

$$d(v, w) := \|v - w\|\,. \tag{25.3}$$

The metric topology induced in this way is called the **norm topology** or **strong topology**.

Definition 9.1.4 can be extended as follows.

Definition 25.1.6 (Isometry). An isometry $f : V \to W$ between normed spaces is a linear map satisfying

$$\|f(v)\|_W = \|v\|_V \tag{25.4}$$

for all $v \in V$. If this only holds on a subspace $V' \leq V$, the map is said to be a **partial isometry**.

Property 25.1.7 (Continuity). Every linear map $\varphi : V \to W$ of topological vector spaces with finite-dimensional domain is continuous. Moreover, if V is normed, every linear isomorphism $\varphi : V \to W$ is a homeomorphism.

Corollary 25.1.8. Since any two finite-dimensional normed vector spaces with the same dimension are homeomorphic, all metrics on a finite-dimensional normed vector space are equivalent.

Definition 25.1.9 (Banach space). A normed vector space that is complete 9.4.2 in the norm topology.

Property 25.1.10 (Duals). The dual of a Banach space is also a Banach space.

Definition 25.1.11 (Reflexive space). A TVS V for which the canonical inclusion $V \hookrightarrow V^{**}$ is an isomorphism.

Property 25.1.12. Every finite-dimensional Banach space is reflexive.

Property 25.1.13. On a reflexive space, the weak and weak-* topologies coincide.

Property 25.1.14 (Weak duals). Consider a TVS V. The dual of $(V^*, \tau_{\text{weak}^*})$ is isomorphic to V.

Theorem 25.1.15 (Open mapping theorem[1]). *Let $f : V \to W$ be a continuous linear operator between two Banach spaces. If f is surjective, it is also open.*

Theorem 25.1.16 (Banach–Alaoglu[2]). *The closed unit ball in the dual of a normed space is compact in the weak-* topology.*

Definition 25.1.17 (Bracket). Let X be a set and consider a normed subset $(\mathcal{F}, \|\cdot\|)$ of $\mathbb{R}^X$. For every two functions $u, l : X \to \mathbb{R}$, the bracket $[l, u]$ consists of all functions $f \in \mathbb{R}^X$ such that

$$\forall x \in X : l(x) \leq f(x) \leq u(x) . \tag{25.5}$$

An ε-bracket, with $\varepsilon > 0$, is a bracket $[l, u]$ such that $\|u - l\| < \varepsilon$. The **bracketing number** $N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|)$ is defined as the least number of ε-brackets needed to cover $\mathcal{F}$.

25.1.1 Convex duality

The Legendre transformation (Definition 16.7.7) can be generalized to arbitrary vector spaces.

Definition 25.1.18 (Legendre–Fenchel transformation). Let V be a topological vector space and consider its (topological) dual V^*. The Legendre–Fenchel transform or **convex conjugate** $f^* : V^* \to \overline{\mathbb{R}}$ of a function $f : V \to \overline{\mathbb{R}}$ is defined as follows:

$$f^*(\varphi) := \sup_{v \in V} \big(\langle \varphi, v \rangle - f(v) \big) . \tag{25.6}$$

Property 25.1.19 (Fenchel–Young inequality). Consider a function $f : V \to \overline{\mathbb{R}}$ and its

[1] Sometimes called the **Banach–Schauder** theorem.
[2] Apparently, at least 12 different mathematicians should be named in this theorem.

convex conjugate f^*. For all $v \in V$ and $\varphi \in V^*$ the following inequality holds:

$$\langle \varphi, v \rangle \leq f(v) + f^*(\varphi) \,. \tag{25.7}$$

Property 25.1.20. The convex conjugate is always lower semicontinuous (Definition 6.2.11). The double convex conjugate f^{**} is equal to the **closed convex hull**, i.e. the greatest lower semicontinuous convex function smaller than f.

Theorem 25.1.21 (Fenchel–Moreau). *Let V be a locally convex Hausdorff space (see Definition 25.3.8 further below). The equality $f = f^{**}$ holds if and only if one of the following conditions holds:*

- *f is identically $+\infty$,*
- *f is identically $-\infty$, or*
- *f is a lower semicontinuous, proper[3], convex function.*

25.2 Hilbert spaces

25.2.1 General

Remark 25.2.1. Let V be an inner product space (Definition 22.3.1). A norm on V can be induced by the inner product in the following way:

$$\|v\|^2 = \langle v \mid v \rangle \,. \tag{25.8}$$

However, the converse is not true. Not every norm induces an inner product. Only norms that satisfy the **parallelogram law**

$$\|v + w\|^2 + \|v - w\|^2 = 2(\|v\|^2 + \|w\|^2) \tag{25.9}$$

can be used to define an inner product. This inner product can be recovered through the **polarization identity**:

$$4\langle v \mid w \rangle = \|v + w\|^2 - \|v - w\|^2 + i\big(\|v + iw\|^2 - \|v - iw\|^2\big) \,. \tag{25.10}$$

Property 25.2.2 (Cauchy–Schwarz inequality).

$$|\langle v \mid w \rangle| \leq \|v\| \, \|w\| \,, \tag{25.11}$$

where the equality holds if and only if v and w are linearly dependent.

Corollary 25.2.3 (Triangle inequality). The Cauchy–Schwarz inequality can be used to prove the triangle inequality. Together with the properties of an inner product, this implies that an inner product space is indeed a normed space as mentioned in the beginning of this section.

[3] It never becomes $-\infty$ and is not identically $+\infty$.

Definition 25.2.4 (Hilbert space). A Banach space where the norm is induced by an inner product.

Example 25.2.5. Consider two square-integrable functions $f, g \in L^2([a,b], \mathbb{C})$. As mentioned in Section 18.3.2, the inner product of f and g is defined as follows:

$$\langle f \mid g \rangle = \int_a^b \overline{f(x)} g(x)\, dx\,. \tag{25.12}$$

It is possible to generalize this inner product with respect to a weight function $\phi \in L^2([a,b], \mathbb{C})$:

$$\langle f \mid g \rangle_\phi := \int_a^b \overline{f(x)} g(x) \phi(x)\, dx\,. \tag{25.13}$$

Formula 25.2.6 (Pythagoras). In an inner product space, the triangle equality reduces to the well-known Pythagorean theorem for orthogonal vectors:

$$\langle v \mid w \rangle = 0 \implies \|v + w\|^2 = \|v\|^2 + \|w\|^2\,. \tag{25.14}$$

This formula can be extended to any set of orthogonal vectors $\{v_1, \dots, v_n\}$ as follows:

$$\left\| \sum_{i=1}^n v_i \right\|^2 = \sum_{i=1}^n \|v_i\|^2\,. \tag{25.15}$$

Theorem 25.2.7 (Riesz's representation theorem). *Let $\mathcal{H}$ be a Hilbert space. For every continuous linear functional $\rho \in \mathcal{H}^*$, there exists a unique element $v_0 \in \mathcal{H}$ such that*

$$\rho(h) = \langle v_0 \,|\, h \rangle \tag{25.16}$$

for all $h \in \mathcal{H}$. This implies that $\mathcal{H}$ and $\mathcal{H}^*$ are isometrically isomorphic.[4] *Furthermore, the operator norm of ρ is equal to the norm of v_0.*

Remark 25.2.8. This theorem justifies the bra-ket notation used in quantum mechanics, where one associates to every *ket* $|\psi\rangle \in \mathcal{H}$, a *bra* $\langle\psi| \in \mathcal{H}^*$.

Remark 25.2.9 (Relation to Riesz–Markov theorem). By the Riesz–Markov theorem 19.1.6, every continuous functional on $C(\widehat{X})$, where X is locally compact, can be written as the integration againt some Radon measure. By using the fact that every Hilbert space is isomorphic to some function space $L^2(X, \mu)$, one can see that Riesz's representation theorem is very similar.

Recall Definition 25.1.6.

Property 25.2.10 (Isometries). If $f : V \to \mathcal{H}'$ is a partial isometry of Hilbert spaces and $V \subseteq \mathcal{H}$ is dense, then their exists a continuous isometric extension $F : \mathcal{H} \to \mathcal{H}'$.

[4] Anti-isomorphic in the convention with conjugate linearity in the second argument.

25.2.2 Generalized Fourier series

Property 25.2.11 (Bessel's inequality). Let $\mathcal{H}$ be a Hilbert space. The following equality holds for all orthonormal vectors $v_1, \dots, v_n \in \mathcal{H}$ and scalars $\lambda_1, \dots, \lambda_n \in \mathbb{C}$:

$$\left\| v - \sum_{i=1}^{n} \lambda_i v_i \right\|^2 = \|v\|^2 - \sum_{i=1}^{n} |\langle v \,|\, v_i \rangle|^2 + \sum_{i=1}^{n} |\langle v \,|\, v_i \rangle - \lambda_i|^2 \,. \tag{25.17}$$

This expression is minimized when the last term vanishes. This leads to Bessel's inequality

$$\sum_{i=1}^{n} |\langle v \,|\, v_i \rangle|^2 \leq \|v\|^2 \,, \tag{25.18}$$

together with the property that the optimal choice in the generalized Fourier series for v is obtained by taking the coefficients to be the projections $\lambda_i = \langle v \,|\, v_i \rangle$.

Corollary 25.2.12. The sum in Eq. (25.18) is bounded for all n, so the series $\sum_{i=1}^{+\infty} |\langle v \,|\, v_i \rangle|^2$ converges for all v. This implies that the sequence $(\langle v \,|\, v_n \rangle)_{n \in \mathbb{N}}$ belongs to l^2.

Theorem 25.2.13. *Consider a Hilbert space $\mathcal{H}$. Let $(v_n)_{n \in \mathbb{N}}$ be an orthonormal sequence in $\mathcal{H}$ and let $(\lambda_n)_{n \in \mathbb{N}}$ be a sequence in $\mathbb{C}$. The series $\sum_{i=1}^{+\infty} \lambda_i v_i$ converges in $\mathcal{H}$ if and only if $(\lambda_n)_{n \in \mathbb{N}} \in l^2$. Furthermore, the series satisfies the following equality:*

$$\left\| \sum_{i=1}^{+\infty} \lambda_i v_i \right\|^2 = \sum_{i=1}^{+\infty} |\lambda_i|^2 \,. \tag{25.19}$$

Bessel's inequality implies that the sequence $(\langle v \mid v_n \rangle)_{n \in \mathbb{N}}$ belongs to l^2, so the generalized Fourier series of $v \in \mathcal{H}$ converges in $\mathcal{H}$.

Remark 25.2.14. Although the convergence of the generalized Fourier series of $v \in \mathcal{H}$ can be established using the previous theorem, it does not follow that the expansion converges to v itself. One can merely say that the Fourier expansion is the best approximation of v with respect to the norm on $\mathcal{H}$.

Definition 25.2.15 (Complete set). Let $\{e_i\}_{i \in I}$ be a set of orthonormal vectors in an inner product space V. This set is said to be complete if every vector $v \in V$ can be expressed as follows:

$$v = \sum_{i \in I} \langle e_i \,|\, v \rangle e_i \,. \tag{25.20}$$

This is also sometimes called an **orthonormal basis** or **Hilbert basis**. Note that it is not necessarily a (Hamel) basis, since the linear combination does not have to be finite.

Alternative Definition 25.2.16. A complete set of orthonormal vectors in a Hilbert space $\mathcal{H}$ is a set $S \subset \mathcal{H}$ such that one cannot add another nonzero vector $w \in \mathcal{H}$ to it satisfying

$$\forall v_i \in S : \langle v_i \,|\, w \rangle = 0 \,. \tag{25.21}$$

Theorem 25.2.17 (Parceval). *Let $(v_n)_{n\in\mathbb{N}}$ be a complete sequence in a Hilbert space $\mathcal{H}$. Every vector $v \in \mathcal{H}$ has a unique Fourier series representation $\sum_{i=1}^{+\infty} \lambda_i v_i$, where the Fourier coefficients $(\lambda_n)_{n\in\mathbb{N}}$ belong to l^2. Conversely, if Bessel's inequality becomes an equality for every $v \in \mathcal{H}$, the sequence $(v_n)_{n\in\mathbb{N}}$ is complete.*

25.2.3 Orthogonality and projections

Property 25.2.18. Let S be a subset (not necessarily a subspace) of a Hilbert space $\mathcal{H}$. The orthogonal complement $S^\perp$ is closed in $\mathcal{H}$.

Corollary 25.2.19. The previous property implies that the orthogonal complement of some arbitrary subset of a Hilbert space is a Hilbert space itself.

Theorem 25.2.20 (Projection theorem). *Let H be a Hilbert space and $S \leq H$ a complete subspace. For every $v \in H$, there exists a unique $v' \in S$ such that $v - v'$ is orthogonal to every $w \in S$, i.e. $v - v' \in S^\perp$.*

Remark 25.2.21. An equivalent definition for this projection is the vector v' satisfying

$$\|v - v'\| = \inf\{\|v - w\| \mid w \in S\}\,. \tag{25.22}$$

It is often denoted by $\mathrm{proj}_S(v)$.

Corollary 25.2.22 (Orthogonal decomposition). It follows that given a complete (or closed) subspace S, the Hilbert space $\mathcal{H}$ can be decomposed as $\mathcal{H} = S \oplus S^\perp$.

Definition 25.2.23 (Trace). Let $\mathcal{H}$ be a Hilbert space with orthonormal basis $\{e_i\}_{i\in I}$. Given a bounded linear operator $S \in \mathcal{B}(\mathcal{H})$, one defines its trace as follows:

$$\mathrm{tr}(S) := \sum_{i\in I} \langle e_i \mid Se_i \rangle\,. \tag{25.23}$$

25.2.4 Separable Hilbert spaces

The definition of separable spaces in the sense of point-set topology is given in Definition 6.5.26. An equivalent definition for Hilbert spaces is the following one (provided that one accepts Zorn's lemma 1.7.14).

Alternative Definition 25.2.24 (Separable Hilbert space). A Hilbert space that contains a Hilbert basis.

Example 25.2.25 (Finite dimensions). By the Gram–Schmidt method, it follows that every finite-dimensional Hilbert space is separable.

The following theorem shows that (up to isomorphism) there are only two distinct types of separable Hilbert spaces.

Theorem 25.2.26. *Let $\mathcal{H}$ be separable. If $\mathcal{H}$ is n-dimensional, it is isometrically isomorphic to $\mathbb{C}^n$. If $\mathcal{H}$ is infinite-dimensional, it is isometrically isomorphic to l^2.*

Property 25.2.27. Every orthogonal subset of a separable Hilbert space is countable.

25.3 Seminorms

Definition 25.3.1 (Seminorm). Let V be a vector space over a normed field K. A function $p : V \to [0, +\infty[$ is called a seminorm if it satisfies the following conditions:

1. **Homogeneity**: $p(\lambda v) = |\lambda| p(v)$ for all scalars $\lambda \in K$ and $v \in V$.
2. **Triangle equality (subadditivity)**: $p(v + w) \leq p(v) + p(w)$ for all $v, w \in V$.

Theorem 25.3.2 (Hahn–Banach). *Let V be a TVS equipped with a seminorm p. If f is a continuous linear functional on V such that $|f(w)| \leq p(w)$ on a subspace $W \leq V$, there exists a linear extension F of f to V such that*

$$|F(v)| \leq p(v) \tag{25.24}$$

for all $v \in V$.

25.3.1 Topology

In this subsection $\mathscr{P}$ denotes a family of seminorms defined on a TVS V with index family I.

Definition 25.3.3 ($\mathscr{P}$-open ball). A $\mathscr{P}$-open ball centred on v_0 is a subset $W \subset V$ such that all points $w \in W$ satisfy the following condition for a finite number of seminorms $p_i \in \mathscr{P}$:

$$p_i(w - v_0) \leq \varepsilon_i \,, \tag{25.25}$$

where $\varepsilon_i > 0$.

Property 25.3.4. The set of $\mathscr{P}$-open balls generates a topology on V. This topology is often called the **$\mathscr{P}$-topology**.

Definition 25.3.5 (Separated family). A family of seminorms $\mathscr{P}$ is said to be separated if for every nonzero point $v \in V$ there exists a seminorm $p \in \mathscr{P}$ such that $p(v) \neq 0$. If $\mathscr{P}$ is separated, then $\sum_{i \in I} p_i$ is a norm.

Property 25.3.6. A family of seminorms $\mathcal{P}$ is separated if and only if it generates a Hausdorff topology on V. Furthermore, the topology is metrizable if and only if $\mathcal{P}$ is countable. The **Fréchet metric**, which is translation-invariant, is given by

$$d(v, w) := \sum_{i \in I} \frac{1}{2^i} \frac{p_i(v - w)}{1 + p_i(v - w)} \,. \tag{25.26}$$

Although the Hahn–Banach theorem 25.3.2 does not imply that the linear extension is unique, one can refine the statement in the case of dense subspaces.

Corollary 25.3.7. Let V be a TVS with a $\mathscr{P}$-topology and let W be a dense subspace. If f is a linear form on W, continuous in the subspace topology, there exists a unique linear extension to V.

25.3.2 Locally convex spaces

Definition 25.3.8 (Locally convex space). Let V be a TVS.

- A **cone** is a subset $U \subseteq V$ such that the line segment connecting any vector to the origin lies in U.
- A subset $U \subseteq V$ is said to be **balanced** if for every vector $v \in U$ the scalar multiples λv, with $|\lambda| \leq 1$, also lie in U. Such a subset is sometimes also called a **circled cone**.
- An **absolutely convex** set is a balanced convex set. Equivalently, this is a subset closed under linear combinations where the absolute values of the coefficients sum at most to 1.
- A subset $U \subseteq V$ is said to be **absorbent** if the union of all sets λU, where λ ranges over the base field, equals the total space.

A locally convex space is a topological vector space where the origin admits a local base of convex sets. Sometimes these sets are also required to be balanced and absolutely convex. However, it can be shown that this is a mere property.

Using the notion of seminorms one can restate this definition as follows.

Alternative Definition 25.3.9 (Locally convex space). A topological vector space is locally convex if its topology is generated by a family of seminorms.

The following example of locally convex spaces is important in functional analysis.

Definition 25.3.10 (Fréchet space). A locally convex topological vector space that admits a complete translation-invariant metric.

By Property 25.3.6, there exists an equivalent formulation.

Alternative Definition 25.3.11 (Fréchet space). A topological vector space that admits a complete metric topology induced by a countable separated family of seminorms.

Locally convex topological vector spaces are important in functional analysis because they are one of the most general types of spaces that lend themselves to the definition of differentiation. A first step in this direction is the following generalization of the (directional) derivative.

Definition 25.3.12 (Gâteaux derivative). Consider two locally convex spaces V, W. The Gâteaux differential of a continuous map $f : U \to W$, where $U \subseteq V$ is open, is defined as follows:

$$\mathrm{d}f(v;h) := \lim_{t\to 0} \frac{f(v+th) - f(v)}{t} = \frac{\mathrm{d}}{\mathrm{d}t} f(v+th)\bigg|_{t=0} . \tag{25.27}$$

If this limit exists for all $h \in V$, the function is said to be **Gâteaux differentiable** at $v \in V$. Moreover, if it is also continuous in both arguments, it is said to be of class C^1. By iterating this construction one can define C^k- and even C^∞-maps:

$$\mathrm{d}^{(k)}f(v;h_1 \otimes \cdots \otimes h_k) := \frac{\mathrm{d}^k}{\mathrm{d}t^k} f(v + th_1 + \cdots + th_k)\bigg|_{t=0} . \tag{25.28}$$

Now, it should be noted that the map $\mathrm{d}f(v;-)$ is not necessarily linear. If it is linear, the function $\delta_v f : V \to W : h \mapsto \mathrm{d}f(v;h)$ is called the **Gâteaux derivative** of f at v. It can be shown that the Gâteaux differential of C^1-functions is always linear and, hence, defines a Gâteaux derivative. In fact, this notion of differentiability is often called **(Michal–)Bastiani differentiability**.

Formula 25.3.13 (Fundamental theorem of calculus). If $f \in C^1(V, W)$, then

$$f(v+h) = f(v) + \int_0^1 \mathrm{d}f(v+th;h)\, dt , \tag{25.29}$$

where the integral is understood in the following sense. Consider a function $f : (X, \Sigma, \mu) \to V$ from a measure space to a vector space V. If for all $\varphi \in V^*$ the pairing $\varphi \circ f$ is (Lebesgue) integrable and there exists for all $A \in \Sigma$ a vector $v_A \in V$ such that

$$\langle \varphi, e_A \rangle = \int_A \varphi \circ f \, d\mu , \tag{25.30}$$

then f is said to be **Gel'fand–Pettis integrable** and v_A is called the **Gel'fand–Pettis integral** of f over A.

On normed spaces, can also introduce alternative notions of differentiability.

Definition 25.3.14 (Hadamard derivative). Let $f : V \to W$ be a function of normed spaces. It is said to be **Hadamard differentiable** at $v \in V$ if there exists a bounded linear operator $\mathrm{D}f_v$ such that

$$\lim_{n\to\infty} \frac{f(v + t_n h_n) - f(v)}{t_n} = \mathrm{D}f_v(h) \tag{25.31}$$

for all sequences $h_n \longrightarrow h$ and $t_n \downarrow 0$.

Property 25.3.15. If f is Hadamard-differentiable, it is also Gâteaux-differentiable and the differentials coincide.

Definition 25.3.16 (Fréchet derivative). Let $f : V \to W$ be a function of normed spaces. It is said to be **Fréchet differentiable** at $v \in V$ if there exists a bounded linear operator $\mathsf{D}f_v$ such that

$$\lim_{\|h\|\to 0} \frac{\|f(v+h) - f(v) + \mathsf{D}f_v(h)\|}{\|h\|} = 0\,. \tag{25.32}$$

If the linear operator $\mathsf{D}f$ exists, it is called the Fréchet derivative of f at v. If f is (Fréchet) differentiable at any point in V and if the map $V \to \mathcal{B}(V, W) : v \mapsto \mathsf{D}f_v$ is continuous, then f is said to be of class C^1.

The relation between Gâteaux and Fréchet derivatives (and the Hadamard derivative since it is squeezed in between) is clarified by the following property.

Property 25.3.17. If a function between normed spaces has a continuous and linear Gâteaux differential (i.e. if it has a Gâteaux derivative), it is also Fréchet differentiable. Furthermore, the Gâteaux derivative and Fréchet derivative coincide.

Although one can extend functional analysis to Fréchet spaces (or even locally convex spaces), they are less well-behaved than Banach spaces.

Property 25.3.18. The dual of a Fréchet space V is Fréchet if and only if V is Banach (and, hence, V^* will also be Banach). Furthermore, the space $\mathcal{L}(V, W)$ of linear maps between Fréchet spaces is Fréchet if and only if W is Banach.

Theorem 25.3.19 (Krein–Milman). *Every compact, convex subset of a locally convex Hausdorff space is equal to the convex hull of its extreme points.*

This theorem can be extended as follows.

Theorem 25.3.20 (Choquet). *For every convex, compact subset C of a normed vector space there exists an assignment*

$$\mu : C \to \mathbb{P}\big(\mathsf{Extr}(C)\big) : c \mapsto \mu_c \tag{25.33}$$

such that

$$f(c) = \int_{\mathsf{Extr}(C)} f\, d\mu_c \tag{25.34}$$

for all affine functions $f : C \to \mathbb{C}$.

Remark 25.3.21. The extension of Choquet's theorem to locally convex topological vector spaces is called the **Choquet–Bishop–de Leeuw theorem**.

Definition 25.3.22 (Strong topology). Let V be a TVS. The strong topology (or **topology of uniform convergence on bounded sets**) on V^* is the locally convex topology defined by the seminorms

$$p_B(\lambda) := \sup_{v\in B} |\lambda(v)|\,, \tag{25.35}$$

where B is a bounded set of V. This is an example of a *polar topology*. A dual space equipped with the strong topology is often called the **strong dual**.

25.3.3 Tensor products

When moving from finite-dimensional vector spaces to general topological vector spaces, the algebraic tensor product from Section 23.3 does not behave in the way one would expect it to. For example, the (algebraic) tensor product of the smooth algebras $C^\infty(\mathbb{R}^m)$ and $C^\infty(\mathbb{R}^n)$ only injects into $C^\infty(\mathbb{R}^{m+n})$, i.e. not all bivariate smooth functions can be written as a finite sum of products of univariate smooth functions. In this section this will be resolved.

Definition 25.3.23 (Tensor product of Hilbert spaces). The algebraic tensor product of two Hilbert spaces V, W can be equipped with an inner product defined on outer products as

$$\langle v_1 \otimes w_1 \mid v_2 \otimes w_2 \rangle_{V \otimes W} := \langle v_1 \mid v_2 \rangle_V \langle w_1 \mid w_2 \rangle_W \tag{25.36}$$

and extended to all of $V \otimes W$ by linearity. The Hilbert space tensor product $V \widehat{\otimes} W$, often denoted by $V \otimes_\sigma W$, is then defined as the completion of $V \otimes W$ with respect to this inner product.

Definition 25.3.24 (Tensor product of Banach spaces). Contrary to the case of Hilbert spaces, the norms on two Banach spaces V and W do not induce a unique natural norm on $V \otimes W$. Two common choices are the following ones:

$$\|x\|_{\text{proj}} := \inf\left\{ \sum_{i=1}^n \|a_i\|\|b_i\| \,\middle|\, x = \sum_{i=1}^n a_i \otimes b_i \right\} \tag{25.37}$$

and

$$\|x\|_{\text{inj}} := \sup\left\{ |(\mu \otimes \nu)(x)| \mid \mu \in V^*, \nu \in W^* : \|\mu\| = \|\nu\| = 1 \right\}. \tag{25.38}$$

These two norms are called the **projective** and **injective** norms, respectively. Accordingly, the completions $V \widehat{\otimes}_\pi W$ and $V \widehat{\otimes}_\varepsilon W$ of the algebraic tensor product $V \otimes W$ with respect to these norms are called the **projective** and **injective** tensor products.[5] The topologies induced by these norms are called the **π-topology** and **ε-topology**, respectively.

Definition 25.3.25 (Tensor products of locally convex spaces). Let V, W be locally convex spaces. Definition 25.3.9 gives rise to a family of projective seminorms as in the definition above. These define the projective tensor products $V \otimes_\pi W$. As mentioned

[5]Note that, in general, the projective tensor product is not complete, even when both V and W are. If both V and W are infinite-dimensional Banach spaces, their tensor product (in this sense) will never be complete.

above, in general, the projective tensor product is not complete, even when both V and W are. The completion is denoted by $V \widehat{\otimes}_\pi W$.

The injective tensor product can be extended to locally convex spaces as follows. Recall Definition 9.5.7 in the setting of linear maps. A collection $\mathcal{F}$ of linear maps is said to be **equicontinuous** if for every neighbourhood U_2 of the origin, there exists a neighbourhood U_1 of the origin such that

$$\forall f \in \mathcal{F} : x \in U_1 \implies f(x) \in U_2 . \tag{25.39}$$

The preceding definition of injective cross norms can be generalized to a family of seminorms on[6] $B_{\text{cont}}(V'_\sigma, W'_\sigma)$ as follows:

$$\|\phi\|_{A,B} := \sup\left\{|\phi(x,y)| \,|\, x \in X, y \in Y\right\}, \tag{25.40}$$

where X, Y range over the equicontinuous subsets of V' and W', respectively. $V \otimes W$ with the induced topology, as a subset of $\mathcal{B}_{\text{cont}}(V, W)$, is called the injective tensor product $V \otimes_\varepsilon W$. An equivalent formulation works with a family of seminorms on the algebraic tensor product $V \otimes W$ given by

$$\|x\|_{p,q} := \sup\left\{|(\mu \otimes \nu)(x)| \,\middle|\, v \in B_p^\circ, w \in B_q^\circ\right\}, \tag{25.41}$$

where B_p° denotes the *polar* of the p-unit ball.[7] Alternatively, the ε-topology is the topology of uniform convergence on equicontinuous subsets $X \times Y$ of $V'_\sigma \times W'_\sigma$.

@@ CHECK IF TRUE (little complete information to be found) @@

Alternative Definition 25.3.26 (Projective tensor product). Let V, W be locally convex spaces. The projective tensor product $V \otimes_\pi W$ carries the finest locally convex topology with respect to the canonical injection $V \times W \to V \otimes W : (v, w) \mapsto v \otimes w$.

Universal Property 25.3.27. The π-topology is the unique topology on the algebraic tensor product $V \otimes W$ such that the linear isomorphism (Universal Property 23.3.3)

$$\mathcal{L}(V \otimes W, Z) \cong B(V, W; Z) \tag{25.42}$$

induces a linear isomorphism

$$\mathcal{L}_{\text{cont}}(V \otimes_\pi W, Z) \cong B_{\text{cont}}(V, W; Z) \tag{25.43}$$

for all locally convex spaces V, W and Z.

Corollary 25.3.28. $B(V, W)$ is the continuous dual of $V \otimes_\pi W$.

Property 25.3.29. The π-topology is finer than the ε-topology.

[6] Note that this set is (linearly) isomorphic to the algebraic tensor product $V \otimes W$.

[7] The relation between these two constructions is given by the fact that a set of linear functionals is equicontinuous if and only if it is contained in the polar of a neighbourhood of the origin.

Definition 25.3.30 (Nuclear space). A locally convex space such that the completed injective and projective tensor products with any other locally convex space are isomorphic.

Property 25.3.31. Nuclear normed spaces are finite dimensional. In particular, infinite-dimensional Banach spaces are never nuclear.

@@ ADD (Schwartz' kernel theorem, ...) @@

25.3.4 Measure theory

Definition 25.3.32 (Distribution). Let V be locally convex vector space. A distribution on V is an equivalence class of linear maps from V^* to the measurable functions on a probability space (Ω, Σ, P), where two maps are identified if for any finite tuple of vectors the joint distributions of their images are equal.

The following remark gives an analogue of the Riesz–Markov theorem 19.1.6.

Remark 25.3.33. For finite-dimensional spaces, the distributions are in a one-to-one correspondence with the regular Borel measures.

@@ COMPLETE @@

25.4 Riesz spaces

Definition 25.4.1 (Riesz space). An ordered vector space (necessarily over $\mathbb{R}$) with the structure of a lattice. This is also called a **vector lattice**.

Definition 25.4.2 (Absolute value). Let $v \in V$ be an element of a Riesz space. Its absolute value is defined as follows:

$$|v| := \sup\{v, -v\}\,. \tag{25.44}$$

Two elements $v, w \in V$ are said to be **disjoint** if

$$\sup\{|v|, |w|\} = 0\,. \tag{25.45}$$

Definition 25.4.3 (Ideal). Let V be a Riesz space. A subspace $W \leq V$ is said to be an ideal if, for all $f \in W$ and $g \in V$, the following condition holds:

$$|g| \leq |f| \implies g \in W\,. \tag{25.46}$$

An ideal generated by a single vector is said to be **principal** (cf. Section 2.4.2).

Definition 25.4.4 (Band). An ideal $W \leq V$ in a Riesz space that contains any element whose absolute value is the supremum of a collection of positive elements in W. A band $W \leq V$ is said to be a **projection band** if

$$V = W \oplus W^{\perp}, \tag{25.47}$$

where $W^{\perp}$ is the disjoint complement of W. If every (principal) band is also a projection band, the Riesz space is said to have the **(principal) projection property**.

Property 25.4.5 (Dedekind-completeness). Every Dedekind-complete Riesz space has the (principal) projection property.

Definition 25.4.6 (Component). Let V be a Riesz space and consider a positive element $e \in V$. A component of e is a positive element $p \in V$ such that

$$\inf\{p, e - p\} = 0. \tag{25.48}$$

A linear combination of pairwise disjoint components is called an ***e*-simple function**.

Definition 25.4.7 (Uniform convergence). Let $e \in V$ be a positive element of a Riesz space. A sequence $(v_n)_{n\in\mathbb{N}} \subset V$ is said to converge e-uniformly to $v \in V$ if, for every $\lambda > 0$, there exists an index $N_\lambda \in \mathbb{N}$ such that

$$n > N_\lambda \implies |x_n - x| < \lambda e. \tag{25.49}$$

Theorem 25.4.8 (Freudenthal spectral theorem). *Let V be a Riesz space with the principal projection property and consider a positive element $e \in V$. For every $f \in I_e$ in the ideal generated by e, there exist sequences $(s_n)_{n\in\mathbb{N}}, (t_n)_{n\in\mathbb{N}}$ of e-simple functions, e-uniformly convergent to f, such that $(s_n)_{n\in\mathbb{N}}$ is increasing and $(t_n)_{n\in\mathbb{N}}$ is decreasing.*

@@ ADD @@

Chapter 26

Operator Theory

The main reference for this chapter is Blackadar (2013).

Chapter Contents

26.1 Operators **505**
- 26.1.1 Operator topologies 505
- 26.1.2 Bounded operators 505
- 26.1.3 Self-adjoint operators 507
- 26.1.4 Compact operators 508
- 26.1.5 Spectrum 510
- 26.1.6 Spectral theorem & functional calculus 510
- 26.1.7 Dixmier trace 513

26.2 C^*-algebras **514**
- 26.2.1 Involutive algebras 514
- 26.2.2 Positive maps 516
- 26.2.3 Traces 517
- 26.2.4 Representations 518
- 26.2.5 Gel'fand duality 518
- 26.2.6 Hilbert modules ♣ 521

26.3 von Neumann algebras **522**
- 26.3.1 Factors 523

26.1 Operators

26.1.1 Operator topologies

Definition 26.1.1 (Weak operator topology). The topology generated by the seminorms $\{T \mapsto |\lambda(Tv)| \mid v \in V, \lambda \in V^*\}$. A net of linear operators $(T_\alpha)_{\alpha \in I}$ on a space V converges to a linear operator T in the weak (operator) topology if $T_\alpha v \longrightarrow Tv$ for all $v \in V$ in the weak topology.

In the case of Hilbert spaces, one can simplify the above definition using Riesz's representation theorem (Theorem 25.2.7). The weak operator topology on a Hilbert space is generated by the seminorms $\{T \mapsto |\langle Tv \mid w\rangle| \mid v, w \in \mathcal{H}\}$.

Definition 26.1.2 (Strong operator topology). The topology generated by the seminorms $\{T \to \|Tv\| \mid v \in V\}$. A net of linear operators $(T_n)_{n\in\mathbb{N}}$ on a space V converges to a linear operator T in the strong (operator) topology if $T_\alpha v \longrightarrow Tv$ for all $v \in V$ in the norm (or strong) topology.

Definition 26.1.3 (Operator norm). The operator norm of L is defined as follows:

$$\|T\|_{\text{op}} = \inf\{\lambda \in \mathbb{R} \mid \forall v \in V : \|Tv\|_W \leq \lambda \|v\|_V\}. \tag{26.1}$$

Equivalent definitions of the operator norm are:

$$\|T\|_{\text{op}} = \sup_{\|v\|\leq 1} \|Tv\| = \sup_{\|v\|=1} \|Tv\| = \sup_{v\neq 0} \frac{\|Tv\|}{\|v\|}. \tag{26.2}$$

Definition 26.1.4 (Norm topology[1]). A sequence of linear operators $(T_n)_{n\in\mathbb{N}}$ on a space V converges to a linear operator T in the norm topology if the sequence $(\|T_n - T\|)_{n\in\mathbb{N}}$ converges to 0. (Sequences suffice since the norm topology is metrizable and, therefore, sequential by Property 6.1.26.)

26.1.2 Bounded operators

Definition 26.1.5 (Bounded operator). Let $T : V \to W$ be a linear operator between two normed spaces. The linear operator is said to be bounded if it satisfies

$$\|T\|_{\text{op}} < +\infty. \tag{26.3}$$

Notation 26.1.6. The space of bounded linear operators from V to W is denoted by $\mathcal{B}(V, W)$.

Property 26.1.7. If V is a Banach space, then $\mathcal{B}(V)$ is also a Banach space.

The following property reduces the problem of continuity to that of boundedness (or vice versa).

[1] Also called the **uniform (operator) topology**.

Property 26.1.8. Consider a linear operator $T \in \mathcal{L}(V, W)$. The following statements are equivalent:

- T is bounded.
- T is continuous at 0.
- T is continuous on V.
- T is uniformly continuous on V.
- T maps bounded sets to bounded sets.

Property 26.1.9 (Bounded eigenvalues). The eigenvalues of a bounded operator are bounded by its operator norm. Furthermore, every bounded linear operator on a Banach space has at least one eigenvalue.

Property 26.1.10 (BLT theorem[2]). Consider a bounded linear operator $f : X \to W$, where X is a dense subset of a normed space V and W is a Banach space. There exists a unique extension $F : V \to W$ such that $\|f\|_{\text{op}} = \|F\|_{\text{op}}$.

Definition 26.1.11 (Schatten class operator). Consider the space of bounded linear operators on a Hilbert space $\mathcal{H}$. The **Schatten p-norm**, for $p \in [1, +\infty[$, is defined as

$$\|T\|_p := \mathrm{tr}\left(\sqrt{T^*T}^{\,p}\right)^{1/p}. \tag{26.4}$$

Linear operators for which this norm is finite constitute the pth Schatten class $\mathcal{J}_p$.

Property 26.1.12. The Schatten classes are Banach spaces with respect to the associated Schatten norms.

Example 26.1.13 (Trace-class operator). The space of trace-class operators on a Hilbert space $\mathcal{H}$ is defined as follows:

$$\mathcal{B}_1(\mathcal{H}) := \{T \in \mathcal{B}(\mathcal{H}) \mid \mathrm{tr}(|T|) < +\infty\}, \tag{26.5}$$

where the trace functional was defined in Definition 25.2.23 and $|T| := \sqrt{T^*T}$.

The following theorem can be seen as the analogue of Riesz's theorem for trace-class operators.

Property 26.1.14. For every bounded linear functional ρ on the space of trace-class operators $\mathcal{B}_1(\mathcal{H})$, there exists a unique bounded linear operator $T \in \mathcal{B}(\mathcal{H})$ such that

$$\rho(S) = \mathrm{tr}(ST) \tag{26.6}$$

for all $S \in \mathcal{B}_1(\mathcal{H})$. This implies that $\mathcal{B}_1^*(\mathcal{H})$ and $\mathcal{B}(\mathcal{H})$ are isometrically equivalent.

[2] BLT stands for 'bounded linear transformation'.

The previous property allows for the following definition.

Definition 26.1.15 (Weak-* operator topology). The weak-* (also **ultraweak** or **σ-weak**) topology on $\mathcal{B}(\mathcal{H})$ with respect to the trace-class operators $\mathcal{B}_1(\mathcal{H})$. This is also called the σ-weak topology on $\mathcal{B}(\mathcal{H})$.

Example 26.1.16 (Hilbert–Schmidt operator). Consider the Hilbert–Schmidt norm $\|\cdot\|_2$ from Formula 22.4.7. A linear operator $T \in \mathcal{B}(\mathcal{H})$ is said to be a Hilbert–Schmidt operator if it satisfies

$$\|T\|_2 < +\infty\,. \tag{26.7}$$

This space is closed under taking adjoints.

A more general, but still well-behaved, class of linear operators is the space of closed operators.

Definition 26.1.17 (Closed operator). A linear operator $T : V \to W$ such that for every sequence $(v_n)_{n\in\mathbb{N}}$ in $\mathsf{dom}(T)$ converging to $v \in V$, where $f(v_n)$ converges to $w \in W$, one finds that $v \in \mathsf{dom}(T)$ and $Tv = w$.

Equivalently, one can define a closed linear operator as a linear operator for which its graph is a closed subset in the direct sum $V \oplus W$.

Definition 26.1.18 (Closure). Let $T : V \to W$ be a linear operator. Its closure (if it exists) is the closed linear operator $\overline{T}$ such that the graph of $\overline{T}$ is the closure of the graph of T in $V \oplus W$.

Theorem 26.1.19 (Closed graph theorem). *A linear operator on a Banach space is closed if and only if it is bounded.*

26.1.3 Self-adjoint operators

There is a multitude of different notions available in the literature that try to indicate in what sense a linear operator is related to its adjoint (not everyone agrees on the definitions). Here, an overview is given in the case of Hilbert spaces where linear operators are allowed to be unbounded.

Definition 22.3.5 for finite-dimensional spaces can be generalized as follows.

Definition 26.1.20 (Adjoint). Let T be a linear operator on a Hilbert space $\mathcal{H}$. A linear operator $T^\dagger$ is said to be the (Hermitian) adjoint of T if the following conditions are satisfied:

1. $\langle v \mid Tw\rangle = \langle T^\dagger v \mid w\rangle$ for all $v \in \mathsf{dom}(T^\dagger)$ and $w \in \mathsf{dom}(T)$.
2. Every other linear operator satisfying this property is a restriction of $T^\dagger$ (i.e. the domain of T^* is maximal with respect to the above property).

Property 26.1.21. Let T be a bounded linear operator. Its adjoint is also bounded and $\|T\|_{\text{op}} = \|T^\dagger\|_{\text{op}}$.

Definition 26.1.22 (Symmetric operator). A linear operator T on a Hilbert space $\mathcal{H}$ such that $\mathsf{dom}(T) \subseteq \mathsf{dom}(T^\dagger)$ and $T = T^\dagger|_{\mathsf{dom}(T)}$.

Definition 26.1.23 (Self-adjoint operator). A linear operator T on a Hilbert space $\mathcal{H}$ such that $\mathsf{dom}(T)$ is dense in $\mathcal{H}$ and $T = T^\dagger$.

The notion of Hermitian operator is the one where almost nobody agrees upon its definition. Here, the definition from Schreiber (2008) is chosen.

Definition 26.1.24 (Hermitian operator). A bounded, symmetric operator.

Theorem 26.1.25 (Hellinger–Toeplitz). *A self-adjoint operator on a Hilbert space $\mathcal{H}$ is bounded if and only if its domain is all of $\mathcal{H}$.*

Theorem 26.1.26 (Stone). *Consider a strongly continuous unitary one-parameter group, i.e. a family of unitary operators $U : \mathbb{R} \to \mathsf{U}(\mathcal{H})$ such that*

- *U is continuous in the strong operator topology:*

$$\lim_{t \to t_0} U(t)v = U(t_0)v \tag{26.8}$$

  for all $t_0 \in \mathbb{R}, v \in \mathcal{H}$.

- *U is a one-parameter group in the sense of Definition 32.1.9.*

There exists a self-adjoint operator T such that $U(t) = e^{itT}$. Furthermore, the linear operator T is bounded if and only if U is continuous in the norm topology.

Definition 26.1.27 (Generator). The linear operator T in the preceding theorem is called the (infinitesimal) generator of the family U. It can be obtained through a formal derivative:

$$T = \left.\frac{\mathrm{d}U(t)}{\mathrm{d}t}\right|_{t=0}. \tag{26.9}$$

26.1.4 Compact operators

Definition 26.1.28 (Compact operator). Let V, W be Banach spaces. A linear operator $T : V \to W$ is said to be compact if the image of any bounded set in V is relatively compact (Definition 6.5.17).

Alternative Definition 26.1.29 (Compact operator). Let V, W be Banach spaces. A linear operator $T : V \to W$ is compact if, for every bounded sequence $(v_n)_{n\in\mathbb{N}}$ in V, the sequence $(Tv_n)_{n\in\mathbb{N}} \subset W$ has a convergent subsequence.

Notation 26.1.30. The space of compact, bounded linear operators between Banach spaces V, W is denoted by $\mathcal{B}_0(V, W)$. If $V = W$, this is abbreviated as $\mathcal{B}_0(V)$ as usual.

Property 26.1.31. $\mathcal{B}_0(V)$ is a two-sided ideal in the (Banach) algebra $\mathcal{B}(V)$. Moreover, $\mathcal{B}_0(V) = \mathcal{B}_1^*(V)$.

Property 26.1.32 (Finite-rank operators). All finite-rank operators are compact. In fact, the space of compact operators is the norm closure of that of finite-rank operators.

Property 26.1.33. Every compact operator is bounded.

Corollary 26.1.34. Every linear map between finite-dimensional Banach spaces is bounded.

Property 26.1.35. If T is a compact self-adjoint operator on a Hilbert space, then $-\|T\|$ or $\|T\|$ are an eigenvalue of T. Furthermore, the set of nonzero eigenvalues is either finite or converges to 0.

Definition 26.1.36 (Calkin algebra). Consider the algebra $\mathcal{B}(V)$ of bounded linear operators on V together with its two-sided ideal $\mathcal{B}_0(V)$ of compact operators. The quotient algebra $\mathcal{Q}(V) = \mathcal{B}(V)/\mathcal{B}_0(V)$ is called the Calkin algebra of V.

Definition 26.1.37 (Fredholm operator). A bounded linear operator for which the kernel and cokernel are finite-dimensional. The space of Fredholm operators on a space V is denoted by $\mathfrak{F}(V)$.

By the following theorem, one can characterize Fredholm operators using the Calkin algebra.

Property 26.1.38 (Atkinson). A bounded linear operator $T : V \to W$ is a Fredholm operator if and only if it is invertible in the Calkin algebra, i.e. there exists a bounded linear operator $S : W \to V$ and compact operators C_1, C_2 such that $\mathbb{1}_W - TS = C_1$ and $\mathbb{1}_V - ST = C_2$. S is called the **parametrix** of T.

Definition 26.1.39 (Fredholm index). The index of a Fredholm operator F is defined as follows:

$$\mathsf{ind}(F) := \dim \ker(F) - \dim \operatorname{coker}(F)\,. \tag{26.10}$$

Property 26.1.40. The induced function

$$\mathsf{ind} : \pi_0\big(\mathfrak{F}(\mathcal{H})\big) \to \mathbb{Z} \tag{26.11}$$

is a group isomorphism:

- $\mathsf{ind}(F^*) = -\mathsf{ind}(F)$, and
- $\mathsf{ind}(FG) = \mathsf{ind}(F) + \mathsf{ind}(G)$.

This theorem is generalized in *K-theory* (see Chapter 41) by the *Atiyah–Jänich theorem* 41.2.17.

26.1.5 Spectrum

Definition 26.1.41 (Resolvent operator). Consider an operator $T \in \mathcal{B}(V)$ on a normed space V. The resolvent operator T_λ for some $\lambda \in \mathbb{C}$ is defined as the linear operator $(T - \lambda \mathbb{1}_V)^{-1}$.

Definition 26.1.42 (Resolvent set). The resolvent set $\rho(T)$ consists of all scalars $\lambda \in \mathbb{C}$ for which the resolvent operator of A is a bounded linear operator on a dense subset of V. These scalars λ are called **regular values** of T.

Definition 26.1.43 (Spectrum). The set of scalars $\mu \in \mathbb{C} \backslash \rho(T)$ is called the spectrum $\sigma(T)$.

From Definition 22.5.2, it is clear that every eigenvalue of T belongs to the spectrum of T. The converse, however, is not true. This is remedied by introducing the following concepts.

Definition 26.1.44 (Point spectrum). The set of scalars $\mu \in \mathbb{C}$ for which $T - \mu \mathbb{1}_V$ fails to be injective is called the point spectrum $\sigma_p(T)$. This set coincides with the set of eigenvalues of T.

Definition 26.1.45 (Continuous spectrum). The set of scalars $\mu \in \mathbb{C}$ for which $T - \mu \mathbb{1}_V$ is injective with dense image but fails to be surjective is called the continuous spectrum of T.

Definition 26.1.46 (Residual spectrum). The set of scalars $\mu \in \mathbb{C}$ for which $T - \mu \mathbb{1}_V$ is injective but fails to have a dense image is called the residual spectrum $\sigma(T)$.

Definition 26.1.47 (Essential spectrum). The set of scalars $\mu \in \mathbb{C}$ for which $T - \mu \mathbb{1}_V$ is not a Fredholm operator is called the essential spectrum $\sigma_{\text{ess}}(T)$.

From Atkinson's theorem 26.1.38, one can derive the following result.[3]

Property 26.1.48. Let T be a bounded linear operator and let C be a compact operator. The essential spectra of T and $T + C$ coincide.

Property 26.1.49 (Bounded spectrum). A self-adjoint operator is bounded if and only if its spectrum is bounded. Furthermore, it is positive if and only if its spectrum lies in $\mathbb{R}^+$.

26.1.6 Spectral theorem & functional calculus

This section focuses on the algebra of bounded operators $\mathcal{B}(\mathcal{H})$ on a (complex) Hilbert space $\mathcal{H}$.

Property 26.1.50 (Closed subspaces). There exists a bijection between the set of closed

[3] In fact, one could (equivalently) define the essential spectrum in terms of the Calkin algebra using Atkinson's theorem. Then, this property would be an obvious consequence.

subspaces of $\mathcal{H}$ and the set of orthogonal projections in $\mathcal{B}(\mathcal{H})$. Furthermore, if the projection p corresponds to a subspace $\mathcal{H}_p$, the projection $\mathbb{1}_{\mathcal{H}} - p$ corresponds to the orthogonal complement $\mathcal{H}_p^{\perp}$.

Definition 26.1.51 (Projection-valued measure). Consider a measurable space (X, Σ). A projection-valued measure (PVM) or **spectral measure**[4] on X is a map $P : \Sigma \to \mathcal{B}(\mathcal{H})$ satisfying the following conditions:[5]

1. P_E is a projection for all $E \in \Sigma$,
2. $P_X = \mathbb{1}_{\mathcal{H}}$,
3. $P_A P_B = P_{A \cap B}$, and
4. for all disjoint $(E_n)_{n \in \mathbb{N}} \subset \Sigma$:

$$\sum_{n=0}^{+\infty} P_{E_n} = P_{\cup_{n \in \mathbb{N}} E_n} \,. \tag{26.12}$$

Property 26.1.52. Let P be a spectral measure on (X, Σ). For every two elements $v, w \in \mathcal{H}$, the map

$$E \mapsto \mu^P_{v,w}(E) := \langle v \mid P_E w \rangle \tag{26.13}$$

defines a (complex) measure $\mu^P_{v,w}$ on X. The square of the norm of an element $v \in \mathcal{H}$ is then simply given by $\mu^P_{v,v}(X)$ due to the second condition above.

Property 26.1.53. Let $f : X \to \mathbb{C}$ be a measurable function on a measurable space (X, Σ). Given a spectral measure P, one defines Δ_f to be the set of all $v \in \mathcal{H}$ for which $f \in L^2(X, \mu^P_{v,v})$. This set is dense in $\mathcal{H}$. Moreover, the map

$$\int_X f(\lambda)\, dP(\lambda) : \Delta_f \to \mathcal{H} \tag{26.14}$$

defined by

$$\left\langle v \middle| \int_X f(\lambda)\, dP(\lambda) w \right\rangle := \int_X f(\lambda)\, d\mu^P_{v,w}(\lambda) \tag{26.15}$$

is closed and normal. It also satisfies the following two equalities:

$$\left(\int_X f(\lambda)\, dP(\lambda) \right)^{\dagger} = \int_X \overline{f(\lambda)}\, dP(\lambda) \,, \tag{26.16}$$

$$\left\| \int_X f(\lambda)\, dP(\lambda) v \right\|^2 = \int_X |f(\lambda)|^2\, d\mu^P_{v,v}(\lambda) \,. \tag{26.17}$$

If f is bounded, the above operator is bounded by the supremum norm of f:

$$\left\| \int_X f(\lambda)\, dP(\lambda) \right\| \leq \|f\|_\infty \,. \tag{26.18}$$

[4]Sometimes also called a **resolution of the identity**.

[5]The third property can in fact be shown to follow from the others.

Theorem 26.1.54 (Spectral theorem). *Let A be a self-adjoint operator on a Hilbert space $\mathcal{H}$. There exists a unique spectral measure $P_A : \Sigma_{\mathbb{R}} \to \mathcal{B}(\mathcal{H})$ on the Borel σ-algebra of the real line such that*

$$A = \int_{\mathbb{R}} \lambda \, dP_A(\lambda) \,. \tag{26.19}$$

If $\mathbb{R}$ is replaced by $\mathbb{C}$, this theorem also holds for normal operators. If A is compact, the measure becomes purely atomic, i.e. there exists a countable orthonormal basis of $\mathcal{H}$ consisting of eigenvectors of A such that $\lambda_i \longrightarrow 0$:

$$A = \sum_{i=1}^{+\infty} \lambda_i e_i^* \otimes e_i \,. \tag{26.20}$$

Property 26.1.55 (Spectrum and support). The spectrum of a self-adjoint operator A coincides with the support of its associated spectral measure P_A. A number $\lambda \in \mathbb{R}$ belongs to the point spectrum of A if and only if P_A does not vanish on $\{\lambda\}$. A number $\lambda \in \mathbb{R}$ belongs to the continuous spectrum of A if P_A vanishes on $\{\lambda\}$ but is nonvanishing on any open set containing λ.

Definition 26.1.56 (Singular value). Let A be a compact operator. The singular values of A are given by the square roots of the eigenvalues of the self-adjoint (and compact) operator A^*A. If A is self-adjoint, its singular values and eigenvalues coincide.

Definition 26.1.57 (Measurable functional calculus). Let (X, Σ) be a measurable space and let $\mathcal{H}$ be a Hilbert space. A measurable functional calculus $(\Phi, \mathcal{H})$ on (X, Σ) is an assignment

$$\Phi : \mathsf{Meas}(X, \mathbb{C}) \to \mathcal{B}_0(\mathcal{H}) \tag{26.21}$$

satisfying the following conditions:

1. **Unitality**: $\Phi(1) = \mathbb{1}_{\mathcal{H}}$.
2. **Sublinearity**: $\Phi(\lambda f + g) \subseteq \lambda \Phi(f) + \Phi(g)$ and the equality holds if either operator is bounded.
3. **Submultiplicativity**: $\Phi(f)\Phi(g) \subseteq \Phi(fg)$ and the equality holds if either operator is bounded. Moreover, the product is commutative if either operator is bounded.
4. **Density**:[6] $\Phi(f)$ is densely defined.
5. **Involutivity**: $\Phi\left(\overline{f}\right) = \overline{\Phi(f)}$.
6. **Boundedness**: $\Phi(f)$ is bounded if f is bounded.
7. **Convergence**: If the bounded sequence $f_n \longrightarrow f$ pointwise, then $\Phi(f_n) \longrightarrow \Phi(f)$ strongly.

[6] This condition is actually redundant.

Property 26.1.58. If $(\Phi, \mathcal{H})$ is a measurable functional calculus on (X, Σ), then

$$P_\Phi : \Sigma \to \mathcal{B}(\mathcal{H}) : E \mapsto \Phi(\mathbb{1}_E) \tag{26.22}$$

is a projection-valued measure. Conversely, every projection-valued measure P gives rise to a measurable functional calculus Φ such that $P = P_\Phi$.

Definition 26.1.59 (Borel functional calculus). A functional calculus on (X, Σ) where X is a topological space and Σ is its Borel σ-algebra. If $X \subseteq \mathbb{C}$ and the identity function $\mathbb{1}_\mathbb{C} : z \mapsto z$ is measurable, $(\Phi, \mathcal{H})$ is called a Borel functional calculus for the operator A if $\Phi(\mathbb{1}_\mathbb{C}) = A$.

The above properties allow to compose self-adjoint operators with (measurable) functions similar to how one can compute $f(X)$ for finite-dimensional operators by applying f to the eigenvalues of X.

Formula 26.1.60 (Borel functional calculus). Consider a bounded operator $A \in \mathcal{B}(\mathcal{H})$. Let $f : \sigma(A) \to \mathbb{C}$ be a measurable function (with respect to the restriction of the Borel algebra on $\mathbb{R}$) and let $g : \mathbb{R} \to \mathbb{C}$ be any other measurable function that coincides with f on $\sigma(A)$.

$$f(A) := \int_{\sigma(A)} f(\lambda)\, dP_A(\lambda) = \int_\mathbb{R} g(\lambda)\, dP_A(\lambda) =: g(A)\,. \tag{26.23}$$

The Spectral Theorem 26.1.54 can also be stated in terms of functional calculus.

Theorem 26.1.61 (Spectral theorem). *A normal operator $A \in \mathcal{B}(\mathcal{H})$ on a Hilbert space $\mathcal{H}$ induces a unique Borel functional calculus on $\sigma(A)$.*

26.1.7 Dixmier trace

Definition 26.1.62 (Dixmier ideal). For every compact operator, one can sort the singular values in decreasing order. The Dixmier ideal is defined as follows:

$$\mathfrak{D}(\mathcal{H}) := \left\{ A \in \mathsf{End}_0(\mathcal{H}) \mid \sigma_n(A) = O(\ln(n)) \right\}, \tag{26.24}$$

where

$$\sigma_n(A) := \sum_{i=1}^{n} s_i(A)\,. \tag{26.25}$$

The following functional gives an alternative for the ordinary trace on operators that are not trace-class.

Definition 26.1.63 (Dixmier trace). For every element in $A \in \mathfrak{D}(\mathcal{H})$, the Dixmier trace is defined as follows:

$$\mathsf{tr}_\mathfrak{D}(A) := \omega\left(\left\{ \frac{\sigma_n(A)}{1 + \ln(n)} \,\middle|\, n \in \mathbb{N} \right\} \right), \tag{26.26}$$

where $\omega : \ell^\infty(\mathbb{N}) \to \mathbb{R}$ is a linear functional satisfying the following conditions:

1. **Positivity**: $\omega(s) \geq 0$ if $s_n \geq 0$ for all $n \in \mathbb{N}$.
2. **Convergence**: $\omega(s) = \lim_{n\to\infty} s_n$ if s is convergent.
3. **Dilation-invariant**: Let $D : \ell^\infty(\mathbb{N}) \to \ell^\infty(\mathbb{N})$ be the following operator:

$$D : (s_1, s_2, \ldots) \mapsto (s_1, s_1, s_2, s_2, \ldots). \tag{26.27}$$

   The state ω is dilation-invariant if $\omega(s) = \omega\big(D(s)\big)$ for all $s \in \ell^\infty(\mathbb{N})$.

If for an operator $A \in \mathfrak{D}(\mathcal{H})$ the Dixmier trace is independent of the chosen state, it is said to be **measurable**.

Property 26.1.64. The Dixmier trace has the following properties:

- $\mathrm{tr}_{\mathfrak{D}}(A + B) = \mathrm{tr}_{\mathfrak{D}}(A) + \mathrm{tr}_{\mathfrak{D}}(B)$ for all $A, B \in \mathfrak{D}(\mathcal{H})$.
- If A is bounded and B is an element of the Dixmier ideal, then $\mathrm{tr}_{\mathfrak{D}}(AB) = \mathrm{tr}_{\mathfrak{D}}(BA)$.
- The Dixmier ideal vanishes on trace-class operators (in particular, it is **singular**, i.e. it vanishes on finite-rank operators):

$$A \in \mathcal{B}_1(\mathcal{H}) \implies \mathrm{tr}_{\mathfrak{D}}(A) = 0\,. \tag{26.28}$$

26.2 C^*-algebras

26.2.1 Involutive algebras

Definition 26.2.1 (Involutive algebra). An involutive algebra is an associative algebra A over a commutative involutive ring $(R, \overline{\cdot})$ together with an algebra involution $\cdot^* : A \to A$ such that:

1. $(a + b)^* = a^* + b^*$,
2. $(ab)^* = b^* a^*$, and
3. $(\lambda a)^* = \overline{\lambda} a^*$,

for all $a, b \in A$ and $\lambda \in R$. These algebras are also sometimes called **$*$-algebras**.

Definition 26.2.2 (Isometry). An element a of a $*$-algebra A satisfying

$$a^* a = 1\,, \tag{26.29}$$

where 1 is the unit of A.

Definition 26.2.3 (Normal element). An element of a $*$-algebra that commutes with its adjoint:

$$a^* a = a a^*\,. \tag{26.30}$$

Property 26.2.4. Every element in a $*$-algebra can be decomposed as the sum of two normal elements:

$$a = \frac{1}{2}\big((a + a^*) + (a - a^*)\big)\,. \tag{26.31}$$

This implies that a linear operator defined on normal elements extends uniquely to the whole algebra.

Definition 26.2.5 (Projection). An element p of a $*$-algebra such that

$$p = p^2 = p^*\,. \tag{26.32}$$

This terminology reflects the property that in an algebra of bounded operators on a Hilbert space, the projections are exactly the operators associated to a orthogonal projections (cf. Section 25.2.3).

Definition 26.2.6 (C^*-algebra). A C^*-algebra is an involutive Banach algebra A (Definition 25.1.9) such that the **C^*-identity**

$$\|a^*a\| = \|a\|\,\|a^*\| \tag{26.33}$$

is satisfied for all $a \in A$.

The Artin–Wedderburn theorem 2.4.71 implies the following decomposition.

Theorem 26.2.7. *Let A be a finite-dimensional C^*-algebra over the field $\mathfrak{K}$. There exist unique integers N and $d_1, \dots, d_N$ such that*

$$A \cong \bigoplus_{i=1}^{N} M_{d_i}(\mathfrak{K})\,. \tag{26.34}$$

This implies that every C^*-algebra can be represented using block matrices.

Example 26.2.8 (Bounded operators). Let $\mathcal{H}$ be a finite-dimensional Hilbert space. The space of bounded operators $\mathcal{B}(\mathcal{H})$ is a C^*-algebra.

Property 26.2.9. Every norm-closed $*$-subalgebra of a C^*-algebra is a C^*-algebra.

Definition 26.2.10 (H^*-algebra). A Hilbert space $\mathcal{H}$ equipped with a unital $*$-algebra structure such that $*$-adjoint and multiplicative adjoints coincide:

1. $\langle ab \mid c\rangle = \langle b \mid a^*c\rangle$, and
2. $\langle ab \mid c\rangle = \langle a \mid cb^*\rangle$,

for all $a, b, c \in \mathcal{H}$.

Example 26.2.11 (Linear operators). The canonical example of H^*-algebras is given by the algebra of linear operators on a Hilbert space $\mathcal{H}$, where the involution is given by

taking adjoints and the inner product is the Hilbert–Schmidt inner product induced by the norm (Formula 22.4.7) (up to a factor $k > 0$):

$$\langle f \mid g \rangle_{\mathrm{HS}} := k \ \mathrm{tr}(f^* g) \,. \tag{26.35}$$

The resulting space is denoted by $L^2(\mathcal{H}, k)$. A result, analogous to the Artin–Wedderburn theorem 2.4.71, states that every H^*-algebra can be decomposed as an orthogonal direct sum of finitely many algebras of the form $L^2(\mathcal{H}_i, k_i)$.

26.2.2 Positive maps

Definition 26.2.12 (Positive element). A self-adjoint element of a C^*-algebra whose spectrum is contained in $[0, +\infty[$. The cone of all positive elements in A is often denoted by A^+.

Property 26.2.13 (Positive decomposition). An element $a \in A$ is positive if and only if $a = b^* b$ for some element $b \in A$. A normal element is positive if and only if it is self-adjoint.

Definition 26.2.14 (Positive map). A morphism $T : A \to B$ of C^*-algebras such that $T(A^+) \subseteq B^+$.

Definition 26.2.15 (Completely positive map). A morphism $T : A \to B$ of C^*-algebras such that the following map is positive for all $k \in \mathbb{N}$:

$$\mathbb{1}_k \otimes T : \mathbb{C}^{k \times k} \otimes A \to \mathbb{C}^{k \times k} \otimes B \,. \tag{26.36}$$

If T satisfies this condition only up to an integer $n \in \mathbb{N}$, it is said to be **n-positive**.

Definition 26.2.16 (State). A positive linear functional of unit norm on a C^*-algebra.

Property 26.2.17 (Convexity). The set of states is a convex set. The extreme points of this set are called **pure states**, all other elements are called **mixed states**.

Definition 26.2.18 (Positivity-improving map). A positive map T that satisfies

$$a \geq 0 \implies T(a) > 0 \,. \tag{26.37}$$

Definition 26.2.19 (Ergodic map). A positive map T that satisfies

$$\forall a > 0 : \exists t_a \in \mathbb{R}_0 : \exp(t_a T)\, a > 0 \,. \tag{26.38}$$

The following theorem can be seen as a Jordan algebra-theoretic analogue of the Gel'fand–Naimark theorem 26.2.42.

Theorem 26.2.20 (Alfsen–Shultz). *The state spaces of two C^*-algebras are isomorphic if and only if the algebras are isomorphic as (special) Jordan algebras (Definition 22.7.8).*

Definition 26.2.21 (Cuntz algebra). The nth Cuntz algebra $\mathcal{O}_n$ is defined as the (universal) unital C^*-algebra generated by $n \in \mathbb{N}$ isometries s_i under the additional relation

$$\sum_{i=1}^{n} s_i s_i^* = 1\,, \tag{26.39}$$

where 1 is the unit element.

26.2.3 Traces

Definition 26.2.22 (Trace). Let A be a C^*-algebra. A trace on A is a linear functional that satisfies the following conditions:

1. **Positivity**: $\mathrm{tr}(A^+) \geq 0$, and
2. **Tracial**: $\mathrm{tr}(ab) = \mathrm{tr}(ba)$ for all $a, b \in A$.

Remark 26.2.23. The tracial property could have been replaced by the following equivalent, but seemingly weaker, condition: $\mathrm{tr}(a^*a) = \mathrm{tr}(aa^*)$ for all $a \in A$.

Definition 26.2.24 (Adjoint map). Assume that a trace functional tr is given on a C^*-algebra and consider a continuous linear operator T defined on the Schatten class $\mathcal{J}_p$ (Definition 26.1.11). One can define the adjoint map T^* on $\mathcal{J}_q$ whenever p, q are Hölder conjugate (Formula 18.3.7). This adjoint is given by the following equation:

$$\mathrm{tr}\big((T^*a)b\big) = \mathrm{tr}\big(a^*Tb\big)\,, \tag{26.40}$$

where $a \in \mathcal{J}_q$ and $b \in \mathcal{J}_p$.

Definition 26.2.25 (Trace-preserving map). A linear operator $T : A \to B$ is said to be trace-preserving if it satisfies

$$\mathrm{tr}_B(Ta) = \mathrm{tr}_A(a) \tag{26.41}$$

for all trace-class elements $a \in A$. Using the above definition, it is easy to see that, on a unital C^*-algebra, this is equivalent to

$$T^*1 = 1\,, \tag{26.42}$$

i.e. to its adjoint being unital.

Property 26.2.26. A completely positive, trace-preserving map T satisfies:

$$\|T\|_1 = 1\,, \tag{26.43}$$

where the subscript 1 indicates that this operator is defined on trace-class elements.

Property 26.2.27 (States). Whenever a C^*-algebra is commutative, every state defines a trace and, after a suitable normalization, every trace defines a state. However, for noncommutative C^*-algebras, only the latter implication holds.

26.2.4 Representations

Definition 26.2.28 (C^*-representation). A representation of a C^*-algebra A on a Hilbert space $\mathcal{H}$ is a unital $*$-morphism $\pi : A \to \mathcal{B}(\mathcal{H})$.

Definition 26.2.29 (Normal state). Consider a Hilbert space $\mathcal{H}$ and a C^*-representation $\pi : A \to \mathcal{B}(\mathcal{H})$. A normal state ω on A is a state such that there exists a trace-class operator $\rho \in \mathcal{B}_1(\mathcal{H})$ with the following property:

$$\omega(a) = \frac{\mathrm{tr}\big(\rho\pi(a)\big)}{\mathrm{tr}(\rho)} . \tag{26.44}$$

When $A = \mathcal{B}(\mathcal{H})$, the normal states are exactly the σ-weakly continuous states (Definition 26.1.15).

Definition 26.2.30 (Folium). Let $\pi : A \to \mathcal{B}(\mathcal{H})$ be a C^*-representation. The space of normal states of this representation is called its folium.

Definition 26.2.31 (Cyclic vector). A cyclic vector for a C^*-algebra representation $\pi : A \to \mathcal{B}(\mathcal{H})$ is a vector $\xi \in \mathcal{H}$ such that $\{\pi(a)\xi \mid a \in A\}$ is (norm-)dense in $\mathcal{H}$.

The injective tensor product of Banach spaces admits an equivalent construction in the case of C^*-algebras.

Definition 26.2.32 (Spatial tensor product). Let A, B be two C^*-algebras with faithful representations $\pi_A : A \to \mathcal{B}(\mathcal{H}_1)$ and $\pi_B : B \to \mathcal{B}(\mathcal{H}_2)$. The spatial tensor product $A \otimes_{\mathrm{sp}} B$ is the norm closure of the algebraic tensor product $A \otimes B$. It can be shown that this definition does not depend on the choice of representation and, moreover, coincides with the injective tensor product of the underlying Banach spaces as induced by Eq. (25.38).

26.2.5 Gel'fand duality

Definition 26.2.33 (Gel'fand spectrum). Consider a commutative C^*-algebra A (in fact any involutive algebra suffices). Its set of **characters**, i.e. the algebra morphisms $A \to \mathbb{C}$, can be equipped with a locally compact Hausdorff topology: the weak-* topology (Definition 25.1.3).

Property 26.2.34. The Gel'fand spectrum of a C^*-algebra is compact if and only if the algebra is unital.

Definition 26.2.35 (Gel'fand representation). Consider a C^*-algebra A and let Φ_A denote its Gel'fand spectrum. The **Gel'fand transformation** of an element $a \in A$ is defined as the morphism $\hat{a} : \Phi_A \to \mathbb{C}$ given by the following formula:

$$\hat{a}(\lambda) = \langle \lambda, a \rangle , \tag{26.45}$$

where $\langle \cdot, \cdot \rangle$ denotes the pairing between A and Φ_A. By definition of the Gel'fand spectrum, the functional $\hat{a}$ is continuous for all $a \in A$. The mapping $a \mapsto \hat{a}$ is called the Gel'fand representation of A.

Theorem 26.2.36 (Gel'fand–Naimark: Commutative case). *Let A be a commutative C^*-algebra. The Gel'fand representation gives an isometric $*$-isomorphism between A and the set $C_0(\Phi_A)$ of continuous complex-valued functions that vanish at infinity (Theorem 19.1.6) on its Gel'fand spectrum.*

Remark 26.2.37. In fact, the Gel'fand–Naimark theorem gives an equivalence between the category of commutative nonunital C^*-algebras and the category of locally compact Hausdorff spaces (with continuous functions that vanish at infinity).

Property 26.2.38 (Connectedness). A compact Hausdorff space is connected if and only if its algebra of continuous functions has no nontrivial projections.

Property 26.2.39 (Metrizability). A commutative C^*-algebra is separable if and only if its Gel'fand spectrum is metrizable.

Formula 26.2.40 (Spectrum). Recall Definition 26.1.43 of the spectrum of an operator. This is related to the spectrum of a unital, commutative C^*-algebra as follows:

$$\sigma(a) = \{a(x) \mid x \in \Phi_A\}, \tag{26.46}$$

for all $a \in A$.

Construction 26.2.41 (Gel'fand–Naimark–Segal). Let A be a C^*-algebra. Given a state ω on A, there exists a C^*-representation $\pi : A \to \mathcal{B}(D)$, where $D \subset \mathcal{H}$ is a dense subspace of a Hilbert space $\mathcal{H}$, such that the following conditions are satisfied:

- There exists a cyclic (unit) vector $\xi \in D$.
- For all elements $a \in A$, the following equality holds:

$$\omega(a) = \langle \pi(a)\xi \mid \xi \rangle. \tag{26.47}$$

@@ COMPLETE CONSTRUCTION @@

Theorem 26.2.42 (Gel'fand–Naimark: General case). *Every C^*-algebra is isometrically $*$-isomorphic to a norm closed (C^*-)algebra of bounded operators on a Hilbert space $\mathcal{H}$.*

The GNS construction can be generalized as follows.

Theorem 26.2.43 (Stinespring[7]). *Consider a linear operator $T : A \to \mathcal{B}(\mathcal{H})$ between C^*-algebras. It is completely positive if and only if there exists a C^*-representation $\pi : A \to \mathcal{B}(\mathcal{K})$ and a bounded operator $V : \mathcal{H} \to \mathcal{K}$ such that*

$$Ta = V^\dagger \pi(a) V \tag{26.48}$$

or, equivalently,

$$T^\dagger b = \pi^\dagger(VbV^\dagger)\,. \tag{26.49}$$

Moreover, V is an isometry if and only if T is unital and, hence, if $T^\dagger$ is trace-preserving by Definition 26.2.25. The triple $(\mathcal{K}, \pi, V)$ is often called a ***Stinespring representation*** *of T.*

Remark 26.2.44. Because the adjoint of a completely positive map is again completely positive, the above two characterizations can be used interchangeably. The latter is mostly used by mathematicians, whereas the former is better known in the physics literature.

The most common situation is where $\mathcal{H} := \mathcal{H}_1 = \mathcal{H}_2$ and

$$\pi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H} \otimes \mathcal{K}) : a \mapsto a \otimes b \tag{26.50}$$

for some density operator $b \in \mathcal{B}(\mathcal{K})$. The adjoint to tensoring by a density operator is taking the partial trace $\mathrm{tr}_{\mathcal{K}}$. This way, one obtains the following expressions:

$$Ta = V^\dagger(a \otimes b)V \tag{26.51}$$

and, equivalently,

$$T^\dagger b = \mathrm{tr}_{\mathcal{K}}(VbV^\dagger)\,. \tag{26.52}$$

Corollary 26.2.45 (GNS construction). The GNS construction follows from the Stinespring theorem by taking T to be a state (hence $\mathcal{H} = \mathbb{C}$).

Corollary 26.2.46 (Choi). Let $\mathcal{H}_1, \mathcal{H}_2$ be two finite-dimensional Hilbert spaces of dimensions $m, n \in \mathbb{N}$, respectively. A linear map $\Phi : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ is completely positive if and only if it can be expressed as

$$\Phi(a) = \sum_{i=1}^{mn} V_i^\dagger a V_i \tag{26.53}$$

for some bounded operators $V_i : \mathcal{H}_2 \to \mathcal{H}_1$. Furthermore, it is trace-preserving if and only if

$$\sum_{i=1}^{mn} A_i^\dagger A_i = \mathbb{1}\,. \tag{26.54}$$

This decomposition is also often called an **operator-sum decomposition** and the operators $\{V_i\}_{i \le mn}$ are called **Kraus operators** of Φ.

[7] Sometimes called the **Stinespring dilation theorem** or **Stinespring factorization theorem**.

26.2.6 Hilbert modules ♣

Definition 26.2.47 (Hilbert C^*-module). Let A be a C^*-algebra and H a vector space. H is a (right) pre-Hilbert A-module if there exists a map

$$\langle \cdot \mid \cdot \rangle_A : H \times H \to A\,, \tag{26.55}$$

sometimes called the **A-inner product**, satisfying the following conditions:

1. **Conjugate linearity**: For all $x, y, z \in H$ and $\lambda \in \mathbb{C}$:

$$\langle x \mid \lambda y + z \rangle_A = \langle x \mid y \rangle_A + \overline{\lambda} \langle x \mid z \rangle_A\,, \tag{26.56}$$

$$\langle x \mid y \rangle_A = \langle y \mid x \rangle_A^*\,. \tag{26.57}$$

2. **Nondegeneracy**: For all $x \in H$:

$$\langle x \mid x \rangle_A \geq 0\,, \tag{26.58}$$

$$\langle x \mid x \rangle_A = 0 \iff x = 0\,. \tag{26.59}$$

3. **Equivariance**: For all $x, y \in H$ and $a \in A$:

$$\langle x \mid y \rangle_A\, a = \langle x \mid y \cdot a \rangle_A\,. \tag{26.60}$$

A left module is obtained by requiring linearity and (left-)equivariance in the first argument. The completion of a pre-Hilbert module with respect to the norm $x \mapsto \sqrt{\|\langle x|x\rangle_A\|}$ is called a **Hilbert A-module**.

Definition 26.2.48 (Hilbert C^*-bimodule). Let A, B be C^*-algebras and consider a Hilbert C^*-module H over B. H is called a Hilbert (A, B)-bimodule if it admits a left $*$-representation of A such that

$$\langle a^* \cdot x \mid y \rangle_B = \langle x \mid a \cdot y \rangle_A \tag{26.61}$$

for all $a \in A$, i.e. adjoints in A correspond to adjoints with respect to the B-inner product.

Definition 26.2.49 (Fredholm module). Let $\pi : A \to \mathsf{End}(\mathcal{H})$ be a $*$-representation of a unital C^*-algebra on a Hilbert space $\mathcal{H}$. When equipped with an operator $F \in \mathsf{End}(\mathcal{H})$, this gives an (**odd**) Fredholm module if

$$F = F^* \qquad \text{and} \qquad F^2 = \mathbb{1}_{\mathcal{H}} \qquad \text{and} \qquad [F, \pi(A)] \subseteq \mathcal{B}_0(\mathcal{H})\,. \tag{26.62}$$

The Fredholm operator is said to be **even** if $\mathcal{H}$ is a super-Hilbert space with grading $\Gamma \in \mathsf{End}(\mathcal{H})$ satisyfing

$$\Gamma^* = \Gamma \qquad \text{and} \qquad \Gamma^2 = \mathbb{1}_{\mathcal{H}} \qquad \text{and} \qquad [\Gamma, \pi(A)] = 0 \qquad \text{and} \qquad \{\Gamma, F\}_+ = 0\,. \tag{26.63}$$

F has the form

$$\begin{pmatrix} 0 & F_+ \\ F_- & 0 \end{pmatrix}, \tag{26.64}$$

where F_+ is Fredholm.

Example 26.2.50. Consider $A = \mathbb{C}$ and let $\pi : \mathbb{C} \to \mathsf{End}(\mathcal{H})$ be the unique unital representation. Fredholm modules with this data correspond to (essentially self-adjoint) Fredholm operators on $\mathcal{H}$.

Remark 26.2.51 (Kasparov K-theory). If A is not unital, one should relax the first two relations up to compact operators, i.e. $F = F^*$ and $F^2 = \mathbb{1}_{\mathcal{H}}$ in the Calkin algebra. This gives a clear relation to *Kasparov's K-theory* (see below), where the homology complex $KK_\bullet$ is defined in terms of such generalized Fredholm modules. However, it can be shown that these induce the same KK-classes.

Definition 26.2.52 (Kasparov bimodule). Let A, B be two C^*-algebras and let H be a super-Hilbert (A, B)-bimodule. H is called a Kasparov (A, B)-bimodule if there exists an odd adjointable operator $F \in \mathcal{B}(H)$ satisfying the following properties for all $a \in A$:

1. $(F^2 - \mathbb{1}_H)\pi_A(a)$ is compact,
2. $(F - F^*)\pi_A(a)$ is compact, and
3. $[F, \pi_A(a)]$ is compact.

If A is unital (and the representation respects units), this implies that F is a projection in the Calkin algebra. If F is a projection on the nose, the bimodule is said to be **normalized**.

Definition 26.2.53 (Kasparov K-theory). The set, in fact Abelian group, of homotopy classes of Kasparov bimodules $KK(A, B)$, where a homotopy between two Kasparov (A, B)-bimodules is a Kasparov $(A, C([0, 1], B))$-bimodule that restricts to the given bimodules on the boundaries of the interval.

Property 26.2.54. Every Kasparov bimodule is homotopic to a normalized bimodule. Hence, for Kasparov K-theory, one can restrict to normalized bimodules.

26.3 von Neumann algebras

Definition 26.3.1 (von Neumann algebra). A $*$-subalgebra of a C^*-algebra equal to its double commutant (Definition 2.4.3): $M = M''$.

Definition 26.3.2 (Concrete von Neumann algebra). A weakly closed unital $*$-algebra of bounded operators on some Hilbert space.

[8] Often called **von Neumann's double commutant theorem**.

Theorem 26.3.3 (Double Commutant theorem[8]). *The above definitions are equivalent.*

Property 26.3.4 (Projections). Every von Neumann algebra is generated by its projections.

Property 26.3.5 (Orthomodular lattices). The set of projections in a von Neumann algebra or, equivalently, the set of closed subspaces of a Hilbert space forms a complete orthomodular lattice (Definition 1.7.42).

Given two closed subspaces, the join and meet are constructed as follows:

- Meet: $A \wedge B = A \cap B$, and
- Join: $A \vee B =$ smallest closed subspace containing $A \cup B$.

Definition 26.3.6 (Murray–von Neumann equivalence). Closed subspaces of a Hilbert space are said to be Murray–von Neumann equivalent if they are isomorphic through a partial isometry. In terms of projections, this means that $p \sim q$ if and only if there exists a partial isometry u such that $p = uu^*$ and $q = u^*u$.

Property 26.3.7 (Partial order). The inclusion of subspaces induces a partial order on the set of projections. A projection p is said to be finite if there exists no smaller projection q that is equivalent to p.

26.3.1 Factors

Definition 26.3.8 (Factor). Consider a von Neumann algebra M. A $*$-subalgebra $A \subseteq M$ is called a factor of M if its center $Z(A)$ is given by the scalar multiples of the identity.

Definition 26.3.9 (Type-I factor). A factor that contains a minimal projection.

Property 26.3.10 (Type-I_n factors). Any type I-factor is isomorphic to the algebra of all bounded operators on a Hilbert space. To indicate the dimension $n \in \overline{\mathbb{N}}$ of this Hilbert space (which may be $+\infty$), one sometimes uses the subclassification of type I_n-factors.

Definition 26.3.11 (Powers index). Consider a Hilbert space $\mathcal{H}$ together with its von Neumann algebra of bounded operators $\mathcal{B}(\mathcal{H})$. A unital $*$-endomorphism α is said to have Powers index $n \in \mathbb{N}$ if the space $\alpha(\mathcal{B}(\mathcal{H}))$ is isomorphic to a type I_n-factor.

Definition 26.3.12 (Type-II factor). A factor that contains nonzero finite projections but no minimal ones. If the identity is finite, the factor is sometimes said to be of type II_1, otherwise it is said to be of type II_∞.

Definition 26.3.13 (Type-III factor). A factor that does not contain any nonzero finite projections.

Chapter 27

Clifford Algebra

The main references for this chapter are Choquet-Bruhat and DeWitt-Morette (2000); Choquet-Bruhat et al. (1991); Gallier (2008). One should note that there are various conventions for the different structures that arise in the study of Clifford algebras and their representations. Even the references given here do not agree on the chosen conventions.

In general, all metrics (and quadratic forms) wil be assumed to be nondegenerate. A part of the theory can also be extended to the degenerate case, but this will not be considered here. See Gallier (2008) for more information.

Chapter Contents

27.1 Clifford algebra **524**

27.2 Geometric algebra **526**

27.3 Bott periodicity ♣ **527**

27.4 Pin group **528**

27.4.1 Clifford group 528

27.4.2 Pin and Spin 530

27.1 Clifford algebra

Definition 27.1.1 (Clifford algebra). Consider a unital associative k-algebra A together with a quadratic form $Q : A \to k$. The Clifford algebra over A associated to Q is defined as the free algebra generated by A under the following relation:

$$a^2 = Q(a)1\,, \tag{27.1}$$

where 1 is the unit element in A. This condition implies that the square of a vector is a scalar.

Notation 27.1.2. The Clifford algebra corresponding to A and Q is often denoted by $C\ell(A, Q)$.

Construction 27.1.3. The previous definition can be given an explicit construction. First, construct the tensor algebra of A:

$$T(A) = \bigoplus_{n \in \mathbb{N}} A^{\otimes n} . \tag{27.2}$$

Then, construct a two-sided ideal I of A generated by

$$\{a \otimes a - Q(a)1 \mid a \in A\} \tag{27.3}$$

as defined in Construction 2.4.35. The Clifford algebra $C\ell(A, Q)$ can then be constructed as the quotient algebra $T(A)/I$.

Remark 27.1.4. Looking at Definition 23.4.22, it can be seen that the exterior algebra $\Lambda^\bullet(A)$ coincides with the Clifford algebra $C\ell(A, 0)$. Note that if $Q \neq 0$, the two algebras are still isomorphic as vector spaces (at least if[1] $\mathrm{char}(A) \neq 2$), but not as algebras.

Property 27.1.5 (Dimension). If A has dimension n, then $C\ell(A, Q)$ has dimension 2^n.

Example 27.1.6 (Inner product spaces). The classical example of a Clifford algebra is given by an inner product space with signature (p, q). The Clifford algebra $C\ell_{p,q}(V)$ or $V_{p,q}$ is then defined as the Clifford algebra generated under the relation

$$v^2 = -g(v, v)1 . \tag{27.4}$$

In physics, this convention would correspond to the 'mostly pluses'-convention, which is mainly adopted in *general relativity* (see Chapter 67).

Notation 27.1.7. The Clifford algebra over $\mathbb{R}^{p+q}$ with respect to the standard Hermitian form of signature (p, q) is often denoted by $\mathbb{R}_{p,q}$.

Formula 27.1.8 (Dimensional reduction).

$$\mathbb{R}_{p+1,q+1} \cong \mathbb{R}_{p,q} \otimes M_2(\mathbb{R}) \tag{27.5}$$

Formula 27.1.9.

$$\mathbb{R}_{p+1,q} \cong \mathbb{R}_{q+1,p} \tag{27.6}$$

Formula 27.1.10.

$$\mathbb{R}_{p,q+2} \cong \mathbb{R}_{q,p} \otimes \mathbb{H} \tag{27.7}$$

[1] This condition will often come back in this chapter.

27.2 Geometric algebra

Definition 27.2.1 (Geometric algebra). Let V be a vector space equipped with a symmetric bilinear form $g : V \times V \to k$. The geometric algebra (GA) over V is defined as the Clifford algebra $C\ell(V, g)$. Here, the classic relation $Q(v) := g(v, v)$ is implicitly used since quadratic forms are required in Definition 27.1.1. This identification is unique as long as $\mathsf{char}(V) \neq 2$.

Definition 27.2.2 (Inner and exterior product). Analogous to the inner product in linear algebra and the wedge product in exterior algebra, one can define (a)symmetric products on a geometric algebra. First of all, it should be noted that the product ab of two vectors a and b can be written as the sum of a symmetric and an antisymmetric part:

$$ab = \frac{1}{2}(ab + ba) + \frac{1}{2}(ab - ba)\,. \tag{27.8}$$

One can then define the inner product as the symmetric part:

$$a \cdot b := \frac{1}{2}(ab + ba) = \frac{1}{2}\big((a + b)^2 - a^2 - b^2\big) = g(a, b)\,. \tag{27.9}$$

Analogously, one can define the exterior (outer) product as the antisymmetric part:

$$a \wedge b := \frac{1}{2}(ab - ba)\,. \tag{27.10}$$

These definitions allow to rewrite the geometric product as follows:

$$ab = a \cdot b + a \wedge b\,. \tag{27.11}$$

Remark. Looking at the last equality in the definition of the inner product (27.9), it can be seen that condition (27.1) is indeed satisfied when $a = b$.

Definition 27.2.3 (Multivector). Any element of a geometric algebra over V is called a multivector. The simple multivectors of grade k, i.e. elements of the form $v_1 v_2 \dots v_k$ with $v_i \in V$ for all $i \leq k$, are called **k-blades**. (This should again remind the reader of the content of Section 23.4.4.) Sums of multivectors of different grades are called **mixed** multivectors (even though these elements do not represent a geometric structure). Taking $\mathsf{dim}(V) = n$, multivectors of grade n and $n - 1$ are also called **pseudoscalars** and **pseudovectors**, respectively.

Definition 27.2.4 (Grade projection operator). Let $a \in \mathcal{G}(V)$ be a general multivector and consider $k \in \mathbb{N}$. The grade (projection) operator $\langle \cdot \rangle_k : \mathcal{G}(V) \to \mathcal{G}(V)_k$ is defined as the projection of a onto the grade-k part of a.

Using these projection operators, one can extend the inner and exterior product to the complete GA as follows.

Formula 27.2.5. Let a, b be two multivectors of grades m and n, respectively. Their inner product is given by

$$a \cdot b := \langle ab \rangle_{|m-n|} \tag{27.12}$$

and their exterior product is given by

$$a \wedge b := \langle ab \rangle_{m+n} . \tag{27.13}$$

An explicit calculation for $a \in \mathcal{G}(V)_1, b \in \mathcal{G}(V)_k$ gives

$$\begin{aligned} a \cdot b &= \frac{1}{2}\left(ab - (-1)^k ba\right), \\ a \wedge b &= \frac{1}{2}\left(ab + (-1)^k ba\right). \end{aligned} \tag{27.14}$$

27.3 Bott periodicity ♣

The following theorem has deep implications in *K-theory* (see Chapter 41). It is also (through K-theory) related to the *tenfold way* of *Altland* & *Zirnbauer* in condensed matter physics.

Theorem 27.3.1 (Bott periodicity). *The classification of (real) Clifford algebras is periodic modulo 8:*

$$\mathbb{R}_{p,q+8} \cong \mathbb{R}_{p+8,q} \cong \mathbb{R}_{p,q} \otimes M_{16}(\mathbb{R}) . \tag{27.15}$$

For complex Clifford algebras, a similar statement exists, but with periodicity 2.

Bott periodicity also has some implications in category theory (Chapter 3). Here, the language of Section 4.3 will be used.

Definition 27.3.2 (Banach category). An additive category enriched over Banach spaces.

For every Banach category $\mathbf{C}$ and finite-dimensional $\mathbb{R}$-algebra A, the category of A-representations $\mathbf{C}^A$ is defined as follows:

1. **Objects**: A-representations in $\mathbf{C}$, i.e. pairs (X, ρ) where $X \in \mathsf{ob}(\mathbf{C})$ and $\rho : A \to \mathsf{End}(X)$ is a Banach algebra morphism.

2. **Morphism**: The A-equivariant morphisms/intertwiners.

Notation 27.3.3. For brevity, some specific notation for the cases $A = \mathbb{R}_{p,q}$ and $A = M_n(\mathbb{R})$ are introduced, respectively: $\mathbf{C}^{p,q}$ and $\mathbf{C}(n)$.

Property 27.3.4 (Morita equivalence). Every pseudo-Abelian Banach category $\mathbf{C}$ is equivalent to $\mathbf{C}(n)$ for some $n \in \mathbb{N}$.

Property 27.3.5. For A, B two finite-dimensional $\mathbb{R}$-algebras and $\mathbf{C}$ a pseudo-Abelian Banach category, the following equivalences exist:

$$\begin{aligned} \mathbf{C}^{A\oplus B} &\cong \mathbf{C}^A \times \mathbf{C}^B\,, \\ \mathbf{C}^{A\otimes B} &\cong \left(\mathbf{C}^A\right)^B\,. \end{aligned} \tag{27.16}$$

Property 27.3.6 (Bott periodicity). For $\mathbf{C}$ a pseudo-Abelian, real Banach category, the equivalence classes of categories $\mathbf{C}^{p,q}$ are determined by $p - q \bmod 8$.

Construction 27.3.7 (Grothendieck group). Consider a **Banach functor** $\varphi : \mathbf{C} \to \mathbf{D}$ between Banach categories, i.e. a functor that acts linearly and continuously on hom-spaces. Furthermore, assume that φ is **quasi-surjective**, i.e. every object in $\mathbf{D}$ is a direct summand of an object in the image of φ.

To this functor, one can assign an Abelian group $K(\varphi)$ as follows. Let $\mathscr{V}(\varphi)$ denote the set of triples (X, Y, f) where $X, Y \in \mathsf{ob}(\mathbf{C})$ and $f : \varphi(X) \to \varphi(Y)$ is an isomorphism. Elements in $\mathscr{V}(\varphi)$ are said to be isomorphic if there exist isomorphisms in $\mathbf{C}$ that make the 'obvious' diagram in $\mathbf{D}$ commute. The sum of such triples is defined elementwise. Let $\mathscr{E}(\varphi)$ denote the subset of $\mathscr{V}(\varphi)$ consisting of triples (X, Y, f) where $X = Y$ and f is homotopic to $\mathbb{1}_{\varphi(X)}$ in $\mathsf{Aut}(\varphi(X))$. The group $K(\varphi)$ is defined as the quotient of $\mathscr{V}(\varphi)$ by the following equivalence relation:

$$v \sim v' \iff \exists e, e' \in \mathscr{E}(\varphi) : v + e \cong_{\mathscr{V}} v' + e'\,. \tag{27.17}$$

Definition 27.3.8 (Restriction of scalars). Consider a pseudo-Abelian Banach category $\mathbf{C}$. The group $K^{p,q}(\mathbf{C})$ is defined as the Grothendieck group of the canonical functor $\mathbf{C}^{p+1,q} \to \mathbf{C}^{p,q}$ (this functor is sometimes called the 'restriction of scalars'-functor since it is contravariantly induced by the inclusion $\mathbb{R}_{p,q} \hookrightarrow \mathbb{R}_{p+1,q}$). Bott periodicity implies that these groups only depend on $p - q \bmod 8$.

Example 27.3.9 ($K^{0,0}$). The functor $\mathbf{C}^{1,0} \to \mathbf{C}^{0,0}$ is (up to equivalence) the direct sum functor $\mathbf{C} \times \mathbf{C} \to \mathbf{C} : (X, Y) \mapsto X \oplus Y$.

Remark 27.3.10 (Complex spaces). The above constructions can be generalized to the setting of complex Banach spaces and complex algebras. However, Bott periodicity will then give a $p - q \bmod 2$ classification.

27.4 Pin group

27.4.1 Clifford group

In this section, it is assumed that V is finite-dimensional and Q is nondegenerate.

Definition 27.4.1 (Transposition). Let $\{e_i\}_{i\leq n}$ be a basis for V. On the tensor algebra

$T(V)$, there exists an anti-automorphism v^t that reverses the order of the basis vectors:

$$\cdot^t : e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_k} \mapsto e_{i_k} \otimes \cdots \otimes e_{i_2} \otimes e_{i_1} . \tag{27.18}$$

Because the ideal in the definition of a Clifford algebra is invariant under this map, it induces an anti-automorphism, called the transposition or **reversal**, on $C\ell(V, Q)$.

Definition 27.4.2 (Main involution). Let V_0, V_1 be the grade-0 and grade-1 components of the Clifford algebra $C\ell(V, Q)$, respectively. Consider the following operator:

$$\hat{v} := \begin{cases} v & \text{if } v \in V_0 , \\ -v & \text{if } v \in V_1 . \end{cases} \tag{27.19}$$

This operator can be generalized to all of $C\ell(V, Q)$ by linearity. The resulting operator is called the main involution or **inversion** on $C\ell(V, Q)$. It turns the Clifford algebra into a *superalgebra* (see Example 23.5.6).

Definition 27.4.3 (Twisted conjugation). Let $C\ell(V, Q)$ be a Clifford algebra and consider its group of units $C\ell^*(V, Q)$. Note that this group consists exactly of the nonzero homogeneous elements of $C\ell(V, Q)$. Twisted conjugation by $s \in C\ell^*(V, Q)$ is defined as follows:

$$\chi : C\ell^*(V, Q) \to \mathsf{Aut}(V) \qquad \text{with} \qquad \chi(s)v = sv\hat{s}^{-1} . \tag{27.20}$$

Definition 27.4.4 (Clifford group[2]). The Clifford group $\Gamma(V, Q)$ is defined as follows:

$$\Gamma(V, Q) := \left\{ s \in C\ell^*(V, Q) \,|\, v \in V \implies sv\hat{s}^{-1} \in V \right\} \tag{27.21}$$

$\Gamma(V, Q)$ inherits the group structure from $C\ell^*(V, Q)$.

Property 27.4.5. On $C\ell^*(V, Q) \cap V$, twisted conjugation is given by a Householder transformation (Definition 22.3.16).

Property 27.4.6. If one interprets the condition $\chi(s)v \in V$ as stating the existence of a linear transformation $L_s \in \mathsf{End}(V)$ such that

$$se_i\hat{s}^{-1} = (L_s)_i^j e_j , \tag{27.22}$$

it can be seen that L_s preserves the norm on V and, accordingly, that the map $s \mapsto L_s$ defines a surjective homomorphism[3]

$$\rho : \Gamma(V, Q) \to \mathsf{O}(V, Q) : s \mapsto L_s . \tag{27.23}$$

Being a group morphism to a matrix group acting on V, it defines a representation called the **vector(ial) representation**. Furthermore, from the first isomorphism theorem 2.1.11 it follows that $\mathsf{O}(V, Q)$ is isomorphic to $\Gamma(V, Q)/\ker(\chi)$, where $\ker(\chi) = \mathbb{R}_0$.

[2]Sometimes called the **Lipschitz group**.

[3]In $\mathsf{char}(K) \neq 2$, the surjectivity of the map χ follows from the *Cartan–Dieudonné theorem*. For characteristic 2, one can prove that the surjectivity holds using different methods.

This isomorphism also implies[4] that the Clifford group is given by the set of finite products of invertible elements in V:

$$\Gamma(V,Q) = \left\{ \prod_i^n v_i \,\middle|\, v_i \text{ invertible in } V \wedge n \in \mathbb{N} \right\}. \tag{27.24}$$

Corollary 27.4.7. By noting that pure rotations can be decomposed as an even number of reflections, one finds that

$$\Gamma^+(V,Q)/\mathbb{R}_0 \cong \mathsf{SO}(V,Q), \tag{27.25}$$

where $\Gamma^+ := C\ell_{\text{even}}(V,Q) \cap \Gamma(V,Q)$ is the intersection of the even Clifford algebra and the Clifford group.

Remark 27.4.8. As was noted in the beginning of this chapter, there is a variety of different conventions in use. One of the important distinctions is the definition (or choice) of conjugation map χ. *Atiyah, Bott* and *Shapiro* have introduced the twisted conjugation map that was used for the definition of the Clifford group. Before them, the usual choice was the ordinary conjugation map[5]

$$\mathsf{ad}_s : v \mapsto svs^{-1}. \tag{27.26}$$

Although the difference between these maps seems to be rather subtle, the implications are important. If the ordinary conjugation would have been chosen for the definition of the Clifford group, only a surjective homomorphism would have been found in the case of $\mathsf{dim}(V)$ being odd. Moreover, the action by a degree-1 element would not be given by a Householder transformation anymore. Instead it would be the negative of this operation. This distinction is, in particular, important for the next section.

27.4.2 Pin and Spin

Formula 27.4.9 (Spinor norm). On $\Gamma(V,Q)$, one can define the spinor norm[6]

$$\mathcal{N}(x) : \Gamma(V,Q) \to K^\times : x \mapsto x^t x, \tag{27.27}$$

where x^t is the Clifford transposition (Definition 27.4.1). On V, $\mathcal{N}$ coincides with the norm induced by Q.

Definition 27.4.10 (Pin and spin groups). Using the spinor norm $\mathcal{N}$, one can now define the pin and spins groups as follows:

$$\mathsf{Pin}(V) := \{ s \in \Gamma(V,Q) \mid \mathcal{N}(s) = \pm 1 \} \tag{27.28}$$

[4] Again using the *Cartan–Dieudonné theorem*, valid only when $\mathsf{char}(K) \neq 2$. In fact this statement is more or less the *Cartan–Dieudonneé* theorem in terms of geometric algebra.

[5] The notation ad_s comes from the fact that this map resembles the adjoint action of a group.

[6] This map can be generalized to the full Clifford algebra, but then the image will not just be the underlying field anymore.

and

$$\mathsf{Spin}(V) := \mathsf{Pin}(V) \cap \Gamma^+(V, Q)\,. \tag{27.29}$$

Remark 27.4.11. In the literature, one can sometimes find the following alternative definition of the spinor norm:

$$\mathcal{N}(x) := \hat{x}^t x\,. \tag{27.30}$$

Alternative Definition 27.4.12. The Pin group can also be defined as the set of elements in $\Gamma(V, Q)$ that can be written as a product of unit Clifford vectors (here, unit indicates unit norm and not just invertible as before). The Spin group is then defined as the elements that can be written as the product of an even number of unit Clifford vectors.

Property 27.4.13. The Pin group satisfies the following isomorphism:

$$\mathsf{Pin}(V, Q)/\mathbb{Z}_2 \cong \mathsf{O}(V, Q)\,. \tag{27.31}$$

An analogous relation holds for the Spin group and $\mathsf{SO}(V, Q)$. These relations imply that the Pin and Spin groups form a double cover (Definition 6.2.15) of, respectively, the orthogonal and special orthogonal groups.

Notation 27.4.14. The Pin groups associated to $\mathsf{O}(n, 0)$ and $\mathsf{O}(0, n)$ are often denoted by $\mathsf{Pin}^+(n)$ and $\mathsf{Pin}^-(n)$, where the signs refer to the sign of squares in the Clifford condition (27.1).

Definition 27.4.15 (Spinor). Consider a vector space V equipped with a (faithful) representation of the group $\mathsf{Spin}(m, n)$. This representation is called the **spin(or) representation**. Elements of V are called spinors.

More precisely, if one considers the complex Clifford algebra $C\ell_{m,n}(\mathbb{C})$, two possibilities exist: either $m + n$ is even or $m + n$ is odd. In the even case ($m + n = 2k$), one can prove (using the Artin–Wedderburn theorem 2.4.71) that the algebra is isomorphic to the matrix algebra $M(2^k, \mathbb{C})$. In the odd case ($m + n = 2k + 1$), the algebra is isomorphic to the direct sum $M(2^k, \mathbb{C}) \oplus M(2^k, \mathbb{C})$.

Inside these matrix algebras, one can find a set of elements satisfying the Clifford condition (27.1) and, thereby, generating the Clifford algebra. These are the so-called **gamma matrices**. The real algebra generated by these elements is isomorphic to the real Clifford algebra $C\ell_{m,n}(\mathbb{R})$.[7] The fundamental representation of this real algebra is often called the **Dirac representation**. If $m + n$ is even, the representation decomposes as the direct sum of two irreducible representations called the **Weyl** or **half-spin(or) representations**.

[7] Note, however, that the matrices themselves are, in general, still complex-valued.

Example 27.4.16 (Exceptional isomorphisms). The following table gives some group isomorphisms for the Spin group in low dimensions $n \in \mathbb{N}$:

n	$\mathsf{Spin}(n)$
1	$\mathsf{O}(1)$
2	$\mathsf{U}(1)$
3	$\mathsf{SU}(2)$
4	$\mathsf{SU}(2) \times \mathsf{SU}(2)$.

For quadratic forms of signature (p, q), the following table is found:

$(1, n)$	$\mathsf{Spin}(1, n)$
$(1, 1)$	$\mathsf{GL}(1, \mathbb{R})$
$(1, 2)$	$\mathsf{SL}(2, \mathbb{R})$
$(1, 3)$	$\mathsf{SL}(2, \mathbb{C})$
$(1, 5)$	$\mathsf{SL}(2, \mathbb{H})$
$(1, 9)$	$\mathsf{SL}(2, \mathbb{O})$.[8]

Formula 27.4.17. Consider the basis of $\mathfrak{su}(2)$ given by the *Pauli matrices* (see Definition 57.6.9). An explicit (double) covering map $\rho : \mathsf{Spin}(3) \cong \mathsf{SU}(2) \to \mathsf{SO}(3)$ is given by:

$$\rho : U \mapsto \frac{1}{2} \operatorname{tr}(U \sigma_i U^\dagger \sigma^j) . \tag{27.32}$$

Property 27.4.18. For all $m, n \in \mathbb{N}$, the following isomorphism exists:

$$\mathsf{Spin}(m, n) \cong \mathsf{Spin}(n, m) . \tag{27.33}$$

Remark 27.4.19 (Physical implications). Note that the above isomorphism only holds for the Spin groups and not for the associated Pin groups. This could have consequences in physics! In general, physicists freely switch between a $(1, 3)$- and $(3, 1)$-signature because all particles are assumed to be proper spinors. However, some pinors can only occur for a specific signature and this way it might be possible to detect the signature of the universe (see Berg, DeWitt-Morette, Gwo, and Kramer (2001)).

Definition 27.4.20 (Semispin group). By definition, every Spin group has a canonical $\mathbb{Z}_2$-subgroup. However, for $n \in 4\mathbb{N}_0$, $\mathsf{Spin}(n)$ also contains a noncanonical central $\mathbb{Z}_2$-subgroup $\iota : \mathbb{Z}_2 \hookrightarrow \mathsf{Spin}(n)$. The quotient $\mathsf{Spin}(n)/\iota$ is called the semispin group $\mathsf{SemiSpin}(n)$.

Property 27.4.21. For $n = 8$, the $\mathbb{Z}_2$-subgroups are isomorphic and, as a consequence

$$\mathsf{SemiSpin}(8) \cong \mathsf{SO}(8) . \tag{27.34}$$

[8] This last isomorphism should not exactly be understood in the sense of matrix algebras, since Spin acts associatively, but the octonions do not. Moreover, the dimensions do not even agree as vector spaces. A suitable definition of $\mathsf{SL}(2, \mathbb{O})$ falls outside the scope of this compendium (see e.g. a paper by *Hitchin*.)

Chapter 28

Noncommutative Algebra

References for this chapter are Beggs and Majid (2020); Sontz (2015b).

Chapter Contents

28.1 Coalgebras **533**

28.2 Hopf algebras **535**

28.2.1 Drinfel'd double 538

28.3 Quantum groups **539**

28.4 Differential structures **541**

28.4.1 Differential calculi 541

28.4.2 Covariance 542

28.1 Coalgebras

The definition of a (unital associative) algebra can be dualized (in the categorical sense) in the following way.

Definition 28.1.1 (Coalgebra). A vector space C together with two linear maps $\Delta : C \to C \otimes C$ and $\varepsilon : C \to \mathfrak{K}$, called the **comultiplication** and **counit**, that satisfy the following two axioms:

1. $(\mathbb{1} \otimes \Delta) \circ \Delta = (\Delta \otimes \mathbb{1}) \circ \Delta$, and
2. $(\mathbb{1} \otimes \varepsilon) \circ \Delta = (\varepsilon \otimes \mathbb{1}) \circ \Delta = \mathbb{1}$.

Example 28.1.2. The simplest example is given by the vector space V with basis $\{e_i\}_{i \in I}$, where the comultiplication and counit are defined as follows:

$$\Delta(e_i) := e_i \otimes e_i \tag{28.1}$$

and

$$\varepsilon(e_i) := 1\,. \tag{28.2}$$

By linearity, these maps can be extended to all of V. Important examples are the tensor algebra and exterior algebra over a vector space (Definitions 23.3.7 and 23.4.21).

Remark 28.1.3. This example shows that every algebra admits a coalgebra structure. However, this does not mean that every algebra admits the structure of a *bialgebra* (see Definition 28.2.1 below).

Definition 28.1.4 (Grouplike element). An element c in a coalgebra (C, Δ, ε) that satisfies $\Delta(c) = c \otimes c$ and $\varepsilon(c) = 1$.

Remark. The name 'grouplike' stems from the fact that the coalgebra structure on the group algebra $k[G]$ is obtained by defining $\Delta(g) = g \otimes g$ for all $g \in G$.

Definition 28.1.5 (Unital coalgebra). A coalgebra (C, Δ, ε) is said to be unital if it comes equipped with a coalgebra morphism $\eta : \mathfrak{K} \to C$. The element $\eta(1)$ is often also denoted by 1.

Definition 28.1.6 (Primitive element). An element c in a unital coalgebra (C, Δ, ε) that satisfies $\Delta(c) = c \otimes 1 + 1 \otimes c$.

Definition 28.1.7 (Frobenius algebra). A tuple $(A, \nabla, \eta, \Delta, \varepsilon)$ that satisfies the following conditions:

1. (A, ∇, η) is an associative algebra.
2. (A, Δ, ε) is a coassociative coalgebra.
3. The **Frobenius law** holds:

$$(\mathbb{1}_A \otimes \nabla) \circ (\Delta \otimes \mathbb{1}_A) = \Delta \circ \nabla = (\nabla \otimes \mathbb{1}_A) \circ (\mathbb{1}_A \otimes \Delta)\,. \tag{28.3}$$

If $\nabla \circ \Delta = \mathbb{1}_A$, the algebra is said to be **special**.

Remark 28.1.8. It is this definition that can be generalized to arbitrary symmetric monoidal categories.

Notation 28.1.9 (Sweedler notation). Let (C, Δ) be a coalgebra. For any element $c \in C$, the comultiplication $\Delta(c)$ is an element of $C \otimes C$ and can thus be written in the following form

$$\Delta(c) = \sum_{i \in I} a_i \otimes b_i$$

for some finite index set I. For lengthy calculations with a lot of different symbols, this notation gets tedious and, hence, the following shorthand[1] is introduced:

$$\Delta(c) = \sum_{(c)} c_{(1)} \otimes c_{(2)} \tag{28.4}$$

[1] Sometimes, the notation $\Delta(c) = \Delta_1(c) \otimes \Delta_2(c)$ is used.

or even

$$\Delta(c) = c_{(1)} \otimes c_{(2)} \,. \tag{28.5}$$

As an example, the coassociativity condition $(\Delta \otimes \mathbb{1}) \circ \Delta = (\mathbb{1} \otimes \Delta) \circ \Delta$ is rewritten:

$$c_{(1)} \otimes c_{(2)} \otimes c_{(3)} = \sum_{(c)} c_{(1)(1)} \otimes c_{(1)(2)} \otimes c_{(2)} = \sum_{(c)} c_{(1)} \otimes c_{(2)(1)} \otimes c_{(2)(2)} \,.$$

Analogously, the counit law becomes

$$c = c_{(1)}\varepsilon(c_{(2)}) = \varepsilon(c_{(1)})c_{(2)}$$

and, hence, one can freely move the counit ε around.

By dualizing the definition of ideals in a ring, one obtains the following notion.

Definition 28.1.10 (Coideal). Let (C, Δ, ε) be a coalgebra. A subcoalgebra I of C is a (left) coideal if

$$\Delta(I) \subseteq C \otimes I \,, \tag{28.6}$$

i.e. it is a comodule with respect to the comultiplication.

28.2 Hopf algebras

Definition 28.2.1 (Bialgebra). Let A be a vector space over a field $\mathfrak{K}$. Suppose that the triple (A, ∇, η) defines a unital associative algebra and that the triple (A, Δ, ε) defines a counital coassociative coalgebra. The quintuple $(A, \nabla, \eta, \Delta, \varepsilon)$ defines a bialgebra if ∇ and Δ satisfy the commutative diagrams in Fig. 28.1. These diagrams state that ∇, η are coalgebra morphisms and Δ, ε are algebra morphisms.

Definition 28.2.2 (Convolution). Let $(A, \nabla, \eta, \Delta, \varepsilon)$ be a bialgebra. The convolution of two morphisms $f, g : A \to A$ is defined as follows:

$$f * g := \nabla \circ (f \otimes g) \circ \Delta \,. \tag{28.7}$$

When equipped with the convolution as multiplication, the space of morphisms on a bialgebra also becomes an algebra. In fact, this extends to morphisms between two algebras, turning $\mathbf{Alg}_{\mathfrak{K}}(A, A')$ into a $\mathfrak{K}$-algebra itself.

Definition 28.2.3 (Hopf algebra). A bialgebra $(A, \nabla, \eta, \Delta, \varepsilon)$ equipped with a linear map $S : A \to A$ that satisfies

$$\nabla \circ (\mathbb{1}_A \otimes S) \circ \Delta = \nabla \circ (S \otimes \mathbb{1}_A) \circ \Delta = \eta \circ \varepsilon \tag{28.8}$$

or, using the convolution on A,

$$\mathbb{1}_A * S = S * \mathbb{1}_A = \eta \circ \varepsilon \,. \tag{28.9}$$

The map S is called the **antipode** or **coinverse**.

$$\begin{tikzcd}
A \otimes A \arrow[r, "\nabla"] \arrow[d, "\Delta \otimes \Delta"'] & A \arrow[r, "\Delta"] & A \otimes A \arrow[d, "\nabla \otimes \nabla"] \\
A \otimes A \otimes A \otimes A \arrow[rr, "\mathbb{1}_A \otimes \sigma_{A,A} \otimes \mathbb{1}_A"'] & & A \otimes A \otimes A \otimes A
\end{tikzcd}$$

$$\begin{tikzcd}
& \mathfrak{K} \arrow[dl, "\eta"'] \arrow[dr, "\eta \otimes \eta"] & \\
A \arrow[rr, "\Delta"'] & & A \otimes A
\end{tikzcd}
\qquad
\begin{tikzcd}
A \otimes A \arrow[rr, "\nabla"] \arrow[dr, "\varepsilon \otimes \varepsilon"'] & & A \arrow[dl, "\eta"] \\
& \mathfrak{K} &
\end{tikzcd}
\qquad
\begin{tikzcd}
& A \arrow[dr, "\varepsilon"] & \\
\mathfrak{K} \arrow[ur, "\eta"] \arrow[rr, "\mathbb{1}_{\mathfrak{K}}"'] & & \mathfrak{K}
\end{tikzcd}$$

Figure 28.1: Bialgebra conditions.

Remark 28.2.4. Some authors require the antipode to be invertible. (In finite dimensions, this is always the case as noted below.)

Property 28.2.5. Given a Hopf algebra structure on a bialgebra, the antipode S is an antihomomorphism. Furthermore, by noting that it is the inverse of the identity under convolutions, one can show that the antipode is unique (if it exists). Being a Hopf algebra is thus a property, not a structure.

Example 28.2.6 (Finite-dimensional bialgebras). Any finite-dimensional bialgebra admits an invertible antipode and, in particular, is a Hopf algebra.

Example 28.2.7 (Connected graded bialgebras). A connected graded bialgebra $H_\bullet$, i.e. a graded bialgebra with $H_0 = \mathfrak{K}$, where $\mathfrak{K}$ is the base field, admits a canonical choice of antipode.

Property 28.2.8 (Group structure). The algebra $\mathbf{Alg}(H, H')$ of algebra morphisms between two Hopf algebras is a group under the convolution product in Definition 28.2.2. The inverse of a morphism ϕ is given by $\phi \circ S_H = S_{H'} \circ \phi$.

Definition 28.2.9 (Quasitriangular Hopf algebra[2]). A Hopf algebra H for which there exists an invertible element $R \in H \otimes H$ that satisfies:

1. $R\Delta(a) = \sigma\big(\Delta(x)\big)R$,
2. $(\Delta \otimes \mathbb{1})(R) = R_{13}R_{23}$, and
3. $(\mathbb{1} \otimes \Delta)(R) = R_{13}R_{12}$,

where $\sigma(x \otimes y) = y \otimes x$ is the braiding on H and where $R_{ij} \in H \otimes H \otimes H$ is defined using the components of R in the ith and j^{th} position and the unit element $1 \in H$ in the other position, i.e. $(a \otimes b)_{13} = a \otimes 1 \otimes b$.

[2]Sometimes called a **braided Hopf algebra**.

The element R is often called the **universal R-matrix**[3]. Any Hopf algebra admitting such an element is said to be **quasicocommutative**.

Example 28.2.10 (Universal enveloping algebra). Let $\mathfrak{g}$ be a Lie algebra and consider its universal enveloping algebra $U(\mathfrak{g})$ from Section 32.4.6. This algebra admits a Hopf algebra structure with the following operations:

$$\begin{aligned} \Delta(e_i) &:= e_i \otimes 1 + 1 \otimes e_i\,, \\ \varepsilon(e_i) &:= 0\,, \\ S(e_i) &:= -e_i\,. \end{aligned} \tag{28.10}$$

It becomes quasitriangular when equipped with the trivial R-matrix $\mathbb{1}_{\mathfrak{g}} \otimes \mathbb{1}_{\mathfrak{g}}$.

Remark 28.2.11 (Tensor product of modules). One could ask where bialgebras and especially Hopf algebras naturally arise. Consider an algebra A together with its category of modules $A\mathbf{Mod}$. Now one would like to define a monoidal structure on $A\mathbf{Mod}$ induced by the tensor product on A. However, this monoidal structure should be compatible with the action of A.

The naive (left) action

$$A \otimes (M \otimes_A N) \to M \otimes_A N : a \otimes m \otimes n \mapsto (am) \otimes n$$

does not admit a suitable tensor unit due to its asymmetric definition. To obtain the correct definition, one could be inspired by group representations: $g \cdot (m \otimes n) = gm \otimes gn$. In this case one has the diagonal map $\Delta : G \to G \times G$ that can be used to act on both sides of the tensor product. One could then ask "Why not just define the action of an algebra in the same way?", i.e.

$$a \otimes (m \otimes n) \mapsto (am) \otimes (an)\,.$$

However, because the action is required to be linear (after all it should be compatible with the algebra morphisms), this definition is not valid. To resolve this issue, the existence of an additional algebra morphism $A \to A \otimes A$ is required with which one can construct a suitable action as follows:

$$A \otimes (M \otimes_A N) \to M \otimes_A N : a \otimes m \otimes n \mapsto (a_{(1)}m) \otimes (a_{(2)}n)\,. \tag{28.11}$$

Together with the usual conditions of an algebra action, one obtains exactly the requirement that A should be a bialgebra. So if A is a bialgebra, then $A\mathbf{Mod}$ will be a monoidal category (this is in fact an equivalence known as **Tannaka duality**).

Now, one could require some more structure on $A\mathbf{Mod}$, for example that it admits duals. Consider an A-module V together with its dual $V^* \cong \mathrm{Hom}(V, \mathbb{C})$. Given a linear

[3] This name is, in general, used for all elements $R \in H \otimes H$ satisfying the first condition above.

map $S : A \to A$ one could define a general action as follows:

$$(af)(v) := f\big(S(a)v\big)\,. \tag{28.12}$$

The requirement that this is indeed an action leads to the requirement $S(ab) = S(b)S(a)$ on S, which is equivalent to requiring that S is an algebra antihomomorphism. Together with the other compatibility conditions, such as that the evaluation and coevaluation maps induced by the underlying vector spaces are also A-module morphisms, one is led to the requirement that A is a Hopf algebra. Hence, if A is a Hopf algebra (with an invertible antipode), then $A\mathbf{Mod}$ will be a rigid monoidal category.

One could even go further and require the representation category to be braided. This requirement then exactly leads to the Hopf algebra being quasitriangular (this also explains why these Hopf algebras are sometimes said to be braided).

28.2.1 Drinfel'd double

One can easily generalize Definition 2.2.51 (when written in terms of group algebras) to the case of bialgebras by replacing the group comultiplication by a general comultiplication.

Definition 28.2.12 (Bicrossed product of bialgebras). Two bialgebras A, B are said to form a **matched pair** (of bialgebras) if there exist actions $- \cdot - : A \otimes B \to B$ and $-^{-} : A \otimes B \to B$, compatible with the coalgebra structures, that satisfy the following equations:

1. $a \cdot (bc) = (a_{(1)} \cdot b_{(1)})(a_{(2)}^{b_{(2)}} \cdot c)$,
2. $a \cdot 1 = \varepsilon(a)1$,
3. $(ab)^c = a^{b_{(1)} \cdot c_{(1)}} b_{(2)}^{c_{(2)}}$,
4. $1^b = \varepsilon(b)1$, and
5. $a_{(1)}^{b_{(1)}} \otimes a_{(2)} \cdot b_{(2)} = a_{(2)}^{b_{(2)}} \otimes a_{(1)} \cdot b_{(1)}$.

Given such a matched pair, one can define a Hopf algebra structure on $A \otimes B$ defined by the following operations:

- **Product**: $(a \otimes b)(c \otimes d) = a(b_{(1)} \cdot c_{(1)}) \otimes b_{(2)}^{c_{(2)}} d$,
- **Coproduct**: $\Delta(a \otimes b) = (a_{(1)} \otimes b_{(1)}) \otimes (a_{(2)} \otimes b_{(2)})$, and
- **Counit**: $\varepsilon(a \otimes b) = \varepsilon_A(a)\varepsilon_B(b)$.

If the bialgebras are equipped with antipodes, the bicrossed product admits an induced

antipode:

$$S(a \otimes b) = S_B(b_{(2)}) \cdot S_A(a_{(2)}) \otimes S_B(b_{(1)})^{S_A(a_{(1)})} . \tag{28.13}$$

Example 28.2.13 (Tensor product). In the case where the bialgebra actions are given by left multiplication with the counits, the bicrossed product is isomorphic to the tensor product.

Construction 28.2.14 (Drinfel'd double[4]). Consider a Hopf algebra H (with invertible antipode). It can be shown that H and $(H^{\mathrm{op}})^*$ form a matched pair of bialgebras. The left and right actions are induced by pullback:

$$\begin{aligned} (a \cdot f)(b) &= f\big(S^{-1}(a_{(2)})ba_{(1)}\big) , \\ a^f &= f\big(S^{-1}(a_{(3)})a_{(1)}\big)a_{(2)} . \end{aligned} \tag{28.14}$$

The resulting bicrossed product $(H^{\mathrm{op}})^* \bowtie H$ is called the Drinfel'd double $D(H)$.

Example 28.2.15 (Drinfel'd double for groups). Consider a finite group G together with its associated group algebra $\mathbb{C}[G]$. On this algebra, one can put a Hopf algebra structure as follows:

$$\begin{aligned} \Delta(g) &= g \otimes g , \\ \varepsilon(g) &= 1 . \end{aligned} \tag{28.15}$$

On the other hand, one can also put a Hopf algebra structure on the dual $\mathbb{C}[G]^*$:

$$\begin{aligned} \Delta(P_g) &= \sum_{hh'=g} P_h \otimes P_{h'} , \\ \varepsilon(P_g) &= \delta_{g,e} , \end{aligned} \tag{28.16}$$

where the basis for $\mathbb{C}[G]^*$ is given by the 'projections' $P_g : h \mapsto \delta_{g,h}$. Antipodes for both algebras are given by $S(g) = g^{-1}$ and $S(P_g) = P_{g^{-1}}$.

@@ COMPLETE @@

28.3 Quantum groups

This section heavily builds upon the theory presented in Chapter 32. The content is partially based on talks by *Henriques*.

Construction 28.3.1 (Jimbo–Drinfeld). Consider a Lie algebra $\mathfrak{g}$ together with its universal enveloping algebra $U(\mathfrak{g})$ constructed using the *Chevalley–Serre relations* (see Construction 32.4.57):

1. $[H_i, H_j] = 0$,

[4] Also known as the **quantum double** (especially in physics).

2. $[H_i, E_j] = a_{ij}E_j$,
3. $[H_i, F_j] = -a_{ij}F_j$,
4. $[E_i, F_j] = \delta_{ij}H_j$,
5. $\mathrm{ad}_{E_i}^{|a_{ij}|-1}(E_j) = 0$, and
6. $\mathrm{ad}_{F_i}^{|a_{ij}|-1}(F_j) = 0$.

To obtain the quantum group $U_q(\mathfrak{g})$, which is also called a **deformation** or **quantization** of $U(\mathfrak{g})$, one replaces the generators H_i by the following generators[5]:

$$K_i := q^{d_i H_i}\,, \tag{28.17}$$

where

$$d_i := \frac{\langle \alpha_i, \alpha_i \rangle}{2} \tag{28.18}$$

is related to the norm of the ith simple root. So, instead of the H_i being functionals on the root lattice, one gets functions from the root lattice to the Laurent polynomials in q, i.e. to $\mathbb{C}[q, q^{-1}]$.

From this functional point of view one can rewrite the second and third relation as follows:

$$\begin{aligned} f \cdot E_i &= E_i \tau_{\alpha_i}(f)\,, \\ f \cdot F_i &= F_i \tau_{-\alpha_i}(f)\,, \end{aligned}$$

where f is a polynomial in the H_i's and $\tau_{\alpha_i}(f)(\lambda) := f(\lambda + \alpha_i)$. Replacing H_i by K_i, one obtains the following relations:

2*. $K_i E_j = q^{d_i a_{ij}} E_j K_i$, and

3*. $K_i F_j = q^{-d_i a_{ij}} F_j K_i$.

The three relations between the E_i's and the F_i's are deformed using q-analog numbers. First, define the q-numbers[6]:

$$[n]_q := \frac{q^n - q^{-n}}{q - q^{-1}}\,. \tag{28.19}$$

Using this definition, Serre relation 4 becomes

4*. $[E_i, F_j] = \delta_{ij}[H_i]_{q^{d_i}} = \delta_{ij}\frac{K_i - K_i^{-1}}{q^{d_i} - q^{-d_i}}$,

[5]To be complete, one should also add generators K_i^{-1} that act formally as inverses of the generators K_i.

[6]Note that q-numbers are often defined differently. This definition is equal to $\frac{1}{q^{n-1}}[n]_{q^2}$ when rewritten using the common definition.

where the factor $[H_i]_{q^{d_i}}$ should be interpreted as first evaluating H_i on a root and then taking the q-analog. The adjoint action relations (5 and 6) on E_i and F_i can be rewritten by replacing binomial coefficients by their q-analogs ($i \neq j$):

5*. $\sum_{k=1}^{1-a_{ij}} (-1)^k \begin{bmatrix} 1-a_{ij} \\ k \end{bmatrix}_{q^{d_i}} E_i^{1-a_{ij}-k} E_j E_i^k = 0$, and

6*. $\sum_{k=1}^{1-a_{ij}} (-1)^k \begin{bmatrix} 1-a_{ij} \\ k \end{bmatrix}_{q^{d_i}} F_i^{1-a_{ij}-k} F_j F_i^k = 0$.

28.4 Differential structures

28.4.1 Differential calculi

Definition 28.4.1 (First-order differential calculus). Let A be an associative algebra and let Γ be an A-bimodule. Together with an A-bimodule morphism $\mathsf{d} : A \to \Gamma$ this structure is called a first-order differential calculus (FODC) if it satisfies the following two conditions:

1. **Leibniz rule**: $\mathsf{d}(ab) = (\mathsf{d}a)b + a(\mathsf{d}b)$.
2. **Standard form**: Every element $g \in \Gamma$ can be written as

$$g = \sum_{i=1}^{n} a_i(\mathsf{d}b_i)$$

for some $n \in \mathbb{N}$ and (not necessarily unique) elements $\{a_i, b_i\}_{i \leq n}$.

If $\ker(\mathsf{d}) \cong \mathfrak{K}$, where $\mathfrak{K}$ is the underlying field, the calculus is said to be **connected**. A calculus is said to be **inner** if the differential d acts through a commutator, i.e. there exists an element $\theta \in A$ such that $\mathsf{d}a = [\theta, a]$ for all $a \in A$.

An algebra morphism $\phi : A \to B$ is said to be **differentiable**, with respect to a choice of FODCs on A and B, if there exists an A-bimodule morphism $\phi_* : \Gamma_A \to \Gamma_B$ (A inherits an action on Γ_B through ϕ) such that $\mathsf{d} \circ \phi = \phi_* \circ \mathsf{d}$. If such a morphism exists, it is given by $\phi_*(a\mathsf{d}b) = \phi(a)\mathsf{d}\phi(b)$.

Remark 28.4.2. The second condition can be rewritten in terms of a right action using the Leibniz rule.

Definition 28.4.3 (Cotangent dimension). If Γ is free over A with a basis of cardinality $n \in \mathbb{N}$, it is said to be **parallelized with cotangent dimension** $\dim(A) - 1$.

This is in analogy with the case of function algebras $C^\infty(M)$ and the first de Rham space $\Omega^1(M)$ on a smooth manifold. If $\Omega^1(M)$ is free over $C^\infty(M)$, this implies that there exists a global basis of one-forms or, equivalently, a global frame of the tangent bundle. This is

the same as saying that M is parallelizable. (The cotangent dimension will be explained later.)

Example 28.4.4 (Universal FODC). Consider an algebra A with multiplication μ. The bimodule $\Omega_{\text{uni}} := \ker(\mu)$, equipped with the operator $\mathrm{d}a := 1 \otimes a - a \otimes 1$, defines a first-order differential calculus on A. Furthermore, every FODC over A can be obtained as the quotient $\Omega_{\text{uni}}/\mathcal{N}$ for a subbimodule $\mathcal{N}$. The universal FODC is inner if and only if there exists a central element $F \in A \otimes A$ such that $\mu(F) = 1$.

Example 28.4.5 (Kähler differentials). In commutative algebra one has an equivalent construction. There one takes the module $\Omega^1(A)$ to be the quotient I/I^2, where I is the again the augmentation ideal $I := \ker(\mu)$. The quotient enforces that the bimodule is symmetric and, hence, that the Leibniz rule becomes

$$\mathrm{d}(ab) = a\mathrm{d}b + b\mathrm{d}a\,. \tag{28.20}$$

The module of Kähler differentials plays the same role as the universal FODC in commutative algebra in that it is universal with respect to modules over commutative algebras equipped with a derivation.

Definition 28.4.6 (Differential calculus). Consider an algebra A. A differential calculus over A is an A-bimodule with the structure of differential graded algebra $(\Gamma^\bullet, \mathrm{d})$ such that $\Gamma^0 = A$. If $\Gamma^\bullet$ is freely generated by A and $\mathrm{d}A$, then $(\Gamma^\bullet, \mathrm{d})$ is often called an **exterior algebra** over A.

Property 28.4.7 (Prolongation). Every FODC (Γ, d) admits a maximal extension (or **prolongation**) to an exterior algebra. In degree $k \in \mathbb{N}$, it is obtained by taking

$$\Gamma^k := \underbrace{\Gamma \otimes_A \cdots \otimes_A \Gamma}_{k \text{ times}}, \tag{28.21}$$

and modding out the ideal generated by the following relations:

$$\left\{ \sum_i da_i \otimes \mathrm{d}b_i + \mathrm{d}u_i \otimes \mathrm{d}v_i = 0 \,\middle|\, \sum_i a_i \mathrm{d}b_i = \sum_i (\mathrm{d}u_i)v_i \right\}. \tag{28.22}$$

If one starts from the universal FODC over an algebra, the resulting differential calculus is again universal in the sense that all other differential calculi can be obtained by taking quotients by differential ideals.

28.4.2 Covariance

Definition 28.4.8 (Covariant FODC). An FODC (Γ, d) over a Hopf algebra H is said to be left-covariant if it is a left H-comodule such that

$$\Phi(a\mathrm{d}b) = \Delta(a)(\mathbb{1}_H \otimes \mathrm{d})\Delta(b)\,. \tag{28.23}$$

Part VI

(Differential) Geometry

Table of Contents

29 Elementary Geometry **551**
29.1 Straight lines . . . 551
29.2 Trigonometry . . . 551

30 Curves & Surfaces **554**
30.1 Curves . . . 554
30.1.1 Arc length . . . 555
30.1.2 Frenet–Serret frame . . . 555
30.2 Surfaces . . . 556
30.2.1 First fundamental form . . . 557
30.2.2 Isometries . . . 558
30.2.3 Second fundamental form . . . 558
30.2.4 Curvature of a surface . . . 559
30.2.5 Christoffel symbols and geodesics . . . 562
30.2.6 Theorema Egregium . . . 562

31 Manifolds **564**
31.1 Charts . . . 564
31.2 Tangent vectors . . . 567
31.3 Submanifolds . . . 569
31.3.1 Immersions and submersions . . . 569
31.3.2 Submanifolds . . . 570
31.4 Manifolds with boundary . . . 572
31.4.1 Cobordism theory ♣ . . . 572
31.5 Surgery theory ♣ . . . 573

32 Lie groups and Lie algebras **574**
32.1 Lie groups . . . 575
32.1.1 Left invariant vector fields . . . 575
32.1.2 One-parameter subgroups . . . 576
32.2 Lie algebras . . . 576
32.2.1 Definitions . . . 576

32.2.2 Examples . . . 579
32.2.3 Exponential map . . . 580
32.2.4 Solvable Lie algebras . . . 581
32.2.5 Simple Lie algebras . . . 582
32.2.6 Central extensions . . . 583
32.3 Representation theory . . . 583
32.3.1 Lie groups . . . 583
32.3.2 Lie algebras . . . 584
32.3.3 Coadjoint orbits . . . 584
32.4 Structure . . . 585
32.4.1 Killing form . . . 585
32.4.2 Weights, roots and Dynkin diagrams . . . 586
32.4.3 Highest weight theory . . . 592
32.4.4 Coxeter groups . . . 593
32.4.5 Kac–Moody algebras ♣ . . . 594
32.4.6 Universal enveloping algebra . . . 596
32.5 Group contractions . . . 598
32.6 Lie algebra cohomology ♣ . . . 598

33 Fibre Bundles **602**
33.1 Bundles . . . 602
33.2 Constructions . . . 605
33.3 Sections . . . 606

34 Vector Bundles **607**
34.1 Tangent bundle . . . 608
34.2 Vector bundles . . . 610
34.2.1 Sections . . . 611
34.2.2 Tubular neighbourhoods . . . 611
34.2.3 Sums and product . . . 611
34.2.4 Associated vector bundles . . . 612
34.3 Vector fields . . . 613
34.3.1 Integral curves . . . 615
34.3.2 Lie derivative . . . 616
34.3.3 Multivector fields . . . 618
34.4 Differential k-forms . . . 619
34.4.1 Definition . . . 619
34.4.2 Exterior derivative . . . 620
34.4.3 Lie derivative . . . 621
34.4.4 Interior product . . . 622
34.4.5 Lie derivative of tensor fields . . . 622

34.4.6 Vector-valued differential forms . . . 623
34.5 Grassmann bundle . . . 625
34.6 Frobenius' theorem . . . 627
34.7 Linear connections . . . 629
34.7.1 Koszul connections . . . 629
34.7.2 Induced connections . . . 630
34.7.3 Curvature & torsion . . . 632
34.8 Integration Theory . . . 633
34.8.1 Orientation and densities . . . 633
34.8.2 Orientation in homology . . . 635
34.8.3 Integration . . . 636
34.8.4 Stokes' theorem . . . 637
34.9 Cohomology . . . 638
34.9.1 de Rham complex . . . 638
34.9.2 Integration . . . 640
34.9.3 Cohomology with compact support . . . 641
34.9.4 Čech–de Rham complex . . . 645
34.9.5 Nonorientable manifolds . . . 646
34.9.6 Generalized cohomology ♣ . . . 648
34.10 Morse theory . . . 648
34.10.1 Morse functions . . . 648
34.10.2 Morse–Bott functions . . . 649
34.10.3 Morse homology . . . 649
34.11 Differential operators . . . 651
34.11.1 Elliptic complexes . . . 651

35 Principal Bundles **652**
35.1 Principal bundles . . . 653
35.1.1 Associated bundles . . . 654
35.1.2 Sections . . . 656
35.2 Universal bundle . . . 657
35.3 Connections . . . 658
35.3.1 Vertical vectors . . . 658
35.3.2 Ehresmann connections . . . 660
35.3.3 Connection forms . . . 661
35.3.4 Local representations . . . 663
35.3.5 Parallel transport . . . 664
35.3.6 Holonomy group . . . 667
35.4 Covariant derivatives . . . 668
35.4.1 Koszul connections . . . 668

35.4.2 Exterior covariant derivative . . . 669
35.4.3 Curvature . . . 671
35.4.4 Torsion . . . 672
35.5 Reduction of the structure group . . . 673
35.6 Characteristic classes . . . 675
35.6.1 Chern–Weil theory . . . 676
35.6.2 Complex bundles . . . 677
35.6.3 Real bundles . . . 679
35.6.4 Lie group cohomology . . . 680
35.6.5 Chern–Simons forms . . . 680
35.7 Differential cohomology ♣ . . . 681
35.7.1 Differential characters . . . 681
35.7.2 Combining singular and de Rham cohomology . . . 682
35.7.3 Deligne cohomology . . . 683
35.8 Cartan geometry . . . 685
35.8.1 Klein geometry . . . 685
35.8.2 Cartan geometry . . . 686
35.9 $\mathcal{D}$-geometry ♣ . . . 688

36 Riemannian Geometry 689
36.1 Riemannian manifolds . . . 689
36.1.1 Metric . . . 689
36.1.2 Riemannian manifolds . . . 692
36.1.3 Levi-Civita connection . . . 692
36.1.4 Killing vectors . . . 694
36.2 Curvature . . . 695
36.3 Spinor bundles . . . 696
36.3.1 Spin structures . . . 696
36.3.2 Stiefel–Whitney classes . . . 698
36.3.3 Dirac operators ♣ . . . 699
36.4 Conformal structures . . . 702
36.5 Distributions ♣ . . . 704

37 Symplectic Geometry 705
37.1 Symplectic manifolds . . . 706
37.2 Poisson manifolds . . . 709
37.3 Lagrangians . . . 711
37.3.1 Lagrangian submanifolds . . . 711
37.3.2 Maslov classes . . . 713
37.3.3 Polarizations . . . 716
37.4 Hamiltonian dynamics . . . 717

37.4.1 Dynamical systems . . . 717
37.4.2 Hamilton–Jacobi equation . . . 719
37.4.3 Integrability . . . 720
37.5 Symplectic reduction . . . 723
37.5.1 Hamiltonian actions . . . 723
37.5.2 Symplectic reduction . . . 725
37.5.3 Poisson reduction . . . 725
37.5.4 Lie–Poisson reduction . . . 726
37.5.5 Infinite-dimensional manifolds . . . 727
37.6 Metaplectic structures . . . 727

38 Contact Geometry **729**
38.1 Contact structures . . . 729
38.1.1 Contact form . . . 729
38.1.2 Reeb vector fields . . . 730

39 Complex Geometry **731**
39.1 Complex structures . . . 731
39.2 Complex differential forms . . . 735
39.3 Holomorphic vector bundles . . . 736
39.4 Kähler manifolds . . . 737
39.4.1 Killing vectors . . . 740
39.4.2 Dirac operators . . . 741
39.5 Complex curves . . . 741
39.5.1 Riemann surfaces and orbifolds . . . 741
39.5.2 Pseudoholomorphic maps . . . 746
39.6 Cohomology . . . 748
39.6.1 Dolbeault cohomology . . . 748
39.6.2 Lagrangian Floer homology ♣ . . . 748

40 Calculus of Variations **753**
40.1 Introduction . . . 754
40.2 Jet bundles . . . 756
40.2.1 Topology . . . 756
40.2.2 Contact structure . . . 758
40.2.3 Differential operators . . . 758
40.3 Variational bicomplex ♣ . . . 759
40.3.1 Differential structure . . . 759
40.3.2 Euler operators . . . 763
40.3.3 Functional complex . . . 765
40.3.4 Structure of the bicomplex . . . 766

40.3.5 Variational problems . . . 770
40.3.6 Inverse problem . . . 772
40.3.7 Finite jet bundles . . . 773
40.4 Partial differential equations . . . 774
40.4.1 Algebraic formulation . . . 774
40.4.2 Symmetries . . . 776
40.4.3 Pseudogroups ♣ . . . 777

41 K-theory ♣ 778
41.1 Preliminaries . . . 779
41.2 Topological K-theory . . . 779
41.2.1 Introduction . . . 779
41.2.2 Classification . . . 780
41.2.3 Negative degree . . . 782
41.2.4 Bott periodicity . . . 783
41.2.5 Clifford modules . . . 785
41.2.6 Cohomology theory . . . 786
41.2.7 Applications . . . 786
41.3 Twisted K-theory . . . 788
41.4 Differential K-theory . . . 788
41.5 Algebraic K-theory . . . 789
41.5.1 Determinant . . . 789
41.6 Operator K-theory . . . 790

42 Noncommutative Geometry 792
42.1 Noncommutative measure theory . . . 792
42.2 Quantum geometry . . . 795
42.3 Spectral geometry . . . 796
42.4 Dunkl operators . . . 800

43 Synthetic Differential Geometry ♣ 801
43.1 Neighbourhoods . . . 801
43.2 Affine connections . . . 802
43.3 Euclidean geometry . . . 802
43.3.1 Infinitesimal elements . . . 802
43.3.2 Calculus . . . 803

44 Higher-dimensional Geometry ♣ 805
44.1 Graded manifolds . . . 806
44.2 Infinite-dimensional geometry . . . 808
44.3 Cohesion . . . 809

44.4 Smooth spaces 813
44.4.1 Frölicher spaces 814
44.4.2 Smooth sets 815
44.4.3 Smooth algebras 819
44.4.4 Supergeometry 819
44.5 Higher Lie theory 824
44.6 Gauge theory ♣ 824
44.6.1 Lie groupoids 825
44.7 Weak groups 826
44.7.1 Spaces 828
44.8 2-Bundles 829
44.9 Space and quantity 831

Chapter 29

Elementary Geometry

Chapter Contents

29.1 Straight lines 551

29.2 Trigonometry 551

29.1 Straight lines

Theorem 29.1.1 (Thales).

29.2 Trigonometry

Figure 29.1: General triangle.

Definition 29.2.1 (Trigonometric functions). Consider a triangle as in Fig. 29.1 with $\gamma = \frac{\pi}{2}$, i.e. a right-angled triangle.

$$\sin(\alpha) := \frac{a}{c} \tag{29.1}$$

$$\cos(\alpha) := \frac{b}{c} \tag{29.2}$$

$$\tan(\alpha) := \frac{a}{b} \tag{29.3}$$

The longest side, c in this case, is called the **hypotenuse**.

Corollary 29.2.2. Since the adjacent side for one angle is the @@ overstaande?? @@ for the other, the formulas above give rise to the following relations:

$$\sin(\alpha) = \cos(\beta)\,, \tag{29.4}$$

$$\cos(\alpha) = \sin(\beta)\,, \tag{29.5}$$

$$\tan(\alpha) = \frac{1}{\tan(\beta)}\,. \tag{29.6}$$

Notation 29.2.3 (Cotangent). Because it occurs relatively often, the expression in the last line above received a distinct notation:

$$\cot(\alpha) := \frac{1}{\tan(\alpha)}\,. \tag{29.7}$$

Formula 29.2.4.

$$\tan(\theta) = \frac{\sin(\theta)}{\cos(\theta)} \tag{29.8}$$

for any angle $\theta \in [0, 2\pi[$.

Theorem 29.2.5 (Pythagoras). *In any right-angled triangle, where c denotes the hypotenuse, the following relation between the sides holds:*

$$a^2 + b^2 = c^2\,.$$

Using the definitions of the trigonometric functions, this can also be rewritten.

Theorem 29.2.6 (Fundamental theorem of trigonometry).

$$\sin^2(\theta) + \cos^2(\theta) = 1 \tag{29.9}$$

for any angle $\theta \in [0, 2\pi[$.

The Pythagorean theorem can be extended as follows.

Formula 29.2.7 (Cosine law).

$$a^2 + b^2 - 2ab\cos(\gamma) = c^2 \tag{29.10}$$

Formula 29.2.8 (Sine law).

$$\frac{\sin(\alpha)}{a} = \frac{\sin(\beta)}{b} = \frac{\sin(\gamma)}{c} \tag{29.11}$$

Formula 29.2.9 (Tangent law).

$$\frac{a-b}{a+b} = \frac{\tan\left(\frac{1}{2}(\alpha-\beta)\right)}{\tan\left(\frac{1}{2}(\alpha+\beta)\right)} \tag{29.12}$$

Formula 29.2.10 (Addition formulas).

$$\sin(\alpha + \beta) = \sin(\alpha)\cos(\beta) + \cos(\alpha)\sin(\beta) \tag{29.13}$$

$$\cos(a + b) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta) \tag{29.14}$$

$$\tan(\alpha + \beta) = \frac{\tan(\alpha) + \tan(\beta)}{1 + \tan(\alpha)\tan(\beta)} \tag{29.15}$$

Besides the addition formulas, there are also product formulas. (Only one is given, since the others can be derived from it.)

Corollary 29.2.11 (Werner–Simpson).

$$2\sin(\alpha)\cos(\beta) = \sin(\alpha - \beta) + \sin(\alpha + \beta) \tag{29.16}$$

Lesser known trigonometric functions are the cotangent-like versions of the sine and cosine.

Definition 29.2.12 (Secant).

$$\sec(\theta) := \frac{1}{\sin(\theta)} \tag{29.17}$$

$$\csc(\theta) := \frac{1}{\cos(\theta)} \tag{29.18}$$

Chapter 30

Curves & Surfaces

Chapter Contents

30.1 Curves **554**

30.1.1 Arc length 555

30.1.2 Frenet–Serret frame 555

30.2 Surfaces **556**

30.2.1 First fundamental form 557

30.2.2 Isometries 558

30.2.3 Second fundamental form 558

30.2.4 Curvature of a surface 559

30.2.5 Christoffel symbols and geodesics 562

30.2.6 Theorema Egregium 562

30.1 Curves

Definition 30.1.1 (Regular curve). Let $c : I \to \mathbb{R}^n$ be a C^1-curve. It is said to be regular at $t \in I$ if $\frac{\mathrm{d}c}{\mathrm{d}t} \neq 0$. (This will be generalized in Corollary 31.3.5.)

Definition 30.1.2 (Geometric property). A property that is invariant under:

1. parameter transformations, and
2. orientation-preserving basis transformations, i.e. transformations in $\mathsf{SO}(n)$.

30.1.1 Arc length

Definition 30.1.3 (Natural parameter). Let $c : I \to \mathbb{R}^n$ be a C^1-curve. The parametrization is said to be natural if

$$\left\| \frac{\mathrm{d}c}{\mathrm{d}t} \right\| = 1 \tag{30.1}$$

for all $t \in I$.

Formula 30.1.4 (Arc length). The following function is a bijective map and gives a natural parameter for the curve c:

$$\phi(t) := \int_{t_0}^{t} \|\dot{c}(t)\| \, dt \,. \tag{30.2}$$

This natural parameter is often denoted by the letter s.

Property 30.1.5. Let c be a curve and let u be any parametrization of c. It is a natural parameter if and only if there exists a constant α such that

$$u = \pm s + \alpha \,, \tag{30.3}$$

where s denotes the arc length.

Remark 30.1.6. This property implies that there does not exist a unique natural parameter.

30.1.2 Frenet–Serret frame

Definition 30.1.7 (Tangent vector). Let $c(s)$ be a curve parametrized by arc length. The tangent vector field $t(s)$ is defined as follows:

$$t(s) := c'(s) \,. \tag{30.4}$$

A different notation for the derivative, e.g. a dot, is used whenever the differentiation is performed with respect to a natural parameter.

Property 30.1.8. From the definition of natural parameters (Definition 30.1.3) and the previous definition it follows that the tangent vector is automatically a unit vector.

Definition 30.1.9 (Principal normal vector). Let $c(s)$ be a curve parametrized by arc length. The principal normal vector field is defined as follows:

$$n(s) := \frac{t'(s)}{\|t'(s)\|} \,. \tag{30.5}$$

Property 30.1.10. From Property 30.1.8 and the definition of the principal normal vector, it follows that the tangent vector and principal normal vector are always orthogonal.

Definition 30.1.11 (Binormal vector). Let $c(s)$ be a curve parametrized by arc length. The binormal vector field is defined as follows:

$$b(s) := t(s) \times n(s)\,. \tag{30.6}$$

Definition 30.1.12 (Frenet–Serret frame). Because the vectors $t(s), n(s)$ and $b(s)$ are mutually orthonormal and linearly independent, they can be used to construct an orthonormal basis. The ordered basis $\{t(s), n(s), b(s)\}$ is called the Frenet–Serret frame.

Remark. This basis does not have to be the same for every value of the parameter s.

Definition 30.1.13 (Curvature). Let $c(s)$ be a curve parametrized by arc length. The curvature of $c(s)$ is defined as follows:

$$\rho(s) := \frac{1}{\|t'(s)\|}\,. \tag{30.7}$$

Definition 30.1.14 (Torsion). Let $c(s)$ be a curve parametrized by arc length. The torsion of $c(s)$ is defined as follows:

$$\tau(s) := \rho(s)^2\,(t\ t'\ t'')\,, \tag{30.8}$$

where $(a\ b\ c)$ denotes the **triple product** $a \cdot (b \times c)$.

Formula 30.1.15 (Frenet formulas). Because the Frenet–Serret vectors form a basis, the derivatives of these vector fields can be written as a linear combination of the vectors themselves:

$$\begin{cases} t'(s) &= & \frac{1}{\rho(s)} n(s)\,, \\ n'(s) &= -\frac{1}{\rho(s)} t(s) & + \quad \tau(s) b(s)\,, \\ b'(s) &= & -\tau(s) n(s)\,. \end{cases} \tag{30.9}$$

Theorem 30.1.16 (Fundamental theorem for curves). *Let $k, w : I \to \mathbb{R}$ be two C^1-functions with k nonnegative. There exists an interval $]-\varepsilon, \varepsilon[\, \subset I$ and a curve $c :]-\varepsilon, \varepsilon[\, \to \mathbb{R}^3$ with natural parameter s such that c has k as its curvature and w as its torsion.*

30.2 Surfaces

Notation 30.2.1. Let $\sigma(q^1, q^2)$ be the parametrization of a surface Σ.[1] The derivative of σ with respect to the coordinate q^i is written as follows:

$$\sigma_i := \frac{\partial \sigma}{\partial q^i}\,. \tag{30.10}$$

[1] The symbol σ denotes the embedding of the surface as a vector field while Σ denotes the geometric image of σ.

Definition 30.2.2 (Tangent plane). Let $p(q_0^1, q_0^2)$ be a point on the surface Σ. The tangent space $T_p\Sigma$ to Σ at p is defined as follows:

$$T_p\Sigma := \left\{ v \in \mathbb{R}^3 \,\middle|\, \left[v - \sigma(q_0^1, q_0^2)\right] \cdot \left[\sigma_1(q_0^1, q_0^2) \times \sigma_2(q_0^1, q_0^2)\right] = 0 \right\}. \tag{30.11}$$

Definition 30.2.3 (Normal vector). The cross product in the preceding definition is closely related to the normal vector to Σ at the point p. The normal vector at the point p is defined as

$$N := \frac{\sigma_1 \times \sigma_2}{\|\sigma_1 \times \sigma_2\|}. \tag{30.12}$$

This way, the tangent plane $T_p\Sigma$ can be seen to consist exactly of the vectors that are orthogonal to the normal vector N.

30.2.1 First fundamental form

Definition 30.2.4 (Metric coefficients). Let σ be the parametrization of a surface. The metric coefficients g_{ij} are defined as follows:

$$g_{ij} := \sigma_i \cdot \sigma_j. \tag{30.13}$$

Definition 30.2.5 (Scale factor). The following factors are often used in vector calculus:

$$g_{ii} =: h_i^2. \tag{30.14}$$

Definition 30.2.6 (First fundamental form). Let σ be the parametrization of a surface. One can define a bilinear form $\mathsf{I}_p : T_p\Sigma \times T_p\Sigma \to \mathbb{R}$ that restricts the inner product on $\mathbb{R}^n$ to $T_p\Sigma$:

$$\mathsf{I}_p(v, w) := v \cdot w. \tag{30.15}$$

This bilinear form is called the first fundamental form or **metric**.

Corollary 30.2.7. All vectors $v, w \in T_p\Sigma$ are linear combinations of the tangent vectors σ_1, σ_2. This allows to relate the first fundamental form and the metric coefficients (Definition 30.2.4):

$$\mathsf{I}_p(v, w) = v^i\sigma_i \cdot w^j\sigma_j = g_{ij}v^iw^j. \tag{30.16}$$

Notation 30.2.8. The (arc) length (Formula 30.1.4) of a curve $c : I \to \mathbb{R}^n$ can be written as follows:

$$s = \int_I \|\dot{c}(t)\|\, dt \equiv \int \sqrt{\mathrm{d}s^2}, \tag{30.17}$$

where the second equality is formally defined. The two equalities together can be combined into the following notation for the metric (which is often used in physics):

$$\mathrm{d}s^2 \equiv g_{ij}\mathrm{d}q^i\mathrm{d}q^j. \tag{30.18}$$

Formula 30.2.9 (Inverse metric). Let (g_{ij}) denote the metric tensor. The matrix (g^{ij}) is defined as its inverse:

$$(g^{ij}) := \frac{1}{\det(g_{ij})}\begin{pmatrix} g_{22} & -g_{12} \\ -g_{12} & g_{11} \end{pmatrix}. \tag{30.19}$$

30.2.2 Isometries

Definition 30.2.10 (Isometry). A smooth distance-preserving function, i.e. a smooth function $\Phi : \Sigma \to \Sigma'$ that maps arc segments in Σ to arc segments with the same length in Σ'. This function is usually assumed to be diffeomorphic.

Property 30.2.11. A diffeomorphism Φ is an isometry if and only if the metric coefficients of σ and σ' are the same.

Definition 30.2.12 (Conformal map). A diffeomorphism $\Phi : \Sigma \to \Sigma'$ is said to be conformal or **isogonal** if it maps two intersecting curves in Σ to intersecting curves in Σ' with the same intersection angle.

Property 30.2.13. A diffeomorphism Φ is conformal if and only if the metric coefficients of σ and σ' are proportional.

Definition 30.2.14 (Area-preserving map). A diffeomorphism $\Phi : \Sigma \to \Sigma'$ that maps a subset of Σ to a subset of Σ' with the same area.

Property 30.2.15. A diffeomorphism Φ is area-preserving if and only if the metric coefficients of σ and σ' satisfy

$$g_{11}g_{22} - g_{12}^2 = g'_{11}g'_{22} - (g'_{12})^2 \tag{30.20}$$

for all points (q^1, q^2), i.e. if it preserves the determinant of the metric.

Corollary 30.2.16. A map that is area-preserving and conformal is also isometric.

30.2.3 Second fundamental form

Definition 30.2.17 (Second fundamental form). Let σ be the parametrization of a surface. The second fundamental form is the bilinear form $\mathrm{II}_p : T_p\Sigma \times T_p\Sigma \to \mathbb{R}$ defined as follows:

$$\mathrm{II}_p(v, w) := L_{ij}(p)v^i w^j\,, \tag{30.21}$$

where $L_{ij} := N \cdot \sigma_{ij}$.

Definition 30.2.18 (Normal curvature). Let c be a curve embedded as

$$c(s) := \sigma\big(q^1(s), q^2(s)\big)\,.$$

The normal curvature of c at the point $\big(q^1(s), q^2(s)\big)$ is defined as

$$\frac{1}{\rho_n(s)} := c''(s) \cdot N(s)\,. \tag{30.22}$$

From the definition of the second fundamental form, it follows that the normal curvature can be written as follows:

$$\frac{1}{\rho_n(s)} = \mathrm{II}(t,t) = \frac{\mathrm{II}(\dot{c}(\lambda), \dot{c}(\lambda))}{\mathrm{I}(\dot{c}(\lambda), \dot{c}(\lambda))}\,, \tag{30.23}$$

where the last equality holds for any given parameter λ.

Theorem 30.2.19 (Meusnier). *Let c, d be two curves defined on a surface σ. The curves have the same normal curvature at the point $\big(q^1(t_0), q^2(t_0)\big)$ if the following two conditions are satisfied:*

- $c(t_0) = d(t_0)$, *and*
- $\dot{c}(t_0) \parallel \dot{d}(t_0)$.

Furthermore, the osculating circles of all curves with the same normal curvature at a given point form a sphere.

Property 30.2.20. The normal curvature of a **normal section**, i.e. the intersection of the surface with a normal plane, at a given point is equal to the curvature of the section at that point.

Definition 30.2.21 (Geodesic curvature). Let c be a curve embedded as

$$c(s) := \sigma\big(q^1(s), q^2(s)\big)\,.$$

The geodesic curvature of c at the point $\big(q^1(s), q^2(s)\big)$ is defined as follows:

$$\frac{1}{\rho_g(s)} := (N(s)\; t(s)\; t'(s))\,. \tag{30.24}$$

Formula 30.2.22. Let c be a curve defined on a surface σ. From the definitions of the normal and geodesic curvature it follows that

$$\frac{1}{\rho^2} = \frac{1}{\rho_n^2} + \frac{1}{\rho_g^2}\,. \tag{30.25}$$

30.2.4 Curvature of a surface

Definition 30.2.23 (Weingarten map[2]). Let p be a point on the surface Σ. The Weingarten map $L_p : T_p\Sigma \to T_p\Sigma$ is the linear map defined as follows:

$$L_p(\sigma_1) := -N_1 \qquad \text{and} \qquad L_p(\sigma_2) := -N_2\,. \tag{30.26}$$

[2]Sometimes called the **shape operator**.

Formula 30.2.24. Let $v, w \in T_p\Sigma$. The following equalities relate the second fundamental form and the Weingarten map:

$$L_p(v) \cdot w = L_p(w) \cdot v = \mathrm{II}_p(v, w)\,. \tag{30.27}$$

Formula 30.2.25. Let $\left(g^{ij}\right)$ be the inverse of the metric. The matrix elements of L_p can be expressed as follows:

$$L_j^k = g^{ki} L_{ij}\,. \tag{30.28}$$

Formula 30.2.26 (Weingarten formulas).

$$N_j = -L_j^k \sigma_k \tag{30.29}$$

Definition 30.2.27 (Principal curvatures). The eigenvalues of the Weingarten map are called the principal curvatures of the surface and they are denoted by

$$\frac{1}{R_1} \quad \text{and} \quad \frac{1}{R_2}.$$

Let h_1, h_2 denote the eigenvectors of L_p. By the formulas above, the principal curvatures are given by $\mathrm{II}_p(h_i, h_i)$ and they are the extreme values of the normal curvature. The associated tangent vectors are called the **principal directions**. Furthermore, they form a basis for the tangent plane.

Property 30.2.28. If the principal curvatures at a point are not equal, the principal directions are orthogonal. If they are equal, the point is said to be an **umbilical point** or **umbilic**.

Definition 30.2.29 (Line of curvature). A curve is said to be a line of curvature if the tangent vector at every point is a principal direction of the surface at that point.

Formula 30.2.30 (Rodrigues's formula). A curve is a line of curvature if and only if it is a solution of the following differential equation:

$$\frac{\mathrm{d}N}{\mathrm{d}t}(t) = -\frac{1}{R(t)} \frac{\mathrm{d}c}{\mathrm{d}t}(t)\,. \tag{30.30}$$

If the curve satisfies this formula, the scalar function $1/R(t)$ coincides with the principal curvature along the curve.

Formula 30.2.31 (Differential equation for curvature lines).

$$\begin{vmatrix} (\dot{q}^2)^2 & -\dot{q}^1\dot{q}^2 & (\dot{q}^1)^2 \\ g_{11} & g_{12} & g_{22} \\ L_{11} & L_{12} & L_{22} \end{vmatrix} = 0 \tag{30.31}$$

Property 30.2.32. From Definition 30.2.27, it follows that the principal directions are determined by orthogonal vectors. This implies that on a surface containing no umbilics, the curvature lines form an orthogonal web.

Definition 30.2.33 (Gaussian curvature). The Gaussian curvature of a surface is defined as the determinant of the Weingarten map:

$$K := \frac{1}{R_1 R_2}\,. \tag{30.32}$$

Definition 30.2.34 (Mean curvature). The mean curvature of a surface is defined as the trace of the Weingarten map:

$$H := \frac{1}{2}\left(\frac{1}{R_1} + \frac{1}{R_2}\right). \tag{30.33}$$

Property 30.2.35. The principal curvatures are the solutions of the following equation:

$$x^2 - 2Hx + K = 0\,. \tag{30.34}$$

This is the characteristic equation (22.64) of the Weingarten map.

Definition 30.2.36. Let p be a point on the surface Σ.

- p is said to be **elliptic** if $K > 0$ in p.
- p is said to be **hyperbolic** if $K < 0$ in p.
- p is said to be **parabolic** if $K = 0$ and $\frac{1}{R_1}$ or $\frac{1}{R_2} \neq 0$ in p.
- p is said to be **flat** if $\frac{1}{R_1} = \frac{1}{R_2} = 0$ in p.
- p is said to be **umbilical** if $\frac{1}{R_1} = \frac{1}{R_2}$ in p.

Property 30.2.37. A surface Σ containing only umbilics is either part of a sphere or part of a plane.

Formula 30.2.38. In the neighbourhood of a point of a surface with principal curvatures $1/R_1$ and $1/R_2$, the surface is locally given by the quadric

$$x_3 = \frac{1}{2}\left(\frac{x_1^2}{R_1} + \frac{x_2^2}{R_2}\right) \tag{30.35}$$

up to order $O(x^2)$.

Formula 30.2.39 (Euler's formula). Let h_1, h_2 be the eigenvectors of the Weingarten map. The normal curvature of a couple (p, v) where $v = h_1 \cos\theta + h_2 \sin\theta \in T_p\Sigma$ is given by

$$\frac{1}{\rho_n} = \frac{\cos^2\theta}{R_1} + \frac{\sin^2\theta}{R_2}\,. \tag{30.36}$$

Definition 30.2.40 (Asymptotic curve). A curve that is at every point tangent to a direction with zero normal curvature.

Formula 30.2.41 (Differential equation for asymptotic curves).

$$L_{11}\left(\dot{q}^1(t)\right)^2 + 2L_{12}\dot{q}^1(t)\dot{q}^2(t) + L_{22}\left(\dot{q}^2(t)\right)^2 = 0 \tag{30.37}$$

Property 30.2.42. A curve on a surface is an asymptotic curve if and only if the tangent plane and the *osculation plane* coincide at every point on the curve.

30.2.5 Christoffel symbols and geodesics

Formula 30.2.43 (Gauss' formulas). The second derivatives of the surface σ are given by

$$\sigma_{ij} = L_{ij}N + \Gamma^k{}_{ij}\sigma_k\,, \tag{30.38}$$

where the **Christoffel symbols** $\Gamma^k{}_{ij}$ are defined as

$$\Gamma^k{}_{ij} := g^{kl}\sigma_l \cdot \sigma_{ij}\,. \tag{30.39}$$

Corollary 30.2.44. From the expression of the Christoffel symbols one can derive an alternative expression only in terms of the metric g_{ij}:

$$\Gamma^k{}_{ij} = \frac{1}{2}g^{kl}\left(\frac{\partial g_{il}}{\partial q^j} - \frac{\partial g_{ij}}{\partial q^l} + \frac{\partial g_{jl}}{\partial q^i}\right). \tag{30.40}$$

Definition 30.2.45 (Geodesic). A curve with zero geodesic curvature.

Property 30.2.46. A curve on a surface is a geodesic if and only if the tangent plane and the *osculation plane* are orthogonal at every point of the surface.

Formula 30.2.47 (Differential equation for geodesic). If the curve is parametrized by arc length, it is a geodesic if the functions $q^1(s)$ and $q^2(s)$ satisfy the following differential equation:

$$q''^k + \Gamma^k{}_{ij}q'^i q'^j = 0\,. \tag{30.41}$$

30.2.6 Theorema Egregium

Formula 30.2.48 (Codazzi–Mainardi equations).

$$\frac{\partial L_{ij}}{\partial q^k} - \frac{\partial L_{ik}}{\partial q^j} = \Gamma^l{}_{ik}L_{lj} - \Gamma^l{}_{ij}L_{lk} \tag{30.42}$$

Definition 30.2.49 (Riemann curvature tensor).

$$R^l{}_{ijk} := \frac{\partial \Gamma^l{}_{ik}}{\partial q^j} - \frac{\partial \Gamma^l{}_{ij}}{\partial q^k} + \Gamma^s{}_{ik}\Gamma^l{}_{sj} - \Gamma^s{}_{ij}\Gamma^l{}_{ks} \tag{30.43}$$

Formula 30.2.50 (Gauss' equations).

$$R^l{}_{ijk} = L_{ik}L^l_j - L_{ij}L^l_k \tag{30.44}$$

Theorem 30.2.51 (Theorema Egregium). *The Gaussian curvature is completely determined by the metric tensor g_{ij} and its derivatives:*

$$K = \frac{R^l{}_{121}g_{l2}}{g_{11}g_{22} - g_{12}^2}\,. \tag{30.45}$$

Remark. This theorem is remarkable due to the fact that the coefficients L_{ij}, which appear in the general formula of the Gaussian curvature, cannot be expressed in terms of the metric.

Property 30.2.52. From the condition of isometries (Property 30.2.11) and the previous theorem it follows that if two surfaces are connected by an isometric map, the corresponding points have the same Gaussian curvature.

Corollary 30.2.53. There exists no isometric projection from the sphere to the plane. This also implies that a perfect (i.e. isometric) map of the Earth cannot be created.

Chapter 31

Manifolds

References for this chapter (and Part VI in general) are Choquet-Bruhat and DeWitt-Morette (2000); Choquet-Bruhat et al. (1991); Kolar, Michor, and Slovak (1999); Nash and Sen (2011); Rudolph and Schmidt (2017); Schuller (2016).

Chapter Contents

31.1 Charts **564**

31.2 Tangent vectors **567**

31.3 Submanifolds **569**

31.3.1 Immersions and submersions 569

31.3.2 Submanifolds 570

31.4 Manifolds with boundary **572**

31.4.1 Cobordism theory ♣ 572

31.5 Surgery theory ♣ **573**

31.1 Charts

Definition 31.1.1 (Chart). Consider a topological space M and consider an open subset of $U \subseteq M$ such that there exists a homeomorphism onto an open subset of $\mathbb{R}^n$. The pair (U, φ) is called a chart on M.

Definition 31.1.2 (Transition map). Let (U_1, φ_1) and (U_2, φ_2) be two charts on M. The mapping $\varphi_1^{-1} \circ \varphi_2$, defined on the intersection $U_1 \cap U_2$, is called the transition map between the charts.

If $\varphi_1^{-1} \circ \varphi_2$ is continuous, the charts are said to be C^0-compatible. However, because the composition of any two continuous functions is also continuous, every two charts on a topological space are automatically C^0-compatible.

Definition 31.1.3 (Atlas). Let M be a topological space and let $\{(U_i, \varphi_i)\}_i$ be a collection of pairwise compatible charts covering M. This collection of charts is called an atlas on M. From the above remark on C^0-compatibility, it follows that every atlas is a C^0-atlas. By requiring the transition functions to satisfy additional conditions, other types of atlases can be defined (e.g. smooth atlases).

Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be two atlases on the same topological space. If $\mathcal{A}_1 \cup \mathcal{A}_2$ is again an atlas, the atlases are said to be **equivalent** or **compatible**. A maximal union of compatible atlases is called a **maximal atlas**.

Definition 31.1.4 (Topological manifold). A topological space equipped with a maximal C^0-atlas. An alternative definition (often used in topology) is that of a locally Euclidean Hausdorff space. Here, the topology is generated by the collection of charts.

Remark. In the literature, second-countability is often added to the definition of a topological manifold. This ensures that the space has (amongst others) the property of paracompactness (Definition 6.5.7) and, hence, lends itself to the construction of partitions of unity (which are, for example, necessary for the introduction of integration theory as in Section 34.8).

Definition 31.1.5 (Smooth manifold[1]). If all transition maps are C^k-diffeomorphisms, the manifold is called a C^k-manifold and, if a manifold admits a C^k-structure for some $k \in \mathbb{N}$, it said to be **differentiable**. The limiting case, a C^∞-manifold, is also called a **smooth manifold**. If the transition maps are not only smooth, but even analytic (Definition 16.5.8), the manifold is called an **analytic** or C^ω-manifold. A topological manifold equipped with a maximal atlas for which the transition maps are piecewise linear is called a **PL manifold**.

Theorem 31.1.6 (Whitney). *Every C^k-atlas on a paracompact space contains a C^∞-atlas. Furthermore, two C^k-atlases are equal if and only if they contain the same C^∞-atlas. It follows that every differentiable manifold is also smooth.*

Theorem 31.1.7 (Radó–Moise). *In dimensions* $1, 2$ *and* 3*, every topological manifold admits a unique smooth structure.*

Theorem 31.1.8. *For dimensions higher than* 4*, there exist only finitely many distinct smooth structures on compact manifolds. In fact, for PL manifolds, the number of smooth structures is fixed for each dimension (except for* 4*).*

Remark 31.1.9. In dimension 4, there only partial results exist. Noncompact manifolds admit uncountably many distinct smooth structures, while for compact manifolds, no complete characterization has been found.

Definition 31.1.10 (Smooth function). A function $f : M \to N$ between two smooth manifolds such that, for all charts (U, φ) and (V, ψ) on M and N with $f(U) \subseteq V$, the

[1] For an alternative definition of manifolds in the context of *smooth spaces*, see Section 44.4.

function

$$f_{\varphi\psi} = \psi \circ f \circ \varphi^{-1} \tag{31.1}$$

between Cartesian spaces is smooth. This function is called a **local representation** of f.

Definition 31.1.11 (Diffeomorphism). A homeomorphism f such that both f and f^{-1} are smooth.

Notation 31.1.12. The set of all C^∞-functions between two smooth manifolds M, N is denoted by $C^\infty(M, N)$, and analogous for C^k-functions. This set forms a commutative algebra when equipped with the usual sum and product (composition) of functions. When $N = \mathbb{R}$, the notation is shortened to $C^\infty(M)$.

Remark 31.1.13. Depending on the choice of chart, one can define other types of functions in the same way, e.g. C^k-functions or piecewise linear functions.

Definition 31.1.14 (Structure sheaf ♣). Let M be a C^k-manifold. The structure sheaf $\mathcal{O}_M$ is defined as the sheaf (Definition 8.2.1) that assigns to every open set $U \subseteq M$ the algebra $C^k(U)$. The set of smooth functions defined on a neighbourhood of $p \in M$, i.e. the stalk (or **germ**) at p, is denoted by $C^k_p(M)$.

Generally, one can define for all $j \leq k$ the sheaf $\mathcal{O}^j_M$ as the sheaf that assigns to every open set $U \subseteq M$ the set of C^j-functions $f : U \to \mathbb{R}$.

From the 'sheafy' point of view, one can equivalently define a smooth manifold as a locally ringed space (Definition 8.6.2) that is locally isomorphic to $\mathbb{R}^n$ equipped with its standard algebra of differentiable functions. (This is an extension of the algebro-geometric constructions from Chapter 10.)

Property 31.1.15. Two C^k-manifolds are isomorphic if and only if their associated structure sheaves are isomorphic. Moreover, if the manifolds are second-countable and paracompact, they are isomorphic if their function algebras are isomorphic as rings, i.e. there exists an embedding

$$\mathbf{Diff} \hookrightarrow \mathbf{CAlg}^{\mathrm{op}}_{\mathbb{R}} : M \to C^\infty(M)\,. \tag{31.2}$$

A smooth manifold can, up to isomorphism, be completely reconstructed from this algebraic data:

$$M \cong \mathrm{Hom}\big(C^\infty(M), \mathbb{R}\big) \setminus \{0\}\,. \tag{31.3}$$

This can be seen as an analogue of the Gel'fand–Naimark theorem 26.2.42. However, no compactness is required here.

Definition 31.1.16 (Differentiably good cover). A good cover (Definition 7.1.8) for which the intersections are diffeomorphic to $\mathbb{R}^n$ for some $n \in \mathbb{N}$.

If a manifold admits a finite (differentiably) good cover, it is said to be of **finite type**.

Property 31.1.17 (Existence). Every paracompact smooth manifold admits a (differentiably) good cover. Furthermore, if the manifold is compact, it admits a finite good cover.

31.2 Tangent vectors

Definition 31.2.1 (Tangent vector). Let M be a smooth manifold and consider a point $p \in M$. A tangent vector to M at p is a differential operator on the germ of smooth functions at p, i.e. a map $v_p : C_p^\infty(M) \to \mathbb{R}$ satisfying

1. **Linearity**: $v_p(\lambda f + g) = \lambda v_p(f) + v_p(g)$, and
2. **Leibniz property**: $v_p(fg) = f(p)v_p(g) + g(p)v_p(f)$

for all $f, g \in C_p^\infty(M)$ and $\lambda \in \mathbb{R}$. Maps with these properties are also called **derivations**. (More generally, every operation that satisfies the Leibniz property is called a derivation.)

Property 31.2.2 (Kernel). Constant functions lie in the kernel of all tangent vectors.

Definition 31.2.3 (Tangent space). The set of all tangent vectors at a point $p \in M$ admits the structure of a vector space T_pM. On a local chart (U, φ) around $p \in U$ with local coordinates $(x^1, \dots, x^n)$, a canonical choice of basis vectors is given by

$$\left.\frac{\partial}{\partial x^i}\right|_p : C_p^\infty(M) \to \mathbb{R} : f \mapsto \frac{\partial}{\partial x^i}\left(f \circ \varphi^{-1}\right)(\varphi(p)) \,. \tag{31.4}$$

These basis vectors are also often denoted by ∂_i. This is called the coordinate-induced basis.

Due to the explicit dependence of the tangent vectors on the point $p \in M$, it is clear that, for general manifolds, the tangent spaces belonging to different points are not the same. However, they are related through the following property.

Property 31.2.4. For a connected smooth manifold, the tangent spaces satisfy

$$\dim(T_pM) = \dim(M) \tag{31.5}$$

for all $p \in M$. Corollary 22.2.15 then implies that the tangent spaces at any two points $p, q \in M$ are isomorphic. (An explicit choice of isomorphism will be presented in Section 34.7 and Section 35.4.)

Alternative Definition 31.2.5 (Tangent space). Let (U, φ) be a chart around the point $p \in M$. Two smooth curves γ_1, γ_2 through $p \in M$ are said to be tangent at p if their local representations are tangent at 0:

$$\frac{\mathrm{d}(\varphi \circ \gamma_1)}{\mathrm{d}t}(0) = \frac{\mathrm{d}(\varphi \circ \gamma_2)}{\mathrm{d}t}(0) \,. \tag{31.6}$$

Because the transition functions (and their Jacobian matrices) are invertible, this defines an equivalence relation on the set of smooth curves through p. The tangent space at p is then defined as the set of equivalence classes. These classes can be explicitly constructed as follows. The tangent vector to the curve $c(t)$ is defined by the following formula:

$$v_p(f) := \left.\frac{\mathrm{d}(f \circ c)}{\mathrm{d}t}\right|_{t=0} . \tag{31.7}$$

Applying the chain rule gives

$$v_p(f) = \frac{\partial(f \circ \varphi^{-1})}{\partial x^i}(\varphi(p))\frac{\mathrm{d}x^i}{\mathrm{d}t}(0) , \tag{31.8}$$

where $x^i := (\varphi \circ c)^i$. The first factor depends only on the point p, while the second factor is equal for all tangent curves through p. It is thus clear that curves satisfying Eq. (31.6) define the same tangent vector.

Proof (Proof of equivalence). Let (U, φ) be a chart around the point $p \in M$. Using the first definition of a tangent vector 31.2.3, i.e.

$$\left.\frac{\partial}{\partial q^i}\right|_p : C_p^\infty(M, \mathbb{R}) \to \mathbb{R} : f \mapsto \frac{\partial}{\partial q^i}\left(f \circ \varphi^{-1}\right)(\varphi(p)) ,$$

one can rewrite Eq. (31.8)

$$v_p(f) = \frac{\partial(f \circ \varphi^{-1})}{\partial q^i}(\varphi(p))\frac{\mathrm{d}q^i}{\mathrm{d}t}(0)$$

as follows:

$$v_p(f) = \left.\frac{\partial f}{\partial q^i}\right|_p \frac{\mathrm{d}q^i}{\mathrm{d}t}(0) .$$

Because the partial derivatives as defined in 31.2.3 form a basis for the tangent space (by construction), one can see that this equation is in fact an expansion of the tangent vector v_p in terms of that basis. It follows that vectors tangent to curves[a] are also tangent vectors according to the first definition.

To prove the other direction one has to show that the partial derivative operators can be constructed as vectors tangent to curves. A tangent vector can be expressed, according to the first construction, in the following way:

$$v_p = v^i \left.\frac{\partial}{\partial q^i}\right|_p ,$$

where the definition $v = (v^1, \dots, v^n)$ was used. One can then construct the curve $\gamma : t \mapsto \varphi^{-1}(q_0 + vt)$. It is obvious that the tangent vector v_p is tangent to the curve γ. From this, it follows that there exists an isomorphism between the tangent vectors

from the first definition and the equivalence classes of vectors tangent to curves from the second definition.

Although the previous equivalence implies that the tangent space construction using germs of curves gives a vector space, one could also check the vector space axioms directly. First, one should prove that the sum of vectors tangent to the curves γ and δ is again a vector tangent to some curve $\chi : \mathbb{R} \to M$. To this end, define the curve

$$\chi(t) \equiv \varphi^{-1} \circ \big(\varphi \circ \gamma(t) + \varphi \circ \delta(t) - \varphi(p)\big)\,,$$

where φ is again the coordinate map in some chart (U, φ) around $p \in M$. Using Eq. (31.8), one can find

$$\begin{aligned} v_{p,\chi}(f) &= \frac{\partial(f \circ \varphi^{-1})}{\partial q^i}(\varphi(p))\frac{\mathrm{d}(\varphi^i \circ \chi)}{\mathrm{d}t}(0) \\ &= \frac{\partial(f \circ \varphi^{-1})}{\partial q^i}(\varphi(p))\frac{\mathrm{d}}{\mathrm{d}t}\big(\varphi^i \circ \gamma + \varphi^i \circ \delta - \varphi^i(p)\big) \\ &= \frac{\partial(f \circ \varphi^{-1})}{\partial q^i}(\varphi(p))\left(\frac{\mathrm{d}(\varphi^i \circ \gamma)}{\mathrm{d}t} + \frac{\mathrm{d}(\varphi^i \circ \delta)}{\mathrm{d}t}\right) \\ &= v_{p,\gamma}(f) + v_{p,\delta}(f)\,. \end{aligned}$$

The constant term $-\varphi(p)$ in the definition of χ is necessary to make sure that $\chi(0) = \gamma(0) = \delta(0) = p$. The axiom of scalar multiplication by a number $\lambda \in \mathbb{R}$ can be proven similarly by defining the curve

$$\chi(t) = \varphi^{-1} \circ \big[\lambda\big(\varphi \circ \gamma(t)\big)\big]\,. \qquad \square$$

[a] More precisely, representatives of equivalence classes of vectors tangent to curves.

31.3 Submanifolds

31.3.1 Immersions and submersions

In this section, arbitrary differentiable functions $f : M \to N$ are considered. By Df_p, the Jacobian of some local representation of f is meant.[2]

Definition 31.3.1 (Regular point). A point $p \in M$ where Df_p is surjective.

Definition 31.3.2 (Regular value). A point $q \in N$ for which every element of the preimage $f^{-1}(q)$ is a regular.

Definition 31.3.3 (Critical point). A point $p \in M$ that is not regular. The image of a critical point is called a **critical value**. In some parts of the literature, a different

[2] A coordinate-independent definition can be given using the notion of *differential* (see Definition 34.1.7). This map will locally be represented by the Jacobian of f.

definition is used. There, a critical point is a point $p \in \mathsf{dom}(f)$ where the rank of Df_p is not maximal. When $\mathsf{dim}(M) < \mathsf{dim}(N)$, every point is critical according to the first definition, but not to the latter. When $\mathsf{dim}(M) \geq \mathsf{dim}(N)$, the definitions coincide.

Property 31.3.4 (Criticality). A point $p \in \mathsf{dom}(f)$ is critical if and only if there exists a chart (U, φ) containing p for which $\partial_i f(p) = 0$.

Corollary 31.3.5. A point $p \in \mathsf{dom}(f)$ is regular if and only if $\partial_i f(p) \neq 0$ for all charts (U, φ) containing p.

Theorem 31.3.6 (Sard). *Consider a differentiable function $f : M \to N$ and let $k_0 := \mathsf{max}\{1, m-n+1\}$, with $\mathsf{dim}(M) = m$ and $\mathsf{dim}(N) = n$. If $f \in C^k(M, N)$ with $k \geq k_0$, the set of critical values of f has Lebesgue measure 0.*

Definition 31.3.7 (Immersion). A differentiable function $f : M \to N$ between smooth manifolds for which the derivative is everywhere injective or, equivalently, such that its derivative has maximal rank everywhere:

$$\mathsf{rk}(Df_p) = \mathsf{dim}(M) \qquad \forall p \in M\,. \tag{31.9}$$

Definition 31.3.8 (Submersion). A differentiable function $f : M \to N$ between smooth manifolds such that all $p \in M$ are regular or, equivalently, such that

$$\mathsf{rk}(Df_p) = \mathsf{dim}(N) \qquad \forall p \in M\,. \tag{31.10}$$

31.3.2 Submanifolds

Definition 31.3.9 (Embedding). A differentiable function between smooth manifolds that is both an immersion and an embedding in the topological sense (Definition 6.2.13). This implies that the (sub)manifold topology coincides with the subspace topology (Example 6.1.4).

Definition 31.3.10 (Embedded submanifold). Let M, N be smooth manifolds. N is called an embedded or **regular submanifold** of M if there exists an embedding $f : M \hookrightarrow N$.

Definition 31.3.11 (Slice). Consider two positive integers $m < n$. The space $\mathbb{R}^m$ can be canonically identified with a subspace of $\mathbb{R}^n$ as follows:

$$\mathbb{R}^m \cong \mathbb{R}^m \times \{0, \dots, 0\} \hookrightarrow \mathbb{R}^m \times \mathbb{R}^{n-m} \cong \mathbb{R}^n\,. \tag{31.11}$$

Subspaces obtained in this way are called slices.

Alternative Definition 31.3.12 (Embedded submanifold). A subset N of M for which there exists a positive integer $k \in \mathbb{N}$ and such that, for every point $p \in N$, there exists a chart (U, φ) that satisfies

$$\varphi(U \cap N) = \varphi(U) \cap (\mathbb{R}^k \times \underbrace{\{0, \dots, 0\}}_{\mathsf{dim}(M)-k})\,. \tag{31.12}$$

The set $U \cap N$ is called a **slice** of (U, φ) in analogy with the previous definition.

Definition 31.3.13 (Immersed submanifold). Let M, N be smooth manifolds. N is said to be an immersed submanifold of M if there exists an immersion $i : N \hookrightarrow M$. Locally, every immersed submanifold looks like a regular submanifold. Globally, however, the topology does not have to coincide with the subspace topology.

Theorem 31.3.14 (Submersion theorem[3]). *Consider a smooth map $f : M \to N$ and let $p \in N$ be regular. Then $M' = f^{-1}(p)$ is a submanifold of M with codimension* $\dim(N)$.

Definition 31.3.15 (Closed embedded manifold). Let N be an immersed submanifold of M. If the inclusion map $i : N \hookrightarrow M$ is closed (or, equivalently, proper), N is an embedded submanifold. These submanifolds are said to be closed embedded.

Example 31.3.16 (Stiefel manifold). Let V be an inner product space (Definition 22.3.1) over a field $\mathfrak{K}$. The set of orthonormal k-frames can be embedded in $\mathfrak{K}^{n\times k}$. It is a compact, embedded submanifold, called the Stiefel manifold of k-frames over V.

Definition 31.3.17 (Transversal intersection). Consider a smooth manifold M. Two submanifolds $X, Y \subset M$ are said to be transversal (or to intersect transversally) if, at each intersection point $p \in X \cup Y$, the following relation holds:

$$T_pX + T_pY = T_pM\,. \tag{31.13}$$

This is denoted by $X \pitchfork Y$. If the dimensions of X and Y are complementary (in M), the sum becomes a direct sum. If two submanifolds do not intersect at all, they are vacuously transversal (independent of their dimension).

Property 31.3.18 (Codimension). The codimension of transversal intersections is equal to the sum of the codimensions of the intersecting submanifolds. It follows that if the submanifolds have complementary dimensions, the intersection consists of isolated points.

Definition 31.3.19 (Intersection number). By the above property, two closed submanifolds $X, Y \subset M$ with complementary dimensions that intersect transversally have a finite number of intersection points. Given an *orientation* on M (see Section 34.8.1), the oriented sum of intersection points is called the intersection number $I(X, Y)$.

To extend this definition to nontransversal intersections, one can observe that the definition is homotopy invariant. Given a homotopy $H : X \times [0,1] \to Z$, if $H(X,0) \pitchfork Y$ and $H(X,1) \pitchfork Y$, then

$$I\big(H(X,0), Y\big) = I\big(H(X,1), Y\big)\,. \tag{31.14}$$

So, to define the intersection number of nontransversally intersecting submanifolds, one simply chooses a transverse (homotopical) deformation. By the invariance property, the result does not depend on the choice of deformation.

[3] Also called the **regular value theorem**.

Property 31.3.20 (Euler characteristic). Consider a closed manifold M. The Euler characteristic (Definition 7.2.18) is given by

$$\chi(M) = I(\Delta_M, \Delta_M)\,, \tag{31.15}$$

where $\Delta_M \in M \times M$ is the diagonal of M.

31.4 Manifolds with boundary

Definition 31.4.1 (Manifold with boundary). Let $\mathbb{H}^n$ denote the upper half space:

$$\mathbb{H}^n := \mathbb{R}^{n-1} \times \mathbb{R}^+ = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid x_n \geq 0\}\,. \tag{31.16}$$

An n-dimensional manifold with boundary is defined as a topological space M equipped with a maximal atlas consisting of 'regular' charts (U, φ), where U is diffeomorphic to $\mathbb{R}^n$ (these points are called **interior points**), and **boundary charts** (V, ϕ), where V is diffeomorphic to $\mathbb{H}^n$ (these points are called **boundary points**).

Remark 31.4.2 (Boundary). The boundary ∂M, consisting of all boundary points of M as defined in the above definition, should not be confused with the topological boundary of M. In general, these are different sets. Similarly, the interior $\mathrm{Int}(M) := M \backslash \partial M$, in the sense of manifolds, should not be confused with the topological interior.

Definition 31.4.3 (Manifold with corners). Analogous to the definition of a manifold with boundaries, one can define a manifold with corners using **corner charts** of the form

$$\varphi : U \to \mathbb{R}^k \times (\mathbb{R}^+)^l\,. \tag{31.17}$$

In contrast to the case of manifolds with boundary, one does need to add an extra requirement when working with higher order corners. For every two charts (U, φ) and (V, ψ), the transition function should preserve the corners:

$$\varphi \circ \psi^{-1}(V \cap \{0\} \times \mathbb{R}^k) \subset \{0\} \times \mathbb{R}^k\,. \tag{31.18}$$

Remark 31.4.4. In the topological setting, every manifold with corners (even higher-order ones) is homeomorphic to a manifold with boundary. However, when working with smooth structures, this result fails. In general, there exists no such diffeomorphism and, consequently, one has to make a distinction between the types of corners.

31.4.1 Cobordism theory ♣

Definition 31.4.5 (Cobordism). Two manifolds X, Y are said to be **(co)bordant** if there exists a manifold with boundary M such that $\partial M = X \sqcup Y$. The manifold M is called a (co)bordism between X and Y.

Remark. In the category of *oriented manifolds* (see Section 34.8.1), one can also define a cobordism, but there the manifolds X, Y should respect the *orientation* of ∂M.

Definition 31.4.6 (Cobordism group). Under the operation of disjoint union, the closed n-dimensional manifolds, modulo cobordisms, form a commutative group Ω_n. Under Cartesian products, these match together to form a graded-commutative ring $\Omega = \bigoplus_{n=0}^{+\infty} \Omega_n$.

@@ COMPLETE @@

31.5 Surgery theory ♣

Definition 31.5.1 (Dehn twist). Consider an *orientable* surface M together with a simple closed curve c. A *tubular neighbourhood* T (see Definition 34.2.10) of c is homeomorphic to an annulus and, hence, allows a parametrization $(e^{i\alpha}, t)$ where $\alpha \in [0, 2\pi[$ and $t \in [0, 1]$. A Dehn twist about c is an automorphism that is given by $(e^{i\alpha}, t) \mapsto (e^{i(\alpha+2\pi t)}, t)$ on T and restricts to the identity outside of it.

@@ COMPLETE @@

Chapter 32

Lie groups and Lie algebras

References for this chapter are Fulton and Harris (2004); Jeevanjee (2015). The main reference for the section on coadjoint orbits is Witten (1988). For some concepts such as vector fields and pushforwards, the reader is referred to Chapter 34.

Chapter Contents

32.1 Lie groups **575**
32.1.1 Left invariant vector fields 575
32.1.2 One-parameter subgroups 576
32.2 Lie algebras **576**
32.2.1 Definitions 576
32.2.2 Examples 579
32.2.3 Exponential map 580
32.2.4 Solvable Lie algebras 581
32.2.5 Simple Lie algebras 582
32.2.6 Central extensions 583
32.3 Representation theory **583**
32.3.1 Lie groups 583
32.3.2 Lie algebras 584
32.3.3 Coadjoint orbits 584
32.4 Structure **585**
32.4.1 Killing form 585
32.4.2 Weights, roots and Dynkin diagrams 586
32.4.3 Highest weight theory 592
32.4.4 Coxeter groups 593

32.4.5 Kac–Moody algebras ♣ 594
32.4.6 Universal enveloping algebra 596
32.5 Group contractions **598**
32.6 Lie algebra cohomology ♣ **598**

32.1 Lie groups

Definition 32.1.1 (Lie group). A group that is also a differentiable manifold such that both the multiplication and inversion maps are smooth. For complex Lie groups, the definition of a *complex manifold* is required (see Chapter 39).

Definition 32.1.2 (Lie subgroup). A subset of a Lie group that is both a subgroup and an immersed submanifold. If it is a regular submanifold, it is sometimes called a **regular Lie subgroup**.

Theorem 32.1.3 (Closed subgroup theorem[1]). *A closed subgroup of a Lie group is a regular Lie subgroup.*

Property 32.1.4 (Generating neighbourhoods). Let G be a connected Lie group. Every neighbourhood U_e of the identity generates G, i.e. every element $g \in G$ can be written as a word in U_e.

Definition 32.1.5 (Isogeny). Two Lie groups are said to be isogenous if one is a covering space (Definition 6.2.15) of the other. The covering map is then called an isogeny between the groups.

32.1.1 Left invariant vector fields

Definition 32.1.6 (Left-invariant vector field). Let G be a Lie group and let X be a vector field on G. X is said to be left-invariant if the following equivariance relation holds for all $g \in G$:

$$L_{g,*}X(h) = X(gh)\,, \tag{32.1}$$

where L denotes the regular (left) action on G. The term 'left-invariant vector field' is often abbreviated as **LIVF**.

Property 32.1.7. The set $\mathfrak{X}^L(G)$ of LIVFs on a (real) Lie group G is a vector space over $\mathbb{R}$.

Property 32.1.8 (Tangent space). The map $L_{g,*}$ is an isomorphism for every $g \in G$. It follows that a LIVF is uniquely determined by its value at the identity $e \in G$. Fur-

[1]Sometimes simply called **Cartan's theorem**.

thermore, for every $v \in T_eG$, there exists a LIVF $X \in \mathfrak{X}^L(G)$ such that $X(e) = v$. This mapping is an isomorphism from T_eG to $\mathfrak{X}^L(G)$.

32.1.2 One-parameter subgroups

Definition 32.1.9 (One-parameter subgroup). A Lie group morphism $\Phi : \mathbb{R} \to G$ from the additive group of real numbers to G.

Property 32.1.10. Let $\Phi : \mathbb{R} \to G$ be a one-parameter subgroup and let $\Psi : G \to H$ be a Lie group morphism, then $\Psi \circ \Phi : \mathbb{R} \to H$ is a one-parameter subgroup of H.

Remark 32.1.11. The above definition and property can be generalized to the topological setting if Lie groups and Lie group morphisms are replaced by topological groups and continuous group morphisms.

Property 32.1.12 (Left-invariant vector fields). All LIVFs X are *complete* (see Definition 34.3.14 for a general statement), i.e. for every LIVF X one can find an integral curve γ_X with initial condition $\gamma_X(0) = e$ for which the maximal *flow domain* (Definition 34.3.12) $D(X)$ is $\mathbb{R}$. This implies that the associated flow σ_t determines a one-parameter subgroup of G. Conversely, for every one-parameter subgroup $\phi(t)$, one can construct a LIVF by taking $X := \phi'(0)$. This correspondence is in fact a bijection.

32.2 Lie algebras

There are two ways to define a Lie algebra. The first one is purely algebraic and consists of a vector space equipped with a multiplication operation satisfying certain conditions. The second one establishes a direct correspondence between Lie groups and Lie algebras.

32.2.1 Definitions

Definition 32.2.1 (Pre-Lie algebra). Let V be a k-vector space equipped with a binary operation $(\cdot,\cdot) : V \times V \to V$. $(V,(\cdot,\cdot))$ is a pre-Lie algebra if the bracket satisfies the following conditions:

1. **Bilinearity**: $(\lambda x + y, z) = \lambda(x,z) + (y,z)$ for all $\lambda \in k$,
2. **Invariance**: $((x,y),z) - (x,(y,z)) = ((x,z),y) - (x,(z,y))$.

Definition 32.2.2 (Lie algebra). Let V be a k-vector space equipped with a binary operation $[\cdot,\cdot] : V \times V \to V$, called the **Lie bracket**. $(V,[\cdot,\cdot])$ is a Lie algebra if the Lie bracket satisfies the following conditions:

1. **Bilinearity**: $[\lambda x + y, z] = \lambda[x,z] + [y,z]$ for all $\lambda \in k$,

2. **Alternativity**: $[y, y] = 0$, and
3. **Jacobi identity**: $[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0$.

Remark 32.2.3. Note that, often, the alternativity condition is replaced by an antisymmetry condition. However, this is only equivalent over fields of characteristic $\mathrm{char}(k) \neq 2$. For $\mathrm{char}(k) = 2$, one only has that alternativity implies antisymmetry. Since the fields of interest will almost exclusively be $\mathbb{R}$ or $\mathbb{C}$, this will not pose any problems and, hence, the antisymmetry condition will always be used.

Example 32.2.4. Every pre-Lie algebra V induces a Lie algebra through the commutator

$$[x, y] := (x, y) - (y, x)\,. \tag{32.2}$$

Note that this does not assume that the pre-Lie structure on V comes from an associative algebra operation. Only the invariance condition is required for the Jacobi identity to hold.

Definition 32.2.5 (Poisson algebra). A vector space V equipped with two bilinear operations $\star$ and $\{\cdot, \cdot\}$ that satisfy the following conditions:

1. The couple $(V, \star)$ is an associative algebra.
2. The couple $(V, \{\cdot, \cdot\})$ is a Lie algebra.
3. The **Poisson bracket** $\{\cdot, \cdot\}$ acts as a derivation 31.2.1 with respect to the operation $\star$, i.e.
$$\{x, y \star z\} = \{x, y\} \star z + y \star \{x, z\}$$
for all x, y and $z \in V$.

Definition 32.2.6 (Structure constants). Because Lie algebras are closed under the Lie bracket, the Lie bracket of two basis elements can again be expressed in terms of the same basis $\{e_k\}_{k \in I}$:

$$[e_i, e_j] = \sum_{k \in I} c_{ij}{}^k e_k\,. \tag{32.3}$$

The coefficients $c_{ij}{}^k$ are called the structure constants of the Lie algebra. (Note that these constants are basis-dependent.)

Property 32.2.7 (Isomorphism). Two Lie algebras are isomorphic if one can find bases for these Lie algebras such that the associated structure constants are equal.

Example 32.2.8 (Lie algebra of LIVFs). Consider the vector space $\mathfrak{X}^L(G)$ of LIVFs on a Lie group G. Using Property 34.3.20, one can show that the commutator also defines a LIVF on G. It follows that $\mathfrak{X}^L(G)$ is closed under Lie brackets and, hence, is a Lie algebra.

For the following alternative definition, Property 32.1.8 is used to relate the above Lie algebra of left-invariant vector fields to the tangent space at the identity.

Alternative Definition 32.2.9 (Lie algebra of Lie group). Let G be a Lie group. The tangent space $\mathfrak{g} := T_e G$ has the structure of a Lie algebra, where the Lie bracket is induced by the *commutator of vector fields* (see Property 34.3.20) in the following way:

$$[x, y]_{\mathfrak{g}} := L_{g^{-1},*}\left[L_{g,*}x, L_{g,*}y\right], \tag{32.4}$$

where $x, y \in T_e G$ and where $[\cdot,\cdot]$ denotes the Lie bracket on $\mathfrak{X}^L(G)$. This induces an isomorphism of Lie algebras $\mathfrak{g} \cong \mathfrak{X}^L(G)$. This isomorphism will be freely used throughout this text.

Notation 32.2.10. Lie algebras are generally denoted by fraktur symbols. For example, as in the previous definition, the Lie algebra associated with the Lie group G is often denoted by $\mathfrak{g}$.

Theorem 32.2.11 (Ado). *Every finite-dimensional Lie algebra can be embedded as a subalgebra of* $\mathfrak{gl}_n \cong M_n$.

Theorem 32.2.12 (Lie's third theorem). *Every finite-dimensional Lie algebra* $\mathfrak{g}$ *is the Lie algebra of a unique simply-connected Lie group* G.

Definition 32.2.13 (Lie algebra morphism). A map $\Phi : \mathfrak{g} \to \mathfrak{h}$ that satisfies the equation

$$\Phi\big([x, y]\big) = \big[\Phi(x), \Phi(y)\big] \tag{32.5}$$

for all $x, y \in \mathfrak{g}$.

Formula 32.2.14 (Induced morphism). Let $\phi : G \to H$ be a Lie group morphism. This morphism induces a Lie algebra morphism $\Phi : \mathfrak{g} \to \mathfrak{h}$ given by

$$\Phi(x) := \frac{\mathrm{d}}{\mathrm{d}t}\phi(e^{tx})\bigg|_{t=0} \tag{32.6}$$

or, equivalently,

$$\phi(e^{tx}) = e^{t\Phi(x)}. \tag{32.7}$$

The morphism Φ is sometimes called the **differential** of ϕ.

Property 32.2.15 (Homomorphism theorem[2]). Let G, H be two Lie groups and assume that G is simply-connected. Every Lie algebra morphism $\Phi : \mathfrak{g} \to \mathfrak{h}$ corresponds to a unique Lie group morphism $\phi : G \to H$ such that $\Phi = \phi_*$.

[2] Also called **Lie's second theorem**.

32.2.2 Examples

Example 32.2.16 (Cross product). The cross product $\times : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ turns $\mathbb{R}^3$ into a Lie algebra.

Example 32.2.17 (General linear group). An interesting example is the Lie algebra associated to the Lie group of invertible complex[3] matrices $\mathsf{GL}_n(\mathbb{C})$. This Lie group is a subset of its own Lie algebra $\mathfrak{gl}_n(\mathbb{C}) = M_n(\mathbb{C})$. It follows that, for every $A \in \mathsf{GL}_n(\mathbb{C})$ and every $B \in \mathfrak{gl}_n(\mathbb{C})$, the following equality holds:

$$L_{A,*}(B) = L_A(B)\,. \tag{32.8}$$

Corollary 32.2.18. By noting that the endomorphism ring $\mathsf{End}(V)$ of an n-dimensional vector space V is given by the matrix ring $M_n(K)$, it can be seen that $\mathsf{End}(V)$ also forms a Lie algebra when equipped with the commutator of linear maps.

Definition 32.2.19 (Derivation). Given a Lie algebra $\mathfrak{g}$, the space of derivations $\mathfrak{Der}(\mathfrak{g})$ is defined as the space of linear maps $\mathsf{d} : \mathfrak{g} \to \mathfrak{g}$ such that

$$\mathsf{d}([x, y]) = [\mathsf{d}x, y] + [x, \mathsf{d}y] \tag{32.9}$$

for all $x, y \in \mathfrak{g}$. This vector space inherits the Lie algebra structure on endomorphisms.

Example 32.2.20 (Isometries). The Lie algebra $\mathfrak{isom}(V)$ associated with the group of isometries $\mathsf{Isom}(V)$ of a nondegenerate Hermitian form satisfies the condition

$$\langle Xv, w\rangle = -\langle v, Xw\rangle \tag{32.10}$$

for all elements $X \in \mathfrak{isom}(V)$. It follows that the Lie algebra of isometries consists of all skew-Hermitian operators.

Two explicit examples for common matrix Lie groups are given.

Example 32.2.21 (Lie algebra of O(3)). This Lie algebra is isomorphic to the set of 3×3 skew-symmetric matrices. It is important to note that $\mathfrak{o}(3) = \mathfrak{so}(3)$. The structure constants of this Lie algebra are given by the Levi-Civita symbol (Definition 23.4.8): $c_{ijk} = \varepsilon_{ijk}$.

Example 32.2.22 (Lie algebra of SU(2)). This Lie algebra is isomorphic to the set of 2×2 traceless skew-Hermitian matrices. This result can be generalized to arbitrary $n \in \mathbb{N}$.

Another important example is obtained by considering the subset of $\mathsf{GL}(2, \mathbb{C})$ consisting of matrices with unit determinant.

[3] This result is also valid for real matrices.

Example 32.2.23 (Lie algebra of SL(2, ℂ)). To compute the Lie bracket of the Lie algebra $\mathfrak{sl}(2,\mathbb{C})$, one needs to find the action of $L_{g,*}$ on any element of $\mathfrak{sl}(2,\mathbb{C})$. This is given by:

$$L_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right),*}\left(\left.\frac{\partial}{\partial x^i}\right|_e\right) = \begin{pmatrix} a & 0 & b \\ -b & a & 0 \\ c & 0 & \frac{1+bc}{a} \end{pmatrix}^m_i \left.\frac{\partial}{\partial x^m}\right|_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)}, \tag{32.11}$$

where the coordinate chart (U, ϕ) defined by

$$U = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathsf{SL}(2,\mathbb{C}) \middle| a \neq 0 \right\} \tag{32.12}$$

and

$$\phi : U \to \mathbb{C}^3 : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto (a, b, c) \tag{32.13}$$

was used. One can then use this formula to work out the Lie bracket of the basis vectors ∂_i to obtain the following structure relations:

$$\begin{aligned} [\partial_1, \partial_2] &= 2\partial_2\,, \\ [\partial_1, \partial_3] &= -2\partial_3\,, \\ [\partial_2, \partial_3] &= \partial_1\,. \end{aligned} \tag{32.14}$$

32.2.3 Exponential map

Formula 32.2.24 (Exponential map). Let X be a LIVF on G. The exponential map is defined as

$$\exp : \mathfrak{g} \to G : X \mapsto \gamma_X(1)\,, \tag{32.15}$$

where γ_X is the associated one-parameter subgroup from Property 32.1.12.

Property 32.2.25 (Uniqueness). The exponential map is the unique map $\mathfrak{g} \to G$ such that $\exp(0) = e$ and for which the restrictions to the lines through the origin in $\mathfrak{g}$ are one-parameter subgroups of G.

Corollary 32.2.26. Because the identity $\mathbb{1}_{\mathfrak{g}} = \exp_*$ is an isomorphism, the *inverse function theorem* (see Theorem 34.1.10) implies that the image of exp will contain a neighbourhood of the identity $e \in G$. If G is connected, Property 32.1.4 implies that the exponential map generates all of G.

Together with the property that $\psi \circ \exp = \exp \circ \psi_*$ for every Lie group morphism $\psi : G \to H$, one can conclude that if G is connected, a Lie group morphism $\psi : G \to H$ is completely determined by its differential ψ_* at the identity $e \in G$.

Example 32.2.27 (Matrix Lie groups). For matrix Lie groups, one can define the classic matrix exponential:

$$e^{tA} := \sum_{k=0}^{+\infty} \frac{(tA)^k}{k!} . \tag{32.16}$$

This operation defines a one-parameter subgroup of G and, from the uniqueness property above, it follows that this matrix exponential is in fact the exponential map for G. It should be noted that this formula converges for every $A \in M_{m,n}(\mathbb{R})$ and that it is invertible with inverse given by $\exp(-A)$. Using Ado's theorem 32.2.11, one can then use this matrix exponential to represent the exponential map for any (finite-dimensional) Lie algebra.

Remark 32.2.28. If G is a (connected) compact Lie group, the exponential map is surjective. However, because the associated Lie algebra $\mathfrak{g}$ is clearly noncompact, the exponential map cannot be homeomorphic and, hence, cannot be injective.

Formula 32.2.29 (Baker–Campbell–Hausdorff formula). Consider the equation

$$z = \ln\big(\exp(x)\exp(y)\big) , \tag{32.17}$$

where $x, y \in \mathfrak{g}$. The solution is given by the following formula:

$$z = x + y + \frac{1}{2}[x,y] + \frac{1}{12}[x,[x,y]] - \frac{1}{12}[y,[x,y]] + \cdots . \tag{32.18}$$

One should note that this formula will only converge if x, y are sufficiently 'small'. For matrix Lie algebras, this means that

$$\|x\| + \|y\| < \frac{\ln(2)}{2} \tag{32.19}$$

in the Hilbert–Schmidt norm (Formula 22.4.7). Due to the closure under the Lie bracket, the exponent in the BCH formula is also an element of the Lie algebra. So, this formula gives an expression for Lie group multiplication in terms of (Lie brackets of) Lie algebra elements (whenever the formula converges).

Corollary 32.2.30 (Lie product formula[4]). Let $\mathfrak{g}$ be a Lie algebra. The following formula applies to any $x, y \in \mathfrak{g}$:

$$e^{x+y} = \lim_{n\to\infty} \left(e^{x/n} e^{y/n}\right)^n . \tag{32.20}$$

32.2.4 Solvable Lie algebras

Definition 32.2.31 (Normalizer). The normalizer of a subset of a Lie algebra $S \subset \mathfrak{g}$ is the space of elements $x \in \mathfrak{g}$ that satisfy $[x, S] \subseteq S$.

[4] Also called the **Lie–Trotter formula**. Later, extensions where given by *Kato* and *Suzuki* for certain unbounded operators.

Definition 32.2.32 (Centralizer). The centralizer of a subset of a Lie algebra $S \subset \mathfrak{g}$ is the space of elements $x \in \mathfrak{g}$ that satisfy $[x, S] = 0$.

Definition 32.2.33 (Derived algebra). Let $\mathfrak{g}$ be a Lie algebra. The derived Lie algebra is defined as follows:

$$[\mathfrak{g}, \mathfrak{g}] := \left\{[x, y] \mid x, y \in \mathfrak{g}\right\}. \tag{32.21}$$

Definition 32.2.34 (Solvable Lie algebra). Consider the sequence of derived subalgebras

$$\mathfrak{g} \geq [\mathfrak{g}, \mathfrak{g}] \geq \big[[\mathfrak{g}, \mathfrak{g}], [\mathfrak{g}, \mathfrak{g}]\big] \geq \cdots . \tag{32.22}$$

If this sequence ends in the zero space, the Lie algebra $\mathfrak{g}$ is said to be solvable.

Remark 32.2.35. In general, one can define the derived series for any ideal and, accordingly, define solvability for ideals.

Definition 32.2.36 (Radical). The largest solvable ideal of a Lie algebra.

32.2.5 Simple Lie algebras

Definition 32.2.37 (Direct sum). The direct sum $\mathfrak{g} \oplus \mathfrak{h}$ of two Lie algebras $\mathfrak{g}, \mathfrak{h}$ is defined as the direct sum (Definition 22.1.15) of the underlying vector spaces where the Lie bracket is extended using the relation

$$[x, y] = 0 \tag{32.23}$$

for all $x \in \mathfrak{g}$ and $y \in \mathfrak{h}$.

Definition 32.2.38 (Semidirect product). The semidirect product (or sum) $\mathfrak{g} \ltimes \mathfrak{h}$ of two Lie algebras $\mathfrak{g}, \mathfrak{h}$ with respect to a Lie algebra morphism $\rho : \mathfrak{g} \to \mathsf{Der}(\mathfrak{h})$ is defined as the direct sum of the underlying vector spaces where the Lie bracket is extended using the relation

$$[g, h] := \rho(g)(h) \tag{32.24}$$

for all $g \in \mathfrak{g}, h \in \mathfrak{h}$. This also turns $\mathfrak{h}$ into an ideal.

Definition 32.2.39 (Simple Lie algebra). A Lie algebra is said to be simple if it is non-Abelian and if it has no nontrivial ideals.

Definition 32.2.40 (Semisimple Lie algebra). A Lie algebra is said to be semisimple if it is the direct sum of simple Lie algebras.

Theorem 32.2.41 (Levi decomposition). *Let $\mathfrak{g}$ be a finite-dimensional Lie algebra. It can be decomposed as follows:*

$$\mathfrak{g} = \mathfrak{R} \ltimes (\mathfrak{L}_1 \oplus \cdots \oplus \mathfrak{L}_n), \tag{32.25}$$

where $\mathfrak{R}$ is the radical of $\mathfrak{g}$ and the algebras $\mathfrak{L}_i$ are simple subalgebras.

Definition 32.2.42 (Levi factor). The semisimple subalgebra $\mathfrak{L}_1 \oplus \cdots \oplus \mathfrak{L}_n$ in the Levi decomposition of $\mathfrak{g}$ is called the **Levi subalgebra** or **Levi factor** of $\mathfrak{g}$.

32.2.6 Central extensions

Definition 32.2.43 (Central extension). A central extension of a Lie algebra $\mathfrak{g}$ by an Abelian Lie algebra $\mathfrak{a}$ is an exact sequence (Definition 5.2.2) of Lie algebras of the form

$$0 \longrightarrow \mathfrak{a} \longrightarrow \mathfrak{h} \longrightarrow \mathfrak{g} \longrightarrow 0\,, \tag{32.26}$$

where the image of $\mathfrak{a}$ lies in the center of $\mathfrak{h}$.

Construction 32.2.44 (Extension by cocycles). Consider a Lie algebra morphism $\Theta : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{a}$ with the following properties:

1. Θ is bilinear,
2. Θ is antisymmetric, and
3. $\Theta\big([u,v],w\big) + \Theta\big([v,w],u\big) + \Theta\big([w,u],v\big) = 0$ for all $u,v,w \in \mathfrak{g}$.

Such a morphism is called a **Lie algebra 2-cocycle** (see Section 32.6 further below for more information). Now, every 2-cocycle $\Theta : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{a}$ induces a central extension of $\mathfrak{g}$ by $\mathfrak{a}$ in the following way. Because the exact sequences characterizing central extensions (of Lie algebras) are, in particular, short exact sequences of vector spaces, they always split and, hence, one can always choose $\mathfrak{h} = \mathfrak{g} \oplus \mathfrak{a}$ to be the underlying vector space. The Lie bracket on this space is then defined as follows:

$$[v \oplus \lambda, w \oplus \mu] := [v,w]_{\mathfrak{g}} \oplus \Theta(v,w)\,. \tag{32.27}$$

32.3 Representation theory

32.3.1 Lie groups

Definition 32.3.1 (Representation of Lie groups). A representation of a Lie group G on a vector space V is a Lie group morphism $\rho : G \to \mathsf{GL}(V)$.

Example 32.3.2 (Adjoint representation of Lie groups). Let G be a Lie group and consider the adjoint action

$$\mathsf{Ad}_g : h \mapsto ghg^{-1}\,. \tag{32.28}$$

The adjoint representation of G on $\mathfrak{g}$ is defined as the differential $\mathsf{Ad}_{g,*}$ at the identity element. For matrix Lie groups, this becomes

$$\mathsf{Ad}_g : T_eG \to T_eG : x \mapsto gxg^{-1}\,. \tag{32.29}$$

32.3.2 Lie algebras

Definition 32.3.3 (Representation of Lie algebras). A representation of a Lie algebra $\mathfrak{g}$ on a vector space V is a Lie algebra morphism $\rho : \mathfrak{g} \to \mathsf{End}(V)$.

The morphism induced by $\mathsf{Ad} : G \to H$ is precisely $\mathsf{ad} : \mathfrak{g} \to \mathfrak{h}$. Informally, one can thus say that the infinitesimal version of the similarity transformation is given by the commutator.

Example 32.3.4 (Adjoint representation of Lie algebras). Using the fact that the adjoint representation of Lie groups is smooth, one obtains:

$$\mathsf{ad}_x = \mathsf{Ad}_{g,*}\,, \tag{32.30}$$

where $g = e^{tx}$. More explicitly, the adjoint map is given by

$$\mathsf{ad}_x(y) = [x, y]\,, \tag{32.31}$$

for all $x, y \in \mathfrak{g}$.

Property 32.3.5 (Faithfulness). The adjoint representation ad_x is faithful for all $x \in \mathfrak{g}$.

Property 32.3.6 (Jacobi identity). Given the antisymmetry of the Lie bracket, the Jacobi identity is equivalent to $\mathsf{ad} : \mathfrak{g} \to \mathsf{End}(\mathfrak{g})$ being a Lie algebra morphism, i.e.

$$\mathsf{ad}_{[x,y]} = [\mathsf{ad}_x, \mathsf{ad}_y]\,. \tag{32.32}$$

Formula 32.3.7 (Structure constants). Let $\{e_i\}_{i \leq n}$ be a basis of the Lie algebra $\mathfrak{g}$. The structure constants are related to the adjoint representation as follows:

$$(\mathsf{ad}_{e_i})^j_k = c_{ik}{}^j\,. \tag{32.33}$$

Corollary 32.3.8 (Commutator). For the algebra of the general linear group GL_n, the Lie bracket is given by the commutator:

$$[A, B] = AB - BA\,. \tag{32.34}$$

32.3.3 Coadjoint orbits

Definition 32.3.9 (Coadjoint representation). The representation of a Lie group G on the dual space $\mathfrak{g}^*$ defined by

$$\langle \mathsf{Ad}^*_g(\alpha), v\rangle := \langle \alpha, \mathsf{Ad}^{-1}_g(v)\rangle\,. \tag{32.35}$$

Infinitesimally, this induces a representation of $\mathfrak{g}$ on its linear dual:

$$\mathsf{ad}^*_x : \alpha \mapsto -\alpha \circ \mathsf{ad}_x\,. \tag{32.36}$$

Definition 32.3.10 (Coadjoint orbit). Given an element $\alpha \in \mathfrak{g}^*$, the coadjoint orbit Ω_α is defined as the orbit of α under the action of G. This orbit can also be defined as the homogeneous space G/G_α.

The following important construction shows that every coadjoint orbit is in fact canonically a *symplectic manifold* (see Chapter 37).

Definition 32.3.11 (Kirillov–Kostant–Souriau form). Consider a coadjoint orbit Ω_α. Because $\alpha \in \mathfrak{g}^*$ is an element of the coadjoint representation of G, the tangent vectors to Ω_α (at α) are elements of the induced representation of $\mathfrak{g}$, i.e. for any tangent vector v one can write $v = \mathsf{ad}^*_x(\alpha)$ for some $x \in \mathfrak{g}$. Now, a symplectic form on Ω_α is defined as follows:

$$\omega_\alpha\left(\mathsf{ad}^*_x(\alpha), \mathsf{ad}^*_y(\alpha)\right) = \langle \alpha, [x, y] \rangle \,. \tag{32.37}$$

32.4 Structure

32.4.1 Killing form

Definition 32.4.1 (Killing form). Let $\mathfrak{g}$ be a finite-dimensional Lie algebra over the field k. The Killing form (also called the **Cartan–Killing** form) on $\mathfrak{g}$ is defined as the following symmetric bilinear form:

$$K : \mathfrak{g} \otimes \mathfrak{g} \to k : x \otimes y \mapsto \mathsf{tr}(\mathsf{ad}_x \circ \mathsf{ad}_y) \,. \tag{32.38}$$

The trace can be calculated by choosing a (finite-dimensional) representation of the Lie algebra using Ado's theorem 32.2.11.

From Formula 32.3.7, one can derive the value of the Killing form on the basis $\{e_i\}_{i \leq n}$:

$$K_{ij} = c_{ik}{}^l c_{jl}{}^k \,, \tag{32.39}$$

where $c_{ij}{}^k$ are the structure constants.

Theorem 32.4.2 (Cartan's criterion). *A Lie algebra is semisimple if and only if its Killing form is nondegenerate.*

Property 32.4.3. If a Lie group G is compact, the Killing form of its Lie algebra $\mathfrak{g}$ is negative-definite.

Corollary 32.4.4. Let G be a compact Lie group. If its Lie algebra is semisimple, the Killing form K induces a metric

$$g : x \otimes y \mapsto -\mathsf{tr}(\mathsf{ad}_x \circ \mathsf{ad}_y) = -K(x, y) \tag{32.40}$$

which turns the corresponding Lie group G into a *Riemannian manifold* (see Chapter 36).

Property 32.4.5 (Killing form is invariant). The Killing-form is Ad-invariant:

$$K\big(\mathsf{Ad}_g(x), \mathsf{Ad}_g(y)\big) = K(x, y) \tag{32.41}$$

for all $g \in G$. More generally, the Killing form is invariant under all automorphisms of $\mathfrak{g}$.

Corollary 32.4.6. The adjoint map ad_z is skewsymmetric with respect to the Killing form:

$$K(\mathsf{ad}_z x, y) = -K(x, \mathsf{ad}_z y)\,. \tag{32.42}$$

Property 32.4.7 (Invariant forms). For a simple Lie algebra, every (ad-)invariant symmetric bilinear form is a scalar multiple of the Killing form.

Example 32.4.8. For $\mathfrak{su}(n)$, the trace can easily be seen to be ad-invariant and, hence, satisfy the above property. The exact relation is given by

$$\mathsf{tr}(AB) = 2nK(A, B)\,. \tag{32.43}$$

Property 32.4.9 (Antisymmetric structure constants). When the Lie algebra $\mathfrak{g}$ is compact and semisimple, i.e. when the Killing form induces a metric, one can find a basis of $\mathfrak{g}$, constructed by orthonormalizing a given basis with respect to the Killing metric, such that the structure constants are invariant under cyclic permutation of the indices:

$$c_{ijk} = c_{jki}\,. \tag{32.44}$$

A corollary of this property is also that the structure constants become totally antisymmetric.

Construction 32.4.10 (Induced Killing form). Let $\mathfrak{g}$ be a Lie algebra and let V be a vector space equipped with a Lie algebra representation $\rho : \mathfrak{g} \to \mathsf{End}(V)$. One can define a Killing form associated with ρ in the following way:

$$K_\rho(x, y) := \mathsf{tr}\big(\rho(x) \circ \rho(y)\big)\,. \tag{32.45}$$

This is a generalization of Definition 32.4.1 which reduces to the Killing form K when choosing V to be $\mathfrak{g}$ in the adjoint representation.

32.4.2 Weights, roots and Dynkin diagrams

From here on, the base field is assumed to be algebraically closed (take $\mathbb{C}$ for simplicity).

Definition 32.4.11 (Cartan subalgebra). Let $\mathfrak{g}$ be a Lie algebra. A subalgebra $\mathfrak{h}$ is called a Cartan subalgebra if it satisfies the following two conditions:

1. **Nilpotency**: Its lower central series terminates:

$$\exists n \in \mathbb{N} : \underbrace{[\mathfrak{h}, [\mathfrak{h}, [\mathfrak{h}, \ldots]]]}_{n \text{ times}} = 0\,. \tag{32.46}$$

2. **Self-normalizing**:

$$\forall x \in \mathfrak{h} : [x, y] \in \mathfrak{h} \implies y \in \mathfrak{h} . \tag{32.47}$$

From here on, it will also be assumed that all Lie algebras are finite-dimensional. This assumption is motivated by the following property.

Property 32.4.12. Every finite-dimensional Lie algebra contains a Cartan subalgebra.

Property 32.4.13. If $\mathfrak{g}$ is semisimple, its Cartan subalgebra is Abelian.

Construction 32.4.14. Let $\mathfrak{g}$ be a semisimple Lie algebra. A Cartan subalgebra $\mathfrak{h}$ can be constructed as follows. For some integer $k \leq \dim(\mathfrak{g})$, choose k linearly independent, commuting vectors $\{h_i\}_{i \leq k} \subset \mathfrak{g}$. (The existence of such a choice, which is equivalent to requiring simultaneous diagonalization, is only guaranteed for semisimple Lie algebras.) If this set can be extended to a basis $\{h_i\}_{i \leq k} \cup \{g_j\}_{j \leq \dim(\mathfrak{g})-k}$ of $\mathfrak{g}$ such that every g_j is a nontrivial eigenvector of the adjoint map ad_{h_i} for all $i \in I$, the algebra $\mathfrak{h} = \mathrm{span}\{h_i\}_{i \leq k}$ is a Cartan subalgebra.

Definition 32.4.15 (Weight space). Let V be a representation of a Lie algebra $\mathfrak{g}$ with Cartan subalgebra $\mathfrak{h}$. For every linear functional λ on $\mathfrak{h}$, the weight space V_λ with **weight** λ is defined as follows:

$$V_\lambda := \{v \in V \mid \forall h \in \mathfrak{h} : h \cdot v = \lambda(H)v\} . \tag{32.48}$$

The nonzero elements of a weight space are called **weight vectors**. If the representation V can be decomposed as a direct sum of weight spaces, it is called a **weight module**:

$$V = \bigoplus_{\lambda \in \mathfrak{h}^*} V_\lambda . \tag{32.49}$$

In the case where V is the adjoint representation, the (nonzero) weights are called **roots**.

Definition 32.4.16 (Root). Let $\mathfrak{g}$ be a Lie algebra with Cartan subalgebra $\mathfrak{h}$. From the definition of a Cartan subalgebra, it follows that for all $h \in \mathfrak{h}$:

$$[h, g_j] = \alpha_j(h) g_j , \tag{32.50}$$

where $\{g_j\}_{j \in J}$ is the basis extension of $\mathfrak{g}$ with respect to $\mathfrak{h}$. The eigenvalues have been written suggestively as maps $\alpha_j(h)$ acting on $\mathfrak{h}$, which is justified because, due to the definition of the g_j's and the structure of the above formula, the α_j's for different h piece together to give linear maps. By comparing the formula to the definition of weights above, it can be seen that the α_j's are the weights of the adjoint representation. The nonzero weights are called the roots of $\mathfrak{g}$ and form the so-called **root system** Φ.

It follows that there exists a weight space decomposition of $\mathfrak{g}$:

$$\mathfrak{g} = \mathfrak{h} \oplus \bigoplus_{\lambda \in \Phi} \mathfrak{g}_\lambda , \tag{32.51}$$

where the one-dimensional spaces $\mathfrak{g}_\lambda$ are the weight spaces associated to the roots λ ($\mathfrak{h}$ is equal to $\mathfrak{g}_0$ in this notation).

Property 32.4.17. If $\alpha \in \Phi$, then $-\alpha \in \Phi$. Furthermore, if $\alpha \in \Phi$ and $c\alpha \in \Phi$, then $c = \pm 1$.

This property says that the root system Φ is not linearly independent. To introduce some kind of basis, the following notion is introduced.

Definition 32.4.18 (Simple root). The set of simple roots Δ is a linearly independent subset of Φ such that every element $\lambda \in \Phi$ can be written as

$$\lambda = \pm \sum_i^n a_i \lambda_i \,, \tag{32.52}$$

where $a_i \in \mathbb{N}$ and $\lambda_i \in \Delta$. (Such a set always exists.) This definition enforces the expansion coefficients a_i of a root λ to be either all positive or all negative.

More generally, one can define the following equivalence relation on a root system.

Definition 32.4.19 (Positive roots). Let Φ be the root system of a given Lie algebra $\mathfrak{g}$. Because the only scalar multiples of a root $\lambda \in \Phi$ in the root system are $\pm\lambda$, one can define a set of positive roots Φ^+ as follows:

- $\lambda \in \Phi^+ \implies -\lambda \notin \Phi^+$, and
- $\alpha, \beta \in \Phi^+ \wedge \alpha + \beta \in \Phi \implies \alpha + \beta \in \Phi^+$.

The simple roots are then exactly the elements in Φ^+ that cannot be written as a sum of other elements in Φ^+.

Definition 32.4.20 (Triangular decomposition). Given a choice of positive roots Φ^+, one can decompose the Lie algebra $\mathfrak{g}$ as follows:

$$\mathfrak{g} = \mathfrak{n}_- \oplus \mathfrak{h} \oplus \mathfrak{n}_+ \,, \tag{32.53}$$

where $\mathfrak{n}_\pm = \bigoplus_{\alpha \in \Phi^\pm} \mathfrak{g}_\alpha$. The subalgebra $\mathfrak{h} \oplus \mathfrak{n}_+$ is called the **Borel subalgebra**. It is the maximal solvable subalgebra of $\mathfrak{g}$ (Definition 32.2.34).

Property 32.4.21 (Rank). Let $\mathfrak{h}$ be a Cartan subalgebra. The set of simple roots Δ forms a basis for the dual space $\mathfrak{h}^*$ (over $\mathbb{C}$). It follows that the cardinality of Δ is equal to the dimension of the Cartan subalgebra. This dimension is called the **rank** of the Lie algebra.

Definition 32.4.22 (Weyl group). For every simple root λ, one can construct a Householder transformation (Definition 22.3.16) as follows:

$$\sigma_\lambda : \mathsf{span}_{\mathbb{R}}(\Delta) \to \mathsf{span}_{\mathbb{R}}(\Delta) : \mu \mapsto \mu - 2\frac{\langle \mu \mid \lambda \rangle}{\langle \lambda \mid \lambda \rangle}\lambda \,, \tag{32.54}$$

where the inner product $\langle \cdot \mid \cdot \rangle$ is the dual Killing form. The Weyl group W is defined as the group generated by these transformations.

Property 32.4.23 (Weyl group symmetries). Every root $\alpha \in \Phi$ can be written as $\alpha = \sigma(\mu)$ for some $\mu \in \Delta$ and $\sigma \in W$. Furthermore, the root system Φ is closed under the action of W. In particular, it can be shown that the Weyl group W is precisely the symmetry group of the root system Φ and the isometry group of the Killing form (and its dual).

Definition 32.4.24 (Coroot). Consider the *sharp map* (see Eq. (36.3)), where the metric g is given by the Killing form K. The dual Killing form K^* is then a proper inner product (when restricted to the real span of Δ) defined as

$$K^*(\cdot,\cdot) = K(\cdot^\sharp, \cdot^\sharp)\,. \tag{32.55}$$

The restriction of the dual Killing form to the real span $\mathfrak{h}_0^*$ of the roots of $\mathfrak{g}$ induces a dual space $\mathfrak{h}_0 \subset \mathfrak{h}$. The coroot $\alpha^\vee \in \mathfrak{h}_0$ associated to a root α is then defined by the following formula (where the metric isomorphism $\mathfrak{h}_0 \cong \mathfrak{h}_0^{**}$ is used):

$$\alpha^\vee := 2\frac{\langle \alpha \mid \cdot \rangle}{\langle \alpha \mid \alpha \rangle} \equiv 2\frac{\alpha}{\langle \alpha \mid \alpha \rangle}\,. \tag{32.56}$$

With this definition the Weyl transformations can be rewritten as follows:

$$\sigma_\lambda : \mathsf{span}_{\mathbb{R}}(\Delta) \to \mathsf{span}_{\mathbb{R}}(\Delta) : \mu \mapsto \mu - \mu(\lambda^\vee)\lambda\,. \tag{32.57}$$

Notation 32.4.25 (Coroot). Sometimes it is more favourable to denote the coroot associated to α by H^α. This convention will be adopted in the remainder of this chapter.

Definition 32.4.26 (Weyl chamber). Given a choice of positive roots Φ^+, the closed (fundamental) Weyl chamber associated to this ordering is defined as the subset of $\mathfrak{h}_0$ that contains the elements w satisfying the following equation for all $\gamma \in \Phi^+$:

$$w(H^\gamma) \geq 0\,. \tag{32.58}$$

Elements of a Weyl chamber are called **dominant weights**.

Property 32.4.27 (Weyl group). The Weyl group acts transitively on the set of Weyl chambers and, accordingly, on the orderings of the root system.

Property 32.4.28. Let $\alpha \in \Phi$ be a root. Choose a generating element E^α of the weight space $\mathfrak{g}_\alpha$ associated to α and let F^α be the generator of the weight space $\mathfrak{g}_{-\alpha}$ such that $\mathsf{span}\{E^\alpha, F^\alpha, [E^\alpha, F^\alpha]\}$ is a one-dimensional simple Lie algebra. The following relations hold (for $\beta \neq \pm\alpha$):

- $\beta(H^\alpha) = 2\frac{\langle \alpha|\beta\rangle}{\langle \alpha|\alpha\rangle} \in \mathbb{Z}$,
- $[H^\alpha, E^\alpha] = \alpha(H^\alpha)E^\alpha = 2E^\alpha$, and

- $[H^\alpha, F^\alpha] = -\alpha(H^\alpha)F^\alpha = -2F^\alpha$.

Definition 32.4.29 (Cartan matrix). Let $\lambda_i, \lambda_j \in \Delta$ be simple roots. Because the Weyl group is the symmetry group of the root system,

$$\sigma_{\lambda_i}(\lambda_j) = \lambda_j - 2\frac{\langle \lambda_i | \lambda_j \rangle}{\langle \lambda_i \mid \lambda_i \rangle}\lambda_i$$

is a root. From the properties above, it then follows that the quantity

$$C_{ij} := 2\frac{\langle \lambda_i \mid \lambda_j \rangle}{\langle \lambda_i \mid \lambda_i \rangle} = \lambda_j(H^{\lambda_i}) \tag{32.59}$$

is an integer. The matrix formed by these numbers is called the Cartan matrix.

Property 32.4.30 (Properties of Cartan matrix). The Cartan matrix C satisfies the following properties:

- $C_{ii} = 2$,
- $C_{ij} \in \mathbb{Z}_{\leq 0}$ if $i \neq j$, and
- $C_{ij} = 0 \iff C_{ji} = 0$.

This last property does not imply that the Cartan matrix is symmetric. The fact that it is not symmetric can immediately be seen from its definition. However,

- it is *symmetrizable*, i.e. there exist a positive diagonal matrix D and a symmetric matrix S such that $C = DS$.
- it is positive definite.

Definition 32.4.31 (Bond number). For all indices $i \neq j$, the bond number n_{ij} is defined as follows:

$$n_{ij} := C_{ij}C_{ji}\,. \tag{32.60}$$

Using the definition of the coefficients C_{ij}, it can be seen that n_{ij} is an integer equal to $4\cos^2 \sphericalangle(\lambda_i, \lambda_j)$. This implies that n_{ij} can only take on the values $0, 1, 2, 3$. The value 4 would only be possible if the angle between λ_i and λ_j is 0, but this can only occur in the case where $i = j$ (which was excluded by definition).

Remark 32.4.32. In the case of $n_{ij} = 2$ or $n_{ij} = 3$, two possibilities arise: $C_{ij} > C_{ji}$ or $C_{ij} < C_{ji}$ for $i < j$. From the definition of the Cartan integers and the symmetry of the dual Killing form, these cases correspond to $\langle \lambda_i \mid \lambda_i \rangle < \langle \lambda_j \mid \lambda_j \rangle$ and $\langle \lambda_i \mid \lambda_i \rangle > \langle \lambda_j \mid \lambda_j \rangle$.

Construction 32.4.33 (Dynkin diagram). For a semisimple Lie algebra $\mathfrak{g}$ with simple roots Δ, one can draw a diagram using the following rules:

1. For every simple root $\lambda \in \Delta$, draw a circle $\circ$.

2. Draw n_{ij} lines between the circles associated to λ_i and λ_j.
3. If $n_{ij} = 2$ or 3, add a $<$ or $>$ sign based on the relation between their lengths (see previous remark).

Property 32.4.34 (Classification). The Dynkin diagrams can be classified as follows (for every type, the first three examples are given):

- A_n:

- $B_n, n \geq 2$:

- $C_n, n \geq 2$:

- $D_n, n \geq 4$:

These are the only possible diagrams for simple Lie algebras with exception of E_6, E_7, E_8, F_4 and G_2, the so-called *exceptional Lie algebras*.

Example 32.4.35 (Special linear group). By looking at the Lie brackets in Eq. (32.14), one can see that the one-element set $\{X_1\}$ forms a Cartan subalgebra of $\mathfrak{sl}(2, \mathbb{C})$. From the same equation, it is also immediately clear that the simple root set Δ is given by the one-element set $\{\lambda \in \mathfrak{sl}^*(2, \mathbb{C}) \mid \lambda(X_1) \mapsto 2\}$. Hence, the Dynkin diagram for $\mathfrak{sl}(2, \mathbb{C})$ is A_1.

Theorem 32.4.36 (Cartan–Killing). *Every finite-dimensional simple Lie algebra over $\mathbb{C}$ can be reconstructed from its set of simple roots Δ.*

Construction 32.4.37 (Chevalley–Serre). Given a Dynkin diagram of a simple Lie algebra, one can reconstruct the original Lie algebra $\mathfrak{g}$ over $\mathbb{C}$ up to isomorphism.

The number of nodes is equal to the number of simple roots and, hence, determines the rank n of $\mathfrak{g}$. First, construct the free Lie algebra on $3n$ generators $\{E_i, F_i, H_i\}_{i \leq n}$. The Cartan subalgebra $\mathfrak{h} \leq \mathfrak{g}$ is constructed from the generators H_i by imposing the following relations:

- $[H_i, H_j] = 0$,
- $[H_i, E_j] = a_{ij} E_j$,

- $[H_i, F_j] = -a_{ij}F_j$, and
- $[E_i, F_j] = \delta_{ij}H_j$,

where the numbers (a_{ij}) form the Cartan matrix obtained by reversing Construction 32.4.33. To complete the reconstruction, one imposes the following additional relations:

- $\mathrm{ad}_{E_i}^{|a_{ij}|+1}(E_j) = 0$, and
- $\mathrm{ad}_{F_i}^{|a_{ij}|+1}(F_j) = 0$.

The first set of relations are called the **Chevalley relations** and the last two are called the **Serre relations**. The full construction is called the **Chevalley–Serre presentation**.

Remark 32.4.38. For composite diagrams that correspond to semisimple Lie algebras, one first has to construct the Lie algebras corresponding to every simple diagram and then take the direct sum.

Property 32.4.39 ($\mathfrak{sl}_2$). For every Cartan matrix of rank $n \in \mathbb{N}$, the triplets $\{E_i, F_i, H_i\}$ with $i \leq n$ generate $\mathfrak{sl}_2$-algebras. All (semi)simple Lie algebras are thus in some sense built up from $\mathfrak{sl}_2$-algebras, similar to how in algebraic topology simplicial complexes are constructed by gluing simplices together.

32.4.3 Highest weight theory

Definition 32.4.40 (Algebraically integral). An element $H \in \mathfrak{h}_0$ is said to be algebraically integral if its value on every root is an integer. The set of all algebraically integral elements is called the **weight lattice**.

Definition 32.4.41 (Fundamental weight). Let $\Delta = \{\alpha_i\}_{i \leq \mathrm{rk}(\mathfrak{g})}$ be the set of simple roots. The fundamental weights $\{\omega_i\}_{i \leq \mathrm{rk}(\mathfrak{g})}$ are defined as the elements of $\mathfrak{h}_0^*$ for which the following formula is satisfied for all $i, j \leq |\Delta|$:

$$\omega_i(H^{\alpha_j}) = \delta_{ij}\,. \tag{32.61}$$

This implies that an element $\lambda \in \mathfrak{h}_0$ is algebraically integral if it is an integral combination of fundamental weights.

Definition 32.4.42 (Ordering of weights). Let Φ^+ be a choice of positive roots. One can define a partial ordering on the set of weights $\mathfrak{h}_0^*$ in the following way:

$$\mu \leq \lambda \iff \lambda - \mu \in \mathrm{span}_{\mathbb{N}}(\Phi^+)\,. \tag{32.62}$$

Definition 32.4.43 (Highest weight vector). Consider a representation V of a Lie algebra $\mathfrak{g}$. An element $v \in V$ is said to be a highest weight vector if it is a weight vector that is annihilated by all positive roots. A highest weight module is a weight module that is generated by a highest weight vector.

Theorem 32.4.44 (Highest weight theorem). *Consider a finite-dimensional Lie algebra. The following statements hold:*

- *Every finite-dimensional irreducible representation has a unique dominant integral highest weight.*
- *If two irreducible representations have the same highest weight, they are isomorphic.*
- *Every dominant integral weight is the highest weight of an irreducible finite-dimensional representation.*

32.4.4 Coxeter groups

The properties about root systems in Section 32.4.2 can be abstracted into the following definition.

Definition 32.4.45 (Root system). Let $(E, \langle \cdot \mid \cdot \rangle)$ denote a finite-dimensional vector space equipped with a positive-definite, symmetric bilinear form. A root system on E is a finite set Φ of nonzero vectors satisfying the following conditions:

1. **Basis**: $E = \mathsf{span}_{\mathbb{R}}(\Phi)$.
2. **Scalar closure**: $\alpha \in \Phi \wedge \lambda\alpha \in \Phi \implies \lambda = \pm 1$.
3. **Reflection closure**: $\sigma_\alpha \Phi = \Phi$, where σ_α is the Householder transformation induced by $\alpha \in \Phi$.
4. **Integrality**: $\alpha, \beta \in \Phi \implies 2\frac{\langle \alpha | \beta \rangle}{\langle \alpha | \alpha \rangle} \in \mathbb{Z}$.

Many of the other constructions, such as the definition of positive and simple roots, also carry over from the case of Lie algebras.

Remark 32.4.46 (Lie algebras). The root system of a Lie algebra $\mathfrak{g}$ relative to a Cartan subalgebra $\mathfrak{h}$ is actually a root system for $\mathfrak{h}^*$ in the sense of the foregoing definition.

Property 32.4.47 (Finite reflections). Every finite reflection group, i.e. a finite group generated by Householder transformations, admits a root system. Conversely, every root system generates a finite reflection group given by its Weyl group (Definition 32.4.22).

The Weyl group induced by a root system is a specific instance of the following notion.

Definition 32.4.48 (Coxeter group). A group admitting a presentation of the form

$$\langle r_1, \dots, r_n \mid (r_i r_j)^{m_{ij}} = e \rangle, \tag{32.63}$$

where $m_{ii} = 1$ and $m_{ij} \geq 2$ for $i \neq j$. If $m_{ij} = \infty$, no relation is imposed.

32.4.5 Kac–Moody algebras ♣

Construction 32.4.49 (Kac–Moody algebra). Consider the Cartan matrix A associated to a finite-dimensional (semi)simple Lie algebra. This matrix has the properties listed in Property 32.4.30. If one drops the positivity condition, the definition of a **generalized Cartan matrix** is obtained. For such matrices, one can construct a (possibly infinite-dimensional) Lie algebra using an analogue of the Chevalley–Serre presentation.

However, since a generalized Cartan matrix might have a vanishing determinant, one cannot use the Chevalley–Serre presentation as given in Construction 32.4.37 as the constructed roots might be linearly dependent. Luckily, this problem can easily be solved. Let A be an $n \times n$ (generalized) Cartan matrix. First, choose a complex vector space $\mathfrak{h}$ equipped with, for every $i \leq n$, a simple root α_i (resp. coroot H^{α_i}) in $\mathfrak{h}^*$ (resp. $\mathfrak{h}$) such that these are linearly independent and satisfy the condition $\alpha_i(H^{\alpha_j}) = A_{ij}$. (Such a **realization** always exists.) It can be shown that $\mathfrak{h}$ satisfies $\dim(\mathfrak{h}) \geq 2n - \mathrm{rk}(A)$. A minimal realization, i.e. one that satisfies $\dim(\mathfrak{h}) = 2n - \mathrm{rk}(A)$, is unique up to isomorphism.

Then, construct the direct sum $\mathfrak{g}$ of the free Lie algebra on $2n$ generators $\{E_i, F_i\}_{i \leq n}$ with $\mathfrak{h}$ and take the quotient by the following relations:

- $[E_i, F_j] = \delta_{ij} H^{\alpha_i}$,
- $[H, H'] = 0$ for $H, H' \in \mathfrak{h}$,
- $[H, E_i] = \alpha_i(H) E_i$ for $H \in \mathfrak{h}$, and
- $[H, F_i] = -\alpha_i(H) F_i$ for $H \in \mathfrak{h}$.

Given this Lie algebra, one can find the unique maximal ideal $\mathfrak{m} \leq \mathfrak{g}$ for which $\mathfrak{m} \cap \mathfrak{h} = \{0\}$. The quotient algebra $\mathfrak{g}/\mathfrak{m}$ is called the Kac–Moody algebra associated to A.[5]

Although an analogue of the Serre relations is not included, it can be shown that these relations still hold (see for example Amini (2014)). In fact, when A is symmetrizable, the ideal $\mathfrak{m}$ is exactly generated by the Serre relations.

Remark 32.4.50 (Classification). There exist three distinct classes of Kac–Moody algebras:

- If A is positive-definite, one obtains a finite-dimensional (semi)simple Lie algebra. A is also said to be of **finite type**.
- If A is positive-semidefinite, one obtains a Kac–Moody algebra of **affine type**. In

[5] *Kac* proved that for symmetrizable matrices this construction is equivalent to the Chevalley–Serre presentation with the generators as given here.

fact, it can be shown for generalized Cartan matrices that affinity is equivalent to the existence of a unique (up to scaling) real vector v such that $Av = 0$.

- If A is indefinite, one obtains a Kac–Moody of **indefinite type**.

Definition 32.4.51 (Loop algebra). Consider a finite-dimensional Lie algebra $\mathfrak{g}$. Its associated **loop algebra** $L\mathfrak{g}$ is the vector space $\mathfrak{g} \otimes \mathbb{C}[t, t^{-1}]$ equipped with the following Lie bracket:

$$[x \otimes t^k, y \otimes t^l] := [x, y]_{\mathfrak{g}} \otimes t^{k+l} . \tag{32.64}$$

Because this strongly resembles the definition of the ring of Laurent polynomials $\mathfrak{K}[t, t^{-1}]$ over a field $\mathfrak{K}$, the loop algebra is sometimes denoted by $\mathfrak{g}[t, t^{-1}]$.

Equivalently, one can obtain the loop algebra as the space of polynomial maps from S^1 to $\mathfrak{g}$ (hence the name). Furthermore, if G is the Lie group associated to $\mathfrak{g}$ and LG denotes its (free) loop group (Definition 7.1.12), then LG has the natural structure of a (infinite-dimensional) Lie group and its Lie algebra is precisely $L\mathfrak{g}$.

Definition 32.4.52 (Affine Lie algebra). Given a simple Lie algebra $\mathfrak{g}$, one can define the affine Lie algebra $\hat{\mathfrak{g}}$ as the central extension of the loop algebra $L\mathfrak{g}$ by $\mathbb{C}$ associated to the cocycle

$$\Theta : (x \otimes t^k, y \otimes t^l) \mapsto kK(x, y)\delta_{k+l,0} , \tag{32.65}$$

where K is the Killing form on $\mathfrak{g}$. The generator c of $\mathbb{C}$ is often called the **central element** since it is mapped to an element in the center of $\hat{\mathfrak{g}}$. This cocycle can also be defined using the residue (Definition 17.5.19) of a Laurent polynomial:

$$\Theta(f, g) := \mathrm{Res}\big[K(f, g)\big]_t , \tag{32.66}$$

where the Killing form is extended from $\mathfrak{g}$ to $L\mathfrak{g}$ by $K(x \otimes t^k, y \otimes t^l) := K(x, y)t^{k+l}$.

However, to obtain a well-behaved affine Kac–Moody algebra, one also needs to extend this affine Lie algebra by a derivation. Observe that the loop algebra and, accordingly, the affine Lie algebra is $\mathbb{Z}$-graded. A well-defined derivation is then obtained through multiplication by the grading. To this end, add a formal generator d together with the following relations (the central element from the previous step is still denoted by c):

$$\begin{aligned} [d, x \otimes P(t)] &= x \otimes t\frac{\mathrm{d}}{\mathrm{d}t}P(t) \\ [d, c] &= 0 , \end{aligned} \tag{32.67}$$

where $P(t) \in \mathbb{C}[t, t^{-1}]$.

Definition 32.4.53 (Extended Cartan matrix). Consider a simple Lie algebra $\mathfrak{g}$ with Cartan matrix A and denote the associated Chevalley generators by $\{E_i, F_i, H_i\}_{i \leq n}$. There

exist unique (up to scaling) nonzero elements E_0, F_0 such that

$$\begin{aligned}[E_0, F_i] &= 0\,,\\ [F_0, E_i] &= 0\end{aligned} \tag{32.68}$$

for all $1 \leq i \leq n$. This also implies that $H_0 := [E_0, F_0]$ is a linear combination of $\{H_i\}_{i\leq n}$. The elements E_0, F_0 can be normalized by enforcing the following Chevalley-type relations:

$$\begin{aligned}[H_0, E_0] &= 2E_0\,,\\ [H_0, F_0] &= -2F_0\,.\end{aligned} \tag{32.69}$$

The extended Cartan matrix $\widehat{A}$ is defined by adjoining a new row and column to A. These new entries are defined by

$$\begin{aligned}[H_0, E_i] &=: a_{0i}E_i\,,\\ [H_i, E_0] &=: a_{i0}E_0\,.\end{aligned} \tag{32.70}$$

Definition 32.4.54 (Twisted Kac–Moody algebra). Consider an indecomposable generalized Cartan matrix. Such a matrix is of affine type if and only if all its proper principal minors are positive-definite. Hence, by deleting the first row and column, one obtains the Cartan matrix B for a finite-dimenionsal simple Lie algebra $\mathfrak{g}(B)$. It can be shown that the affine Kac–Moody algebra $\mathfrak{g}(A)$, as defined in Construction 32.4.49, is isomorphic to the affine Kac–Moody algebra as constructed above, starting from the simple Lie algebra $\mathfrak{g}(B)$, if and only if $A = \hat{B}$.

All affine Kac–Moody algebras that are isomorphic to Lie algebras defined in this way are said to be **untwisted**. All other affine Kac–Moody algebras are said to be **twisted**.

Property 32.4.55. Let $\mathfrak{g}$ be a simple Lie algebra with Cartan matrix A. The affine Lie algebra $\hat{\mathfrak{g}}$ is isomorphic to the derived subalgebra $[\mathfrak{g}(\widehat{A}), \mathfrak{g}(\widehat{A})]$ of the Kac–Moody algebra $\mathfrak{g}(\widehat{A})$.

32.4.6 Universal enveloping algebra

Definition 32.4.56 (Universal enveloping algebra). Let $\mathfrak{g}$ be a Lie algebra and consider its tensor algebra $T(\mathfrak{g})$. The universal enveloping algebra $U(\mathfrak{g})$ is defined as the quotient of $T(\mathfrak{g})$ by the two-sided ideal generated by $\{g \otimes h - h \otimes g - [g, h] \mid g, h \in \mathfrak{g}\}$.

Construction 32.4.57. If the Chevalley–Serre presentation from Construction 32.4.37 is regarded as a presentation for a unital associative algebra instead of as a Lie algebra presentation (by replacing the Lie bracket by the commutator constructed from the algebra multiplication), the universal enveloping algebra $U(\mathfrak{g})$ of $\mathfrak{g}$ is obtained.

Theorem 32.4.58 (Poincaré–Birkhoff–Witt). *Let $\mathfrak{g}$ be a Lie algebra with a totally ordered basis $\{g_i\}_{i\leq \dim(\mathfrak{g})}$. The monomials of the form $g_1^{m_1} g_2^{m_2} \cdots g_N^{m_N}$ constitute a basis for $U(\mathfrak{g})$.*

Definition 32.4.59 (Casimir invariant[6]). Let $\mathfrak{g}$ be a Lie algebra. A Casimir invariant is an element in the center of $U(\mathfrak{g})$.

Formula 32.4.60 (Quadratic Casimir invariant). Consider a Lie algebra representation $\rho : \mathfrak{g} \to \mathsf{End}(V)$ and let $\{X_i\}_{i \leq \dim(\mathfrak{g})}$ be a basis for $\mathfrak{g}$. The (quadratic) Casimir invariant associated with ρ is given by

$$\Omega_\rho := \sum_{i=1}^{\dim(\mathfrak{g})} \rho(X_i) \circ \rho(\xi_i)\,, \tag{32.71}$$

where the set $\{\xi_i\}_{i \leq n}$ is defined by the relation $K_\rho(X_i, \xi_j) = \delta_{ij}$ using the Killing form (Construction 32.4.10).

Property 32.4.61 (Casimir invariants of irreducible representations). When the representation $\rho : \mathfrak{g} \to \mathsf{End}(V)$ is irreducible, Schur's lemma 24.2.3 says that

$$\Omega_\rho = c_\rho \mathbb{1}_V\,. \tag{32.72}$$

By taking the trace of this formula and using Construction 32.4.10, it can be seen that

$$c_\rho = \frac{\dim(\mathfrak{g})}{\dim(V)}\,. \tag{32.73}$$

Definition 32.4.62 (Verma module). Consider a finite-dimensional Lie algebra $\mathfrak{g}$ with Borel subalgebra $\mathfrak{b}$. The Verma module with highest weight λ is defined as follows[7]:

$$V(\lambda) := U(\mathfrak{g}) \otimes_{U(\mathfrak{b})} \mathbb{C}_\lambda\,, \tag{32.74}$$

where $\mathbb{C}_\lambda$ is the one-dimensional left $\mathfrak{b}$-module on which the Cartan subalgebra acts by weight λ and $\mathfrak{n}_+ \subset \mathfrak{b}$ acts trivially. $U(\mathfrak{g})$ contains $U(\mathfrak{b})$ as a subalgebra by the Poincaré–Birkhoff–Witt theorem 32.4.58 and, hence, becomes a right $U(\mathfrak{b})$-module through right multiplication. Since $U(\mathfrak{g})$ is trivially a left module over itself, the Verma module also becomes a left $U(\mathfrak{g})$-module.

The Verma module with highest weight λ can also be defined using a quotient construction.

Alternative Definition 32.4.63. Let $I_\lambda \subset U(\mathfrak{g})$ be the left ideal generated by the following elements (these relations precisely give the conditions for a highest weight vector):

- $x_\alpha \in \mathfrak{g}_\alpha$ for all positive roots α, and
- $h - \lambda(h)1$ for all $h \in \mathfrak{h}$.

The Verma module $V(\lambda)$ is isomorphic to the quotient $U(\mathfrak{g})/I_\lambda$.

[6] Also known as a **Casimir operator** or **Casimir element**.

[7] This can be seen as an 'extension of scalars'-procedure where a $U(\mathfrak{b})$-module is turned into a $U(\mathfrak{g})$-module.

The importance of Verma modules is given by the following property.

Property 32.4.64 (Highest weight modules). The Verma module $V(\lambda)$ is a highest weight module with highest weight vector $1 \otimes 1$, where the former 1 indicates the unit of $U(\mathfrak{g})$ and the latter indicates the unit of $\mathbb{C}$. Furthermore, every highest weight module with highest weight λ is a quotient of the Verma module $V(\lambda)$.

Property 32.4.65 (Basis of Verma module). A basis for $V(\lambda)$ is given by the monomials

$$F_{\alpha_1}^{m_1} F_{\alpha_2}^{m_2} \cdots F_{\alpha_N}^{m_N} v_\lambda \,, \tag{32.75}$$

where v_λ is the highest weight vector, α_i are negative roots, $m_i \in \mathbb{N}$ and $F_{\alpha_i} \in \mathfrak{g}_{\alpha_i}$.

32.5 Group contractions

Definition 32.5.1 (Inönü–Wigner contraction). Consider an n-dimensional Lie group G with Lie algebra $\mathfrak{g}$ and choose a basis $\{e_i\}_{i \leq n}$ for $\mathfrak{g}$. A nonsingular transformation of the basis would leave the structure of the group unchanged. However, this nonsingular transformation can be rewritten in terms of a singular transformation:

$$U = u + \varepsilon w \,. \tag{32.76}$$

The group contraction is obtained by taking the limit $\varepsilon \longrightarrow 0$. In terms of the structure constants, this is equivalent to setting some of the structure constants to zero, thereby obtaining a subalgebra (and its associated subgroup). It can be shown that there exists a bijection between continuous subgroups and group contractions. The Lie algebra elements belonging to the contracted subalgebra form an Abelian invariant subalgebra and, hence, generate an Abelian invariant subgroup. The group contraction $\widetilde{G}$ is obtained as the quotient group of G with respect to this Abelian subgroup.

Example 32.5.2 (Galilei group). The *Galilei group* in $d \in \mathbb{N}$ dimensions can be obtained as a group contraction of the *inhomogeneous Lorentz group* in $d + 1$ (spacetime) dimensions with respect to time displacements and spatial rotations.

32.6 Lie algebra cohomology ♣

Although the construction of Lie algebra-cohomology can be generalized almost verbatim to the infinite-dimensional case, it is only stated for finite dimensions (the following definition, for example, is only valid for finite-dimensional algebras).

Definition 32.6.1 (Chevalley–Eilenberg algebra). Let $\mathfrak{g}$ be a finite-dimensional Lie algebra. Consider a basis $\{t_a\}_{a \leq \dim(\mathfrak{g})}$ of $\mathfrak{g}$ and its linear dual $\{t^a\}_{a \leq \dim(\mathfrak{g})}$. The Chevalley–Eilenberg algebra $\mathsf{CE}(\mathfrak{g})$ is defined as the Grassmann algebra $\Lambda^\bullet \mathfrak{g}^*$ with a dg-algebra

structure induced by the differential[8]

$$\mathrm{d}t^a := -\frac{1}{2} c_{bc}{}^a t^b \wedge t^c \,, \tag{32.77}$$

where $c_{bc}{}^a$ are the structure constants of $\mathfrak{g}$.

By analogy with the case of group (co)homology as in Section 5.6.1, the (co)homology of a Lie algebra is defined through the Tor- and Ext-functors. The natural choice of ring in the case of Lie algebras is the universal enveloping algebra $U(\mathfrak{g})$. The tensor and hom-operations underlying the construction are defined with respect to the trivial $U(\mathfrak{g})$-module $\mathfrak{K}$ (the underlying field of the Lie algebra). This gives

$$\begin{aligned} H^i_{\mathrm{Lie}}(\mathfrak{g};M) &:= \mathsf{Ext}^i_{U(\mathfrak{g})}(\mathfrak{K},M)\,, \\ H_i^{\mathrm{Lie}}(\mathfrak{g};M) &:= \mathsf{Tor}_i^{U(\mathfrak{g})}(\mathfrak{K},M)\,, \end{aligned} \tag{32.78}$$

where M is a $\mathfrak{g}$-module and, by extension, a $U(\mathfrak{g})$-module.

For simplicity, only cohomology will be considered here. The chapter on homological algebra, in particular Section 5.4.1, says that one has to find a projective resolution of $\mathfrak{K}$ to determine the Ext-functor in terms of hom-sets $\mathsf{Hom}(\cdot,M)$. It can be shown that such a resolution is given by the tensor product $U(\mathfrak{g}) \otimes \Lambda^\bullet\mathfrak{g}$:

$$\mathsf{Ext}^i_{U(\mathfrak{g})}(\mathfrak{K},M) = H^i\big(\mathsf{Hom}_{\mathfrak{g}}(U(\mathfrak{g}) \otimes_{\mathfrak{K}} \Lambda^\bullet\mathfrak{g}, M)\big) \cong H^i\big(\mathsf{Hom}_{\mathfrak{K}}(\Lambda^\bullet\mathfrak{g}, M)\big)\,, \tag{32.79}$$

where the differential of the (middle) complex is given by

$$\begin{aligned} \mathrm{d}(u \otimes g_1 \wedge \cdots \wedge g_n) := &\sum_i (-1)^{i+1} u g_i \otimes (g_1 \wedge \cdots \wedge \hat{g}_i \wedge \cdots \wedge g_n) \\ &+ \sum_{i<j} (-1)^{i+j} u \otimes [g_i, g_j] \wedge \cdots \wedge \hat{g}_i \wedge \cdots \wedge \hat{g}_j \wedge \cdots \wedge g_n\,, \end{aligned} \tag{32.80}$$

where as usual the caret $\hat{}$ indicates the omission of a factor. In the case $M = \mathfrak{K}$, the hom-complex can easily be seen to be the Chevalley–Eilenberg algebra $\mathsf{CE}(\mathfrak{g}) \cong \Lambda^\bullet\mathfrak{g}^*$ (as stated above, this latter identification is only valid for finite-dimensional algebras). By a change of coefficients, the general case can be shown to be isomorphic to $\Lambda^\bullet\mathfrak{g}^* \otimes M$, where the usual differential on $\Lambda^\bullet\mathfrak{g}^*$ gets extended by an additional term $(-1)^n \mathrm{d}m \otimes \omega$ with $\mathrm{d}m(g) := g \cdot m$.

Property 32.6.2 (H^0 and H^1). The zeroth cohomology group $H^0(\mathfrak{g};M)$ is equal to the algebra of $\mathfrak{g}$-invariants in M:

$$H^0(\mathfrak{g};M) = \{m \in M \mid \mathfrak{g} \cdot m = m\}\,. \tag{32.81}$$

The first cohomology group with coefficients in $\mathfrak{K}$ is isomorphic to the quotient of $\mathfrak{g}$ by its first derived ideal:

$$H^1(\mathfrak{g}) \cong \mathfrak{g}/[\mathfrak{g},\mathfrak{g}]\,. \tag{32.82}$$

[8] In Section 4.9.2 it was explained how this differential can be obtained as the dual of the Lie bracket.

Property 32.6.3 (Whitehead lemma). If $\mathfrak{g}$ is semisimple, the cohomology groups $H^1(\mathfrak{g})$ and $H^2(\mathfrak{g})$ vanish. Conversely, a Lie algebra $\mathfrak{g}$ is semisimple if and only if $H^1(\mathfrak{g}; M)$ vanishes for all finite-dimensional $\mathfrak{g}$-modules M.

Property 32.6.4 (Classification of central extensions). The 2-cocycles (with values in $\mathfrak{K}$) from Construction 32.2.44 define classes in $H^2(\mathfrak{g})$. Furthermore, a central extension is trivial if and only if its associated cocycle is a coboundary. This says that the central extensions of $\mathfrak{g}$ by $\mathfrak{K}$ are classified by $H^2(\mathfrak{g})$.

Corollary 32.6.5. Semisimple Lie algebras do not admit nontrivial central extensions.

Definition 32.6.6 (Weil algebra). Consider a Lie algebra $\mathfrak{g}$. Its Weil algebra is defined as the semifree dg-algebra $\Lambda^\bullet(\mathfrak{g}^* \oplus \mathfrak{g}^*[1])$ with differential $\mathsf{d}_W := \mathsf{d}_{\mathrm{CE}} + \Pi$, where d_{CE} is the differential on the Chevalley–Eilenberg subalgebra $\mathsf{CE}(\mathfrak{g}) \subset W(\mathfrak{g})$ and Π shifts the degree by 1. The action of d_{CE} on shifted generators is defined through the relation $[\Pi, \mathsf{d}_{\mathrm{CE}}] = 0$.

Definition 32.6.7 (Horizontal elements). The elements of the subalgebra $\Lambda^\bullet\mathfrak{g}^*[1]$ are sometimes called the horizontal elements.

From here on, the subscript will be dropped and the differential of the Weil algebra will be denoted by d. It is clear that the above constructions fit in a short exact sequence.

$$0 \to \ker(p) \to W(\mathfrak{g}) \overset{p}{\twoheadrightarrow} \mathsf{CE}(\mathfrak{g}) \to 0\,, \tag{32.83}$$

where p is the obvious projection map. An important subpace of $\ker(p)$ is given by the algebra of invariant polynomials $\mathsf{inv}(\mathfrak{g})$.

Definition 32.6.8 (Invariant polynomial). A horizontal element ω for which $\mathsf{d}\omega$ is also horizontal. (Sometimes, the horizontality condition is replaced by $\mathsf{d}\omega = 0$.)

It should be noted that, although this definition might seem complicated, it is (for ordinary Lie algebras[9]) equivalent to the definition in terms of Ad-invariant polynomials.

Alternative Definition 32.6.9 (Invariant polynomial). Let G be a Lie group with Lie algebra $\mathfrak{g}$. A polynomial $P \in \mathfrak{K}[\mathfrak{g}]$ is said to be invariant (or Ad-invariant) if

$$P(x) = P(gxg^{-1}) \tag{32.84}$$

for all $x \in \mathfrak{g}$ and $g \in G$. This subalgebra of $\mathfrak{K}[\mathfrak{g}]$ is denoted by $\mathfrak{K}[\mathfrak{g}]^G$.

A concept that will be important later on in the study of characteristic classes on fibre bundles is the transgression map.

Definition 32.6.10 (Transgression). The exact sequence (32.83) induces a long exact sequence in cohomology. An invariant polynomial is said to be in transgression with

[9] The above definition leads to a straightforward generalization in the context of L_∞-algebras.

a cocycle in $\mathsf{CE}(\mathfrak{g})$ if their cohomology classes are related by the connecting morphism. More explicitly, by definition of invariant polynomials, one has $\mathsf{d}\omega = 0$ and, since $W(\mathfrak{g})$ has vanishing cohomology, there exists an element c_ω such that $\omega = \mathsf{d}c_\omega$. By restricting c_ω to $\mathsf{CE}(\mathfrak{g})$, one obtains a $\mathfrak{g}$-cocycle, since $\mathsf{d}_{\mathrm{CE}}c_\omega = 0$.

Example 32.6.11 (Killing form). Consider the invariant polynomial $\langle\cdot,\cdot\rangle$ induced by the Killing form on a semisimple Lie algebra. By transgression one obtains the canonical 3-cocycle $\langle\cdot,[\cdot,\cdot]\rangle$.

Chapter 33

Fibre Bundles

This chapter is formulated in sufficient generality so as to encompass both the topological and smooth setting (or any other setting one might find useful). To this end, the generic terms 'space', 'group' and 'morphism' are used. The reader should choose in which category to work, e.g. topological space, topological group and continuous map in the case of **Top**.

Chapter Contents

33.1 Bundles 602

33.2 Constructions 605

33.3 Sections 606

33.1 Bundles

Definition 33.1.1 (Bundle). A triple (E, B, π) where E and B are spaces and $\pi : E \to B$ is a morphism. Sometimes, π is required to be surjective. However, under this additional restriction, one cannot make the association $\mathbf{Bundle}(B) \cong \mathbf{C}_{/B}$ of categories.

An explicit example in the category **Diff** is the following.

Example 33.1.2 (Fibred manifold). A surjective submersion (Definition 31.3.8)

$$\pi : E \to B,$$

where E is called the **total space**, B the **base space** and π the **projection**. For every point $p \in B$, the set $\pi^{-1}(p)$ is called the **fibre over** p.

The most important example of a bundle is a fibre bundle. Before being able to give the definition, an important concept needs to be introduced.

Definition 33.1.3 (Cocycle). Let B be a space and G a group. A G-valued cocycle on B with respect to an open cover $\{U_i\}_{i\in I}$ is a family of morphisms $g_{ij} : U_i \cap U_j \to G$ that satisfy the **Čech cocycle condition**

$$g_{ij} = g_{ik} \circ g_{kj} \tag{33.1}$$

for all $i, j, k \in I$. Two cocycles (U_i, g_{ij}) and (V_i, h_{ij}) are said to be equivalent if there exist morphisms $\lambda_{i,j} : U_i \cap V_j \to G$ such that

$$\lambda_{i,r} g_{ij} \lambda_{j,s}^{-1} = h_{rs} \tag{33.2}$$

whenever this is well defined. The resulting quotient set is denoted by $\check{H}^1(B; G)$.[1]

Property 33.1.4 (Normalization). Let $\{g_{ij}\}_{i,j\in I}$ be a cocycle on B. It satisfies the following properties for all $x \in B$:

- $g_{ij}(x) = g_{ji}(x)^{-1}$, and
- $g_{ii}(x) = e$.

Definition 33.1.5 (Fibre bundle). A tuple (E, B, π, F, G) where E, B and F are spaces and G is a group, called the **structure group**, such that there exists a surjective morphism

$$\pi : E \to B$$

and an open cover $\{U_i\}_{i\in I}$ of B together with a family of isomorphisms $\{\varphi_i : \pi^{-1}(U_i) \to U_i \times F\}_{i\in I}$ that make the following diagram commute for all $i \in I$:

$$\begin{tikzcd} \pi^{-1}(U_i) \arrow[rr, "\varphi_i"] \arrow[dr, "\pi"'] & & U_i \times F \arrow[dl, "\mathrm{pr}_1"] \\ & U_i & \end{tikzcd}$$

As for general bundles, one calls E and B the **total space** and **base space**, respectively. The space F is called the **(typical) fibre**. The pair (U_i, φ_i) is sometimes called a **bundle chart** and the set $\{(U_i, \varphi_i)\}_{i\in I}$ is often called a **local trivialization**[2]. The cover $\{U_i\}_{i\in I}$ itself is called a **trivializing cover** of the bundle.

The **transition maps** $\varphi_j \circ \varphi_i^{-1} : (U_i \cap U_j) \times F \to (U_i \cap U_j) \times F$ can be identified with a cocycle $g_{ji} : U_i \cap U_j \to G$ as follows. The transition maps restrict to the identity on B and, hence, act only on the fibres:

$$\varphi_j \circ \varphi_i^{-1}(b, x) = \big(b, g_{ji}(b) \cdot x\big) . \tag{33.3}$$

[1]The notation stems from the fact that this is the first Čech cohomology group with values in G (Section 8.4.2).

[2]This terminology follows from the fact that the bundle is locally isomorphic to a (trivial) product space: $E \cong U \times F$.

The compatibility conditions satisfied by the functions g_{ji}, obtained by considering triple intersections, are exactly the cocycle conditions (33.1). Moreover, it can be shown that this action of G on every fibre is faithful (Definition 2.3.12).

Remark 33.1.6. One should pay attention to the fact that the bundle charts are not coordinate charts in the sense of manifolds (Definition 31.1.1), because the image of φ_i is not an open subset of $\mathbb{R}^n$. However, they serve the same purpose as they are used to locally describe the total space P.

Notation 33.1.7. A fibre bundle (E, B, π, F, G) is often denoted by $F \hookrightarrow E \xrightarrow{\pi} B$ or even by the morphism $\pi : E \to B$ if the fibre is not important. A drawback of such notations is that the structure group of the bundle is not immediately clear.

Definition 33.1.8 (Numerable fibre bundle). A fibre bundle that admits a local trivialization over a numerable open cover (Definition 6.5.11).

Definition 33.1.9 (Compatible bundle charts). A bundle chart (V, ψ) is said to be **admissible** or compatible with a trivializing cover $\{(U_i, \varphi_i)\}_{i\in I}$ if, whenever $V \cap U_i \neq \emptyset$, there exists a map $h_i : V \cap U_i \to G$ such that

$$\psi \circ \varphi_i^{-1}(b, x) = \big(b, h_i(b)x\big) \tag{33.4}$$

for all $b \in V \cap U_i$ and $x \in F$. Two trivializing covers are said to be equivalent if all bundle charts are mutually compatible. As in the case of manifolds, this gives rise to the notion of a **G-atlas**. A **G-bundle** is then defined as a fibre bundle equipped with an equivalence class of G-atlases.

Definition 33.1.10 (Bundle map). A bundle map between two fibre bundles $\pi_1 : E_1 \to B_1$ and $\pi_2 : E_2 \to B_2$ is a pair (f_E, f_B) of morphisms that make Diagram 33.1 commute. The map f_E is said to **cover** f_B. If such a couple exists, the base map f_B is uniquely determined by f_E and, therefore, a bundle map is often just denoted by $f_E : E_1 \to E_2$.

$$\begin{tikzcd} E_1 \arrow[r, "f_E"] \arrow[d, "\pi_1"'] & E_2 \arrow[d, "\pi_2"] \\ B_1 \arrow[r, "f_B"'] & B_2 \end{tikzcd}$$

Figure 33.1: Bundle map between fibre bundles.

Definition 33.1.11 (Equivalent fibre bundles). Two fibre bundles $\pi_1 : E_1 \to B$ and $\pi_2 : E_2 \to B$ with the same typical fibre and structure group are said to be equivalent if there exist trivializations $\{(U_i, \varphi_i)\}_{i\in I}$ and $\{(U_i, \varphi_i')\}_{i\in I}$ such that the associated cocycles are equivalent (note that the cover $\{U_i\}_{i\in I}$ is the same for both trivializations). An

explicit form of the functions λ is given by

$$\lambda_i := \varphi'_i \circ \varphi_i^{-1} . \tag{33.5}$$

Property 33.1.12 (Isomorphism). Two fibre bundles over the same base space are equivalent if and only if they are isomorphic.

Definition 33.1.13 (Trivial bundle). A fibre bundle (E, B, π, F) is said to be trivial if there exists an equivalence $E \cong B \times F$.

33.2 Constructions

Construction 33.2.1 (Construction theorem[3]). Let B and F be spaces and let G be a group equipped with a faithful (left) action on F. Suppose that a cover $\{U_i\}_{i\in I}$ of B and a collection of morphisms $\{g_{ji} : U_i \cap U_j \to G\}$ that satisfy the cocycle condition (33.1.3) are given. A fibre bundle over B can be constructed as follows:

1. Construct, for every set U_i, the Cartesian product $U_i \times F$.
2. Construct the disjoint union $\bigsqcup_{i\in I} U_i \times F$ and equip it with the disjoint union topology (Construction 6.3.2).
3. From this disjoint union, construct a quotient space, equipped with the quotient space topology (Construction 6.3.3), induced by the following equivalence relations for all $i, j \in I$:

$$(b, f) \sim \big(b, g_{ji}(b) \cdot f\big) , \tag{33.6}$$

   where $b \in U_i \cap U_j$ and $f \in F$ (note that the disjoint union indices are suppressed in this notation).

The fibre bundle E is equal to this quotient space, where $\pi : E \to B : [(b, f)] \mapsto b$ is the quotient space projection. Local trivializations are given by the maps $\varphi_i : \pi^{-1}(U_i) \to U_i \times F$ that satisfy

$$\varphi_i^{-1} : (b, f) \mapsto [(b, f)] . \tag{33.7}$$

Remark 33.2.2 (Topology). Although the resulting fibre bundle is, by construction, bijective (as a set) to the Cartesian product $B \times F$ or the disjoint union $\bigsqcup_{b\in B} F$, this does not hold on the level of topologies. It is not equipped with the disjoint union topology!

Property 33.2.3 (Homotopy invariance). Homotopic transition functions give rise to equivalent bundles.

[3]Sometimes also called the **clutching theorem**. See further on for an explanation.

Remark 33.2.4 (Clutching). The above construction is often called the **clutching construction**, especially when constructing vector bundles over a (hyper)sphere S^n. In that case, the covering consists of two hemispheres that intersect on the equator S^{n-1} and the function g_{21} is called the **clutching function**.

Construction 33.2.5 (Pullback bundle). Let $\pi : E \to B$ be a fibre bundle and let $f : B' \to B$ be a morphism of spaces. The pullback (Definition 3.4.66) of π and f gives the total space of the pullback bundle f^*E:

$$f^*E := \left\{(b', e) \in B' \times E \mid f(b') = \pi(e)\right\}. \tag{33.8}$$

The topology on f^*E is induced by the subspace topology of $B' \times E$. The projection onto the second factor gives a map of total spaces $f^*E \to E$.

Construction 33.2.6 (Fibre product). Let (E_1, B, π_1) and (E_2, B, π_2) be two fibre bundles over the same base space B. Their fibre product is defined as follows:

$$E_1 \times_B E_2 := \left\{(p, q) \in E_1 \times E_2 \mid \pi_1(p) = \pi_2(q)\right\}. \tag{33.9}$$

It is the pullback of one bundle along another.

33.3 Sections

Definition 33.3.1 (Section). A (**global**) section of a fibre bundle $\pi : E \to B$ is a morphism $s : B \to E$ such that $\pi \circ s = \mathbb{1}_B$, i.e. a section of π in the sense of Definition 3.4.1. For any open subset $U \subset B$, a **local** section is defined as a morphism $s_U : U \to E$ such that $\pi \circ s_U = \mathbb{1}_U$.

Notation 33.3.2. The set of all global sections of a bundle E is denoted by $\Gamma(E)$. The set of local sections over U is sometimes denoted by $\Gamma(U, E)$. With this latter notation, one also has $\Gamma(E) = \Gamma(B, E)$.

Property 33.3.3 (Pullback of sections). Consider a fibre bundle $\pi : E \to B$ together with a morphism of spaces $f : B' \to B$. The sections of E pullback to the pullback bundle f^*E by defining $f^*s := s \circ f$.

Chapter 34

Vector Bundles

The main reference for this chapter is Bott and Tu (1995).

Chapter Contents

34.1 Tangent bundle **608**

34.2 Vector bundles **610**

34.2.1 Sections 611

34.2.2 Tubular neighbourhoods 611

34.2.3 Sums and product 611

34.2.4 Associated vector bundles 612

34.3 Vector fields **613**

34.3.1 Integral curves 615

34.3.2 Lie derivative 616

34.3.3 Multivector fields 618

34.4 Differential k-forms **619**

34.4.1 Definition 619

34.4.2 Exterior derivative 620

34.4.3 Lie derivative 621

34.4.4 Interior product 622

34.4.5 Lie derivative of tensor fields 622

34.4.6 Vector-valued differential forms 623

34.5 Grassmann bundle **625**

34.6 Frobenius' theorem **627**

34.7 Linear connections **629**

34.7.1 Koszul connections 629

34.7.2 Induced connections 630
34.7.3 Curvature & torsion 632
34.8 Integration Theory **633**
34.8.1 Orientation and densities 633
34.8.2 Orientation in homology 635
34.8.3 Integration 636
34.8.4 Stokes' theorem 637
34.9 Cohomology **638**
34.9.1 de Rham complex 638
34.9.2 Integration 640
34.9.3 Cohomology with compact support 641
34.9.4 Čech–de Rham complex 645
34.9.5 Nonorientable manifolds 646
34.9.6 Generalized cohomology ♣ 648
34.10 Morse theory **648**
34.10.1 Morse functions 648
34.10.2 Morse–Bott functions 649
34.10.3 Morse homology 649
34.11 Differential operators **651**
34.11.1 Elliptic complexes 651

34.1 Tangent bundle

The tangent space, as introduced in Section 31.2, can also be introduced in a more abstract and general way. Because it is the most important example of a vector bundle, tangent bundles are introduced first.

Construction 34.1.1 (Tangent bundle). Let M be an n-dimensional manifold with atlas $\{(U_i, \varphi_i)\}_{i \leq n}$. Construct, for every open set U, an associated set $TU := U \times \mathbb{R}^n$ and construct, for every smooth function $f : U \to \mathbb{R}$, an associated smooth function on TU, called the **differential** or **derivative** of f, as follows:

$$Tf : U \times \mathbb{R}^n \to f(U) \times \mathbb{R}^n : (p, v) \mapsto \big(f(p), Df(p)v\big), \tag{34.1}$$

where $Df(p) : \mathbb{R}^n \to \mathbb{R}^n$ is the Jacobian of f at p.

By applying this definition to the transition functions ψ_{ji}, one obtains a new set of functions $T\psi_{ji} : TU_i \to TU_j$ given by

$$T\psi_{ji}(\varphi_i(p), v) := \left(\varphi_j(p), D(\varphi_j \circ \varphi_i^{-1})(\varphi_i(p))v\right). \tag{34.2}$$

Because the transition functions are diffeomorphisms, the associated Jacobians are invertible. This implies that the maps $T\psi_{ji}$ are represented by elements of $\mathsf{GL}(\mathbb{R}^n)$. The tangent bundle is then obtained by applying the Construction Theorem 33.2.1 to the triple $(M, \mathbb{R}^n, \mathsf{GL}(\mathbb{R}^n))$ together with the cover $\{U_i\}_{i\leq n}$ and the cocycle $\{T\psi_{ji}\}_{i,j\leq n}$.

Definition 34.1.2 (Natural chart). The charts in the atlas of this bundle are sometimes called natural charts or **adapted charts** because the first n coordinates are equal to the coordinates on the base manifold.

Definition 34.1.3 (Tangent space). Consider a point $p \in M$. The definition of the tangent space in the bundle setting is given by the fibre

$$T_pM := \pi_{TM}^{-1}(p). \tag{34.3}$$

If one uses the natural charts identify T_pM with the set $\varphi_i(p) \times \mathbb{R}^n$, it can be seen that T_pM is isomorphic to $\mathbb{R}^n$ (as a vector space).

Property 34.1.4 (Smooth structure). An atlas on TM is given by the charts (TU_i, θ) with

$$\theta : TU_i \to \mathbb{R}^{2n} : (p, X) \mapsto (\varphi_i \circ \pi(p), X^1, \dots, X^n), \tag{34.4}$$

where (U_i, φ_i) is a chart around $p \in M$ such that a tangent vector X is trivialized as $X^i\partial_i \in T_pM$.

Property 34.1.5 (Dimension). Let M be an n-dimensional manifold. Using the charts on M and the adapted charts on TM, one can see that TM is locally isomorphic to $\mathbb{R}^{2n}$. This implies that

$$\dim(TM) = 2\dim(M). \tag{34.5}$$

Remark 34.1.6 (Physics). Now, it should be clear that the statement "*a vector is something that transforms like a vector*", which one often hears in introductory physics courses, comes from the fact that

$$\text{a vector } v \in T_pM \text{ is tangent to } \varphi_i(p) \text{ in a chart } (U_i, \varphi_i)$$

if and only if

$$D(\varphi_j \circ \varphi_i^{-1})(\varphi_i(p))v \text{ is tangent to } \varphi_j(p) \text{ in a chart } (U_j, \varphi_j).$$

Definition 34.1.7 (Differential). The map T defined in Eq. (34.1) can be generalized to arbitrary smooth manifolds as the map $Tf : TM \to TN$, which is locally represented by the Jacobian. Furthermore, let $p \in U \subseteq M$ and let $V = f(U)$. By looking at the restriction of Tf to T_pM, one can see that it maps T_pU to $T_{f(p)}V$ linearly.

Property 34.1.8. The map $Tf : TM \to TN$ has the following properties:

- T preserves identities: $T\mathbb{1}_M = \mathbb{1}_{TM}$.
- Let f, g be two smooth functions on smooth manifolds, then $T(f \circ g) = Tf \circ Tg$.

This turns the map T into an endofunctor (Definition 3.2.4) on the category of smooth manifolds. One can view T as a 'functorial derivative'.

Definition 34.1.9 (Rank). Let $f : M \to N$ be a differentiable function between smooth manifolds. Using the fact that Tf is fibrewise linear, the rank of f at $p \in M$ is defined as the rank of the differential $Tf : T_pM \to T_{f(p)}N$ in the sense of Definition 22.2.12.

Theorem 34.1.10 (Inverse function theorem). *A smooth function $f : M \to N$ between smooth manifolds is a local diffeomorphism at $p \in M$ if and only if its differential $Tf : T_pM \to T_{f(p)}N$ is an isomorphism at p.*

Definition 34.1.11 (Parallelizable manifold). A manifold with a trivial tangent bundle.

34.2 Vector bundles

Instead of restricting the typical fibre to be a Euclidean space with the same dimension as the base manifold, one can generalize the construction of the tangent bundle in the following way.

Construction 34.2.1 (Vector bundle). Consider a (topological) manifold with an atlas $\{(U_i, \varphi_i)\}_{i\in I}$, a cocycle $\{g_{ji} : U_i \cap U_j \to G\}_{i,j\in I}$ with values in a group G and a representation $\rho : G \to \mathsf{GL}(V)$ on a (finite-dimensional) vector space V. This data can be used to construct a fibre bundle through Construction 33.2.1. The dimension of the typical fibre V is called the **rank** of the vector bundle.

Remark 34.2.2. As was also the case for tangent bundles, the choice of charts on E is not random. To preserve the linear structure of fibres, the use of the natural charts is imperative.

Example 34.2.3 (Line bundle). A vector bundle with a one-dimensional fibre. A common example in quantum mechanics is the $\mathbb{C}$-line bundle over some smooth manifold whose sections correspond to the *wave functions* of a given system. (See Section 59.3 on geometric quantization for more information.)

Property 34.2.4 (Vector bundles over a sphere). The Clutching Theorem 33.2.4 and the homotopy invariance imply that rank-p vector bundles over the sphere are determined by homotopy classes of functions $S^{n-1} \to \mathsf{GL}_p(\mathfrak{K})$, i.e. they are classified by the homotopy group $\pi_{n-1}(\mathsf{GL}_p(\mathfrak{K}))$.

34.2.1 Sections

Definition 34.2.5 (Frame). A frame of a vector bundle E is a tuple $(s_1, \dots, s_n)$ of sections such that $(s_1(b), \dots, s_n(b))$ is a basis for the fibre $\pi^{-1}(b)$ for all $b \in B$. By passing to local sections, one obtains a local frame.

Property 34.2.6 (Trivial bundles). A vector bundle is trivial if and only if it admits a global frame.

Theorem 34.2.7 (Serre & Swan). *The set of smooth sections of a smooth vector bundle over a smooth manifold M is a finitely generated projective $C^\infty(M)$-module. More generally, the set of sections of a vector bundle over a compact Hausdorff manifold B is a finitely generated projective $C(B)$-module.*

Definition 34.2.8 (Zero section). The section s_0 of a vector bundle $E \to B$ that assigns to every point $b \in B$ the zero vector of the associated fibre E_b. For every vector bundle $\pi : E \to B$, one can embed the base manifold B in the bundle E through the zero section $s_0 : B \to E$. The complement of the image of this section is often denoted by E_0.

34.2.2 Tubular neighbourhoods

Definition 34.2.9 (Normal bundle). Consider a smooth manifold M with a submanifold S and consider, for every point $p \in S$, the tangent spaces T_pS and T_pM. Since T_pS is a subspace of T_pM, one can construct the quotient space $N_pS := T_pM/T_pS$. The normal bundle of S in M is defined as the vector bundle with fibres N_pS.

Definition 34.2.10 (Tubular neighbourhood). Consider a smooth manifold M with an embedded submanifold S. A tubular neighbourhood of S in M is a vector bundle $\pi : E \to S$ such that (an open neighbourhood of the zero section of) E is diffeomorphic to an open neighbourhood of S in M.

Theorem 34.2.11 (Tubular neighbourhood theorem). *Every embedded submanifold admits a tubular neighbourhood, namely its normal bundle. Furthermore, all tubular neighbourhoods are diffeomorphic.*

Corollary 34.2.12 (NDR pairs). Consider a smooth manifold M with a submanifold S. The pair (M, S) is an NDR pair (Property 7.2.33). In particular, consider a smooth fibre bundle $\pi : E \to B$. If π admits a global section, one can embed B in E as a submanifold and, hence, the pair (E, B) is an NDR pair.

34.2.3 Sums and product

Definition 34.2.13 (Whitney sum). Consider two vector bundles E, E' with typical fibres V, V' over the same base space. One can construct a new vector bundle $E \oplus E'$ by taking the typical fibre to be the direct sum $V \oplus V'$, i.e. the fibre over p is given by

$V_p \oplus V'_p$. This operation is called the Whitney sum or **direct sum** of vector bundles.

The existence property for complements (Property 22.1.20) from linear algebra can be generalized in the following way.

Property 34.2.14. Let B be a paracompact Hausdorff manifold and let E be a vector bundle over B. Every vector subbundle F of E admits an orthogonal complement $F^\perp$.

Property 34.2.15. Let B be a compact Hausdorff manifold. Every vector bundle E over B admits a complementary vector bundle E^c such that $E \oplus E^c \cong B \times \mathbb{R}^n$ for some $n \in \mathbb{N}$.

Definition 34.2.16 (Stable isomorphism). Two vector bundles E, E' over a base space B are said to be stably isomorphic if there exist integers $m, n \in \mathbb{N}$ such that

$$E \oplus (B \times \mathbb{R}^m) \cong E' \oplus (B \times \mathbb{R}^n)\,. \tag{34.6}$$

Definition 34.2.17 (Tensor product). The tensor product $E \otimes E' \to B$ of two vector bundles $\pi : E \to B$ and $\pi' : E' \to B$ is given by the fibrewise tensor product.

Definition 34.2.18 (Exterior tensor product). The exterior tensor product $E \boxtimes E' \to B \times B'$ of two vector bundles $\pi : E \to B$ and $\pi' : E' \to B'$ is given by the tensor product of fibres over the Cartesian product $B \times B'$.

34.2.4 Associated vector bundles

Construction 34.2.19 (Associated vector bundle). Consider a representation

$$\rho : \mathsf{GL}(\mathbb{R}^n) \to \mathsf{GL}(\mathbb{R}^l)$$

together with the tangent bundle cocycle $\{t_{ji} := D(\psi_{ji}) \circ \varphi_i\}_{i,j \leq n}$. The composite

$$\rho \circ t_{ji} : U_i \cap U_j \xrightarrow{t_{ji}} \mathsf{GL}(\mathbb{R}^n) \xrightarrow{\rho} \mathsf{GL}(\mathbb{R}^l)$$

is again a cocycle and can be used to define a new vector bundle on M through the fibre bundle construction theorem. The vector bundle $\rho(TM)$ is called the associated (vector) bundle of the tangent bundle induced by ρ.

Example 34.2.20 (Contravariant vectors). By noting that the kth tensor power $\otimes^k$ induces a representation given by the tensor product of representations, one can construct the bundle of order-k (contravariant) tensors $\otimes^k(TM)$.

Example 34.2.21 (Cotangent bundle). Another useful construction is given by the contragredient representation $(\rho^T)^{-1} = (\rho^{-1})^T$. The vector bundle constructed this way, where the cocycle is given by $(t_{ji}^T)^{-1}$, is called the cotangent bundle on M and is denoted by T^*M. Elements of the fibres are called **covariant vectors** or **covectors**.

Notation 34.2.22. A combination of the cocycle t_{ji} and its dual $(t_{ji}^T)^{-1}$ can also be used to define the bundle of (k, l)-tensors on M. This bundle is denoted by $T^{(k,l)}M$.

Definition 34.2.23 (Twisted bundle). Given a vector bundle $\pi : E \to B$ and a line bundle $\psi : L \to B$, one calls the tensor product $E \otimes L$ the L-twisted version of E.

For every vector bundle, one can define a canonical line bundle.

Construction 34.2.24 (Determinant line bundle). Consider a rank-n vector bundle $\pi : E \to B$. The determinant map induces an associated line bundle $\det(\pi) : \bigwedge^n E \to B$, where the transition functions on the fibres are given by the determinant of the transition functions of E. Bundles twisted by a determinant line bundle $\det(E)$ are called **densitized bundles**.[1]

Example 34.2.25 (Canonical bundle). Consider a smooth manifold M. The canonical (line) bundle of M is given by $\det(T^*M) = \Omega^{\dim(M)}(M)$, the determinant line bundle of the cotangent bundle of M.

34.3 Vector fields

From here on, the theory will be specialized to the smooth setting, i.e. all manifolds and all morphisms will be assumed to be smooth (unless stated otherwise).

Definition 34.3.1 (Vector field). A section $s \in \Gamma(TM)$ of the tangent bundle. By the Serre–Swan theorem 34.2.7, the set of vector fields forms a $C^\infty(M)$-module.

Notation 34.3.2. The set of all vector fields on a manifold M is often denoted by $\mathfrak{X}(M)$.

Definition 34.3.3 (Index). Consider a vector field X on an n-dimensional manifold M and let $p \in M$ be an isolated zero of X. Because p is isolated, one can find a small $(n-1)$-sphere around p that does not contain any other zeroes of X. The index $\operatorname{ind}_X(p)$ of X at p is defined as the degree (Definition 7.2.38) of the function

$$f : S^{n-1} \to S^{n-1} : m \mapsto \frac{X(m)}{\|X(m)\|} . \tag{34.7}$$

Property 34.3.4 (Winding number). The winding number of a vector field X along a curve γ on which X does not vanish is equal to the sum of indices of zeroes of X lying inside γ.

Theorem 34.3.5 (Poincaré–Hopf). *Let M be a compact manifold and consider a vector field X having only isolated zeroes. The Euler characteristic (Definition 7.2.18) is given by*

$$\chi(M) = \sum_{p \in X^{-1}(0)} \operatorname{ind}_X(p) . \tag{34.8}$$

Remark 34.3.6. This is essentially a restatement of Property 31.3.20, where one submanifold is given by the zero section and the other is given by the graph of X.

[1] See also Remark 34.8.11.

An immediate consequence of the Poincaré–Hopf theorem is the following well-known result.

Theorem 34.3.7 (Hairy ball theorem). *There exists no nowhere-vanishing vector field on an even-dimensional sphere S^{2n}.*

Definition 34.3.8 (Pullback). Consider a diffeomorphism $\varphi : M \to N$ and let X be a vector field on N. The pullback of X along φ is defined as follows:

$$(\varphi^* X)_p := T\varphi^{-1}(X_{\varphi(p)}) \,. \tag{34.9}$$

Definition 34.3.9 (Pushforward). Let X be a vector field on M and let $\varphi : M \to N$ be a diffeomorphism. The pushforward of X along φ is defined as follows:

$$(\varphi_* X)_{\varphi(p)} := T\varphi(X_p) \,. \tag{34.10}$$

This can be rewritten using the pullback as follows:

$$\varphi_* X = (\varphi^{-1})^* X \,. \tag{34.11}$$

Equivalently, one can define a vector field on N as

$$(\varphi_* X)_q(f) := X_{\varphi^{-1}(q)}(f \circ \varphi) \,. \tag{34.12}$$

Example 34.3.10 (Projective space). Consider the manifold $\mathbb{CP}^n$ for some $n \in \mathbb{N}$. Local coordinate patches are given by charts

$$\mathbb{C}^n \to \pi_i : U_i : (x_1, \dots, x_n) \mapsto \left[x_1 : \dots : x_i : 1 : x_{i+1} : \dots : x_n\right] , \tag{34.13}$$

where $U_i := \{[x] \in \mathbb{CP}^n \mid x \in \mathbb{C}^{n+1}, x_i \neq 0\}$. To determine the vector fields induced by the basis vector fields ∂_i, however, it is more useful to start from the quotient map $\pi : \mathbb{C}^{n+1}/\{0\} \to \mathbb{CP}^n$. This map acts on the subset $\{x \in \mathbb{C}^{n+1} \mid x_i \neq 0\}$ as

$$(x_1, \dots, x_{n+1}) \mapsto \left[\frac{x_1}{x_i} : \dots : \frac{x_{i-1}}{x_i} : 1 : \frac{x_{i+1}}{x_i} : \dots : \frac{x_{n+1}}{x_i}\right] \tag{34.14}$$

to give the chart U_i. Locally on this subset, pushing the $n+1$ basis vector fields ∂_i forward along π gives

$$\begin{cases} \pi_* \partial_j = \dfrac{1}{x_i} \tilde{\partial}_j & \text{if } j \neq i \,, \\[2ex] \pi_* \partial_i = -\displaystyle\sum_{j \neq i} \frac{x_j}{x_i^2} \tilde{\partial}_j & \text{otherwise} \,. \end{cases} \tag{34.15}$$

The kernel of this projection is given by the Euler vector field

$$\mathbb{E} := x^i \partial_i \tag{34.16}$$

from Definition 16.5.13.

34.3.1 Integral curves

Definition 34.3.11 (Integral curve). Let $X \in \mathfrak{X}(M)$ and let $\gamma :]a, b[\to M$ be a curve on M. γ is called an integral curve of X if

$$\gamma'(t) = X(\gamma(t)) \tag{34.17}$$

for all $t \in]a, b[$, where $\gamma'(t) := T\gamma(t, 1)$.

This equation can be viewed as a system of ordinary differential equations. Using the Picard–Lindelöf existence theorem 20.2.1, together with the initial value condition $\gamma(0) = p$, one can find a unique maximal curve satisfying the above equation. This solution, denoted by γ_p, is called the **integral curve of** X **through** p.

Definition 34.3.12 (Flow). Let $X \in \mathfrak{X}(M)$ and consider its integral curve γ_p through a point $p \in M$. The function σ_t defined by

$$\sigma_t(p) := \gamma_p(t)\,, \tag{34.18}$$

is called the flow of X at time t. The **flow domain** is defined as the set

$$D(X) := \left\{(t, p) \in \mathbb{R} \times M \mid p \in M, t \in]a_p, b_p[, a_p < 0 < b_p\right\}, \tag{34.19}$$

where $]a_p, b_p[$ is the maximal interval on which γ_p is defined.

Property 34.3.13. Suppose that $D(X) = \mathbb{R} \times M$. The flow σ_t has the following properties for all $s, t \in \mathbb{R}$:

- σ_t is smooth,
- $\sigma_0 = \mathbb{1}_M$,
- $\sigma_{s+t} = \sigma_s \circ \sigma_t$, and as a consequence
- $\sigma_{-t} = (\sigma_t)^{-1}$.

These properties say that σ_t is a diffeomorphic group action of the additive group of real numbers on M. This implies that σ_t is a (smooth) flow in the general mathematical sense.

Definition 34.3.14 (Complete vector field). A vector field on a manifold M for which the flow domain for every flow is all of $\mathbb{R} \times M$.

Property 34.3.15. If the manifold M is compact, every vector field $X \in \mathfrak{X}(M)$ is complete.

Property 34.3.16 (Winding number). The winding number of a vector field along a closed integral curve is 1.

34.3.2 Lie derivative

Definition 34.3.17 (Lie derivative for smooth functions). Consider $X \in \mathfrak{X}(M)$ and $f \in C^\infty(M)$. The Lie derivative of f with respect to X at $p \in M$ is defined as

$$\mathcal{L}_X f(p) := \lim_{t\to 0} \frac{f\big(\gamma_p(t)\big) - f(p)}{t}\,. \tag{34.20}$$

The definition of the Lie derivative closely resembles the definition of the ordinary derivative on Euclidean space. This is not a coincidence.

Formula 34.3.18. Working out the definition of the Lie derivative and rewriting it as an operator equality gives

$$\mathcal{L}_X = \sum_k X^k \frac{\partial}{\partial x^k}\,. \tag{34.21}$$

It is clear that this is just the vector field X expressed in the basis of Definition 31.2.3. This way, one also recovers the behaviour of a tangent vector as a derivation. For smooth functions $f : M \to \mathbb{R}$, this gives

$$\mathcal{L}_X f(p) = X_p(f)\,. \tag{34.22}$$

Proof (Explanation). In this derivation, Landau's little-o notation is used:

$$\lim_{t\to 0} \frac{o(t)}{t} = 0. \tag{34.23}$$

Now, assume that X is a smooth vector field and f is a smooth function. Because the Lie derivative is a local operation one can work in a local chart such that γ is, again locally, equivalent to a curve[a] $\beta_p : U \to \mathbb{R}^n$ and such that one can expand $\beta_p(t)$ around $p \in U$:

$$\begin{aligned}
\mathcal{L}_X f(p) &= \lim_{t\to 0}\left[\frac{f(\beta_p(0) + t\beta_p'(0) + o(t)) - f(p)}{t}\right] \\
&= \lim_{t\to 0}\left[\frac{f(p + tX(p) + o(t)) - f(p)}{t}\right] \\
&= \lim_{t\to 0}\left[\frac{f(p) + tDf(p)\cdot X(p) + o(t) - f(p)}{t}\right] \\
&= \sum_k \frac{\partial f}{\partial x^k}(p) X_k(p) + \lim_{t\to 0}\frac{o(t)}{t} \\
&= \sum_k \frac{\partial f}{\partial x^k}(p) X_k(p)\,,
\end{aligned} \tag{34.24}$$

where Definition 34.3.11 for integral curves was used on the second line. If this

equation is rewritten as an operator equality, one obtains

$$\mathcal{L}_X = \sum_k X_k \frac{\partial}{\partial x^k} . \tag{34.25}$$

□

[a]The vector field $X(p) = (p, Y(p))$, where Y is a smooth vector field on $\mathbb{R}^n$, can also be identified with Y itself. This is implicitly done in the derivation by using the notation X for both vector fields.

Definition 34.3.19 (Lie derivative for vector fields). Let $X, Y \in \mathfrak{X}(M)$.

$$\mathcal{L}_X Y(p) := \frac{\mathrm{d}}{\mathrm{d}t}(\sigma_t^* X)(\gamma_p(t))\Big|_{t=0} . \tag{34.26}$$

Proof (Explanation). For vector fields one cannot just take the difference at two different points because the tangent spaces generally do not coincide. This can be resolved by using the flow (Definition 34.3.12):

$$\mathcal{L}_X Y = \lim_{t\to 0} \frac{(T\sigma_t)^{-1} X(\gamma_p(t)) - X(p)}{t} , \tag{34.27}$$

where $T\sigma_t$ is the differential (Definition 34.1.7) of the flow, which satisfies $(T\sigma)^{-1} = T\sigma_{-t}$. To see that this definition makes sense, one has to show that $(T\sigma_t)^{-1}[X(\gamma_p(t))] \in T_pM$. This goes as follows:

$$\begin{aligned}
(T\sigma_t)^{-1} X(\gamma_p(t))(f) &= T\sigma_{-t} X(\gamma_p(t))(f) \\
&= X(\sigma_{-t} \circ \gamma_p(t))(f \circ \sigma_{-t}) \\
&= X(\sigma_{-t} \circ \sigma_t(p))(f \circ \sigma_{-t}) \\
&= X(p)(f \circ \sigma_{-t}) \\
&\in T_pM
\end{aligned}$$

for all $f \in C^k(M, \mathbb{R})$. On the third line, the definition of the flow was used. One can also rewrite the second term in the numerator of Eq. (34.27) using the flow:

$$X(p) = X(\sigma_0(p)) = T\sigma_0(X) .$$

Using the definition of the pushforward of vector fields (Definition 34.3.9), the Lie derivative can be rewritten as follows:

$$\begin{aligned}
\mathcal{L}_X Y &= \lim_{t\to 0} \frac{\sigma_{-t*} X(\gamma_p(t)) - \sigma_{0*} X(\gamma_p(0))}{t} \\
&= \frac{\mathrm{d}}{\mathrm{d}t}(\sigma_{-t*} X)(\gamma_p(t))\Big|_{t=0} .
\end{aligned}$$

Finally, by using the relation (34.11) between pushforward and pullback, this becomes

$$\mathcal{L}_X Y = \frac{\mathrm{d}}{\mathrm{d}t}(\sigma_t^* X)(\gamma_p(t))\Big|_{t=0} . \tag{34.28}$$

□

Property 34.3.20 (Lie bracket). Let $X, Y \in \mathfrak{X}(M)$ be C^k-vector fields. The Lie derivative has the following properties:

- $\mathcal{L}_X Y$ is a vector field.
- The Lie derivative of vector fields coincides with the commutator:

$$\mathcal{L}_X Y = [X, Y] . \tag{34.29}$$

  The fact that this is indeed a derivation on $C^{k-1}(M, \mathbb{R})$ follows from Schwarz's theorem 16.5.11. This result shows that the Lie derivative on vector fields turns the space $\mathfrak{X}(M)$ into a (real) Lie algebra.

- The previous item also implies that the Lie derivative is antisymmetric:

$$\mathcal{L}_X Y = -\mathcal{L}_Y X . \tag{34.30}$$

Definition 34.3.21 (Holonomic basis). Consider a smooth manifold M and an open subset $U \subseteq M$. A local frame $\{e_i\}_{i \leq \dim(M)}$ for TU is said to be holonomic if all the Lie derivatives vanish on U:

$$\mathcal{L}_{e_i} e_j = 0 . \tag{34.31}$$

Equivalently, a basis is holonomic if the associated structure coefficients of the Lie algebra $\mathfrak{X}(M)$ vanish on U.

Property 34.3.22. For every holonomic basis, there exists a coordinate system on M such that the basis coincides with the coordinate-induced basis.

34.3.3 Multivector fields

Similar to the construction of *differential forms* in Section 34.4 below, one can also construct products of vector fields.

Definition 34.3.23 (Multivector field). A section of the exterior algebra $\Lambda^\bullet TM$.

Definition 34.3.24 (Schouten–Nijenhuis bracket). The unique extension of the Lie bracket to multivector fields that turns $\Gamma(\Lambda^\bullet TM)$ into a Gerstenhaber algebra (Definition 4.9.4).

A more explicit description is given as follows. Let $f \in C^\infty(M)$ and $\{X_i\}_{i \leq n} \in \mathfrak{X}(M)$ for some $n \in \mathbb{N}$.

$$[f, X_1 \wedge \cdots \wedge X_n] := -\iota_{\mathrm{d}f} X_1 \wedge \cdots X_n , \tag{34.32}$$

where $\mathrm{d}f$ is the *exterior derivative* (see Definition 34.4.7) and ι is defined analogous to Definition 34.4.15, and

$$[X_1 \wedge \cdots \wedge X_{i_k}, X_{i_{k+1}} \wedge \cdots \wedge X_n] := \sum_{i,j} [X_i, X_j] X_1 \wedge \cdots \wedge \hat{X}_i \wedge \cdots \wedge \hat{X}_j \wedge \cdots \wedge X_n \tag{34.33}$$

for all $k \leq n$, where the caret indicates that this vector field is omitted. The Schouten–Nijenhuis bracket of two smooth functions is defined to be identically zero.

34.4 Differential k-forms

34.4.1 Definition

Definition 34.4.1 (Differential form). A differential k-form is a map

$$\omega : T^k M \to \mathbb{R} \tag{34.34}$$

such that the restriction of ω to each fibre of the bundle $T^k M$ is multilinear and anti-symmetric. The space of all differential k-forms on a manifold M is denoted by $\Omega^k(M)$. Just like $\mathfrak{X}(M)$, it forms a $C^\infty(M)$-module.

Differential forms can also be constructed as sections of an associated vector bundle.

Alternative Definition 34.4.2. Consider the representation

$$\rho_k : \mathsf{GL}(\mathbb{R}^{m*}) \to \mathsf{GL}(\Lambda^k \mathbb{R}^{m*}) : A \mapsto A \wedge \cdots \wedge A.$$

This representation induces an associated vector bundle $\rho_k(\pi_{T^*M})$ of the cotangent bundle on M. A differential k-form is given by a section of $\rho_k(\pi_{T^*M})$:

$$\Omega^k(M) := \Gamma(\rho_k(\pi_{T^*M}))\,. \tag{34.35}$$

Construction 34.4.3 (Exterior algebra). One can construct a Grassmann algebra (Definition 23.4.21) by equipping the graded vector space

$$\Omega^\bullet(M) := \bigoplus_{k \geq 0} \Omega^k(M) \tag{34.36}$$

with the wedge product of differential forms that is induced by the wedge product on $\Lambda^k \mathbb{R}^m$. This graded algebra is associative, graded-commutative and unital with the constant function $1 \in C^\infty(M)$ as the identity element.

Definition 34.4.4 (Pullback). Let $f : M \to N$ be a smooth function between manifolds and let ω be a differential k-form on N. The pullback of ω by f is defined as

$$f^* \omega := \omega \circ f_*\,. \tag{34.37}$$

This defines a map $f^* : \Omega^\bullet(N) \to \Omega^\bullet(M)$.

Definition 34.4.5 (Pushforward). Let $f : M \to N$ be a diffeomorphism between manifolds and let ω be a differential k-form on M. The pushforward of ω by f is defined as

$$f_*\omega := \omega \circ (f^{-1})_* \,. \tag{34.38}$$

Remark. Note that the pushforward of differential k-forms is only defined for diffeomorphisms, in contrast to pullbacks which only require smooth functions. This also explains why differential forms are the most valuable elements in differential geometry. (Vector fields cannot even be pulled back by general smooth maps.)

Formula 34.4.6 (Dual basis). Consider the coordinate basis from Definition 31.2.3 for the tangent space T_pM. From this set, one can construct a natural dual basis for the cotangent space T_p^*M using the natural pairing:

$$\left\langle \frac{\partial}{\partial x^i}, \mathrm{d}x^j \right\rangle = \delta_i^j \,. \tag{34.39}$$

It should be noted that $\mathrm{d}x^i$ is not just a notation. In the next section, it will be shown that these basis vectors can be obtained by applying the *exterior derivative* to the coordinate functions x^i.

34.4.2 Exterior derivative

Definition 34.4.7 (Exterior derivative). The exterior derivative d_k is a morphism constructed on the graded algebra of differential k-forms:

$$\mathrm{d}_k : \Omega^k(M) \to \Omega^{k+1}(M) \,. \tag{34.40}$$

For $k = 0$, it is defined by

$$\mathrm{d}f := \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \mathrm{d}x_i \tag{34.41}$$

The object $\mathrm{d}f \in \Omega^1(M)$ is often called the **differential** of f. This formula can be generalized to higher degree forms as follows:

$$\mathrm{d}(f \mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_k}) := \mathrm{d}f \wedge \mathrm{d}x_{i_1} \wedge \cdots \wedge \mathrm{d}x_{i_k} \,. \tag{34.42}$$

Remark 34.4.8. Equation (34.41) should be compared with the (informal) formula for the differential of a function that is often used in physics. The main difference is that, here, the quantities $\mathrm{d}x^i$ are not infinitesimal quantities but vectors of unit norm.

Property 34.4.9. The exterior derivatives satisfy the following properties for all $k \in \mathbb{N}$:

- **Nilpotency**: For all $\omega \in \Omega^k(M)$:

$$\mathrm{d}_k \circ \mathrm{d}_{k+1} = 0 \,. \tag{34.43}$$

- **Linearity**: d_k is $\mathbb{R}$-linear.

These two items say that $(\Omega^\bullet(M), \mathsf{d})$ is not just a graded algebra, but in fact a dg-algebra (Definition 5.1.9).

- **Graded Leibniz rule**: (hence d is a graded derivation):

$$\mathsf{d}(\omega_1 \wedge \omega_2) = \mathsf{d}\omega_1 \wedge \omega_2 + (-1)^j \omega_1 \wedge \mathsf{d}\omega_2 \,, \tag{34.44}$$

  where $\omega_1 \in \Omega^j(M)$ and $\omega_2 \in \Omega^k(M)$.

- **Naturality**: If $f \in C^\infty(M)$, then $f^*(\mathsf{d}\omega) = \mathsf{d}(f^*\omega)$.

Formula 34.4.10. Let $f \in C^\infty(M, \mathbb{R})$ and let γ be a curve on M. From Formula 34.4.6 of the basis $\{\mathsf{d}x_k\}_{k \leq n}$, one obtains the following result:

$$\langle \mathsf{d}f(x), \gamma'(t) \rangle = \sum_k \frac{\partial f}{\partial x_k}(x)\gamma_k'(t) = (f \circ \gamma')(t) \,. \tag{34.45}$$

Formula 34.4.11. An explicit formula for the exterior derivative of a k-form ω is

$$\begin{aligned} \mathsf{d}\omega(X_1, \dots, X_{k+1}) = &\sum_{i=1}^{k+1} (-1)^{i+1} X_i\big(\omega(X_1, \dots, \hat{X}_i, \dots, X_{k+1})\big) \\ &+ \sum_{i<j} (-1)^{i+j} \omega\big([X_i, X_j], X_1, \dots, \hat{X}_i, \dots, \hat{X}_j, \dots, X_{k+1}\big) \,, \end{aligned} \tag{34.46}$$

where $\hat{X}$ indicates that this argument is omitted.

34.4.3 Lie derivative

Formula 34.4.12 (Lie derivative of smooth functions). Using Eq. (34.41) for the exterior derivative of smooth functions and Formula 34.4.6 for the dual basis, one can rewrite the Lie derivative 34.3.17 as follows:

$$Xf(p) = \mathsf{d}f_p(X(p)) \,. \tag{34.47}$$

Formula 34.4.13 (Lie derivative of differential forms).

$$\mathcal{L}_X \omega(p) := \lim_{t \to 0} \frac{\sigma_t^* \omega - \omega}{t}(p) \tag{34.48}$$

Property 34.4.14. The Lie derivative has the following Leibniz-type property with respect to differential forms:

$$\mathcal{L}_X(\omega(Y)) = (\mathcal{L}_X \omega)(Y) + \omega(\mathcal{L}_X Y) \,, \tag{34.49}$$

where X, Y are two vector fields and ω is a one-form.

34.4.4 Interior product

Definition 34.4.15 (Interior product). Aside from the exterior derivative, one can also define another operation on the algebra of differential forms:

$$\iota_X : (\iota_X \omega)(v_1, \dots, v_{k-1}) \mapsto \omega(X, v_1, \dots, v_{k-1}) . \tag{34.50}$$

This antiderivation of degree -1 is called the **interior product** (or **interior derivative**). It can be seen as a generalization of the contraction map 23.3.15.

Notation 34.4.16. In certain situations, the above notation might become cumbersome. For this reason, the notation $X \lrcorner \omega$ is frequently used.

Property 34.4.17 (Pullback). Let $f : M \to N$ be a diffeomorphism.

$$f^*(\iota_X \omega) = \iota_{f^* X} f^* \omega \tag{34.51}$$

Formula 34.4.18 (Cartan's magic formula[2]). Let X be a vector field and let ω be a differential k-form. The Lie derivative of ω along X is given by the following formula:[3]

$$\mathcal{L}_X = \iota_X \circ \mathrm{d} + \mathrm{d} \circ \iota_X = \{\iota_X, \mathrm{d}\}_+ . \tag{34.52}$$

Combining the above two properties also shows that Lie derivatives pullback under diffeomorphisms.

Property 34.4.19 (Pullback). Let $f : M \to N$ be a diffeomorphism.

$$f^*(\mathcal{L}_X \omega) = \mathcal{L}_{f^* X} f^* \omega \tag{34.53}$$

34.4.5 Lie derivative of tensor fields

Formula 34.4.20 (Lie derivative of tensor fields). By comparing Definition 34.3.19 of the Lie derivative of vector fields to that of differential forms (Formula 34.4.13), one can see that both definitions are identical upon replacing X by ω. This leads to the following definition of the Lie derivative of a general tensor field $\mathcal{T} \in \Gamma(T^{(k,l)}M)$:

$$\mathcal{L}_X \mathcal{T}(p) := \frac{\mathrm{d}}{\mathrm{d}t} \sigma_t^* \mathcal{T}(\gamma_p(t)) \bigg|_{t=0} . \tag{34.54}$$

Alternative Definition 34.4.21 (Lie derivative of tensor fields). The Lie derivative of tensor fields can also be defined as the unique differential operator satisfying the following axioms:

1. $\mathcal{L}_X$ coincides with X on $C^\infty(M)$.

[2] Sometimes called **Cartan's (infinitesimal) homotopy formula**.
[3] The last step could be rewritten as a commutator in the algebra of graded derivations.

2. $\mathcal{L}_X$ satisfies the Leibniz rule with respect to tensor products.
3. $\mathcal{L}_X$ satisfies the Leibniz rule with respect to the contraction of forms and vector fields.
4. $\mathcal{L}_X$ commutes with the exterior derivative.

Property 34.4.22 (Derivations). Every derivation D of the tensor algebra can be decomposed as

$$D = \mathcal{L}_X + S \tag{34.55}$$

for some vector field X and some endomorphism S.

34.4.6 Vector-valued differential forms

Definition 34.4.23 (Vector-valued form). Consider a vector space V and let $E \to M$ be a vector bundle with typical fibre V. A V-valued k-form on M is a map $\omega : \Gamma(T^kM) \to V$. More generally, an **$E$-valued differential form** is a section of the associated bundle

$$\Omega^k(M; E) := \Gamma(E \otimes \Lambda^k T^*M)\,. \tag{34.56}$$

Construction 34.4.24 (Wedge product). Let $\omega \in \Omega^p(M; E_1)$ and $\nu \in \Omega^q(M; E_2)$. The wedge product $\omega \wedge \nu \in \Omega^{p+q}(M; E_1 \otimes E_2)$ of these differential forms is defined as follows:

$$\omega \wedge \nu(v_1, \dots, v_{p+q}) := \frac{1}{p!q!} \sum_{\sigma \in S_{p+q}} \mathrm{sgn}(\sigma)\omega(v_{\sigma(1)}, \dots, v_{\sigma(p)}) \otimes \nu(v_{\sigma(p+1)}, \dots, v_{\sigma(q)})\,. \tag{34.57}$$

This is a direct generalization of the formula for the wedge product of ordinary differential forms, where the scalar product (product on the algebra $\mathbb{R}$) is replaced by the tensor product (product on the tensor algebra). It should be noted that the result of this operation is not a section of any of the original bundles E_1 or E_2, but rather of the tensor product bundle $E_1 \otimes E_2$.

At this point, the reason why the pushforward (Definition 34.3.9) is also sometimes called the differential (as in Definition 34.1.7) can be given.

Property 34.4.25 (Differential vs. pushforward). It can be shown that for any vector bundle E and any manifold M, the space of sections of the homomorphism bundle $\mathrm{Hom}(TM, E)$ is isomorphic to the space of E-valued differential forms $\Omega^1(M; E)$. Now, consider a smooth function $f : M \to N$. Its pushforward is a map $f_* : TM \to TN$. Locally, the corresponding differential form is given by

$$\mathrm{d}f := \mathrm{d}f^i \otimes \partial_i\,, \tag{34.58}$$

where $(f^1, \dots, f^{\dim(N)})$ is a local expression for f and $\{\partial_1, \dots, \partial_{\dim(N)}\}$ is a local frame for TN. It is straightforward to show that acting with this differential form on a vector field in $\mathfrak{X}(M)$ gives the same result as acting with the pushforward f_*.

Construction 34.4.26 (Exterior derivative). The definition of an exterior derivative on E-valued differential forms is more involved than in the case of ordinary forms. The naive thing to do would be defining a derivative through the Leibniz formula. However, without further structure on E there is no natural way of differentiating sections of E.

If E is *flat*, i.e. if its transition functions are locally constant, one can choose a frame of sections $e^i : U \to E|_U$ induced by the trivializing maps $E|_U \to U \times \mathbb{R}^n$. Locally, one can then express any E-valued differential form as $\omega|_U = \sum_i \omega_i \otimes e^i$, where the ω_i are ordinary differential forms. After defining $\mathsf{d}e^i := 0$, one can again construct an exterior derivative through the Leibniz formula.

Remark 34.4.27. It should be noted that the definition of d depends on the choice of trivialization since the sections e^i depend on this choice.

Definition 34.4.28 (Lie algebra-valued form). A vector-valued differential form where the vector space V is equipped with a Lie algebra structure.

Formula 34.4.29 (Wedge product). Consider Lie algebra-valued differential forms $\omega \in \Omega^p(M;\mathfrak{g})$ and $\nu \in \Omega^q(M;\mathfrak{g})$. The wedge product of these differential forms is defined as follows:

$$[\omega \wedge \nu](v_1, \dots, v_{p+q}) := \frac{1}{p!q!} \sum_{\sigma \in S_{p+q}} \mathsf{sgn}(\sigma) \left[\omega(v_{\sigma(1)}, \dots, v_{\sigma(p)}), \nu(v_{\sigma(p+1)}, \dots, v_{\sigma(q)})\right], \quad (34.59)$$

where $[\cdot,\cdot]$ denotes the Lie bracket on $\mathfrak{g}$.

Formula 34.4.30. Let $\{e_a\}_{a \leq \mathsf{dim}(\mathfrak{g})}$ be a basis for the Lie algebra $\mathfrak{g}$. One can write any two Lie algebra-valued differential forms as $\phi = \phi^a \otimes e_a$ and $\psi = \psi^b \otimes e_b$, where ϕ^a and ψ^b are ordinary differential forms. The above formula for the wedge product can now be rewritten more elegantly as

$$[\phi \wedge \psi] = (\phi^a \wedge \psi^b) \otimes [e_a, e_b], \quad (34.60)$$

where $\wedge$ is the wedge product on $\Omega^\bullet(M)$.

Corollary 34.4.31 (Graded algebra). Using the above formula, it is easy to verify a number of properties similar to the ones of ordinary differential forms. As an example, the analogue of the graded-commutativity property on $\Omega^\bullet(M)$ is given:

$$[\phi \wedge \psi] = (-1)^{pq+1}[\psi \wedge \phi], \quad (34.61)$$

where $\phi \in \Omega^p(M;\mathfrak{g})$ and $\psi \in \Omega^q(M;\mathfrak{g})$. Here, the extra factor -1 comes from the antisymmetry of the Lie bracket.

Analogously, one can prove that the Lie algebra-valued wedge product satisfies a graded Jacobi-type identity:

$$(-1)^{pr}[\phi \wedge [\psi \wedge \theta]] + (-1)^{pq}[\psi \wedge [\theta \wedge \phi]] + (-1)^{qr}[\theta \wedge [\phi \wedge \psi]] = 0, \quad (34.62)$$

where $\theta \in \Omega^r(M;\mathfrak{g})$.

Definition 34.4.32 (Frölicher–Nijenhuis bracket). The interior product can also be generalized to vector-valued differential forms. Let $\omega \in \Omega^l(M)$ and $\phi \in \Omega^k(M;TM)$.

$$\iota_\phi \omega(X_1,\dots,X_{k+l-1}) := \frac{1}{p!q!}\sum_{\sigma\in S_{p+q}} \mathrm{sgn}(\sigma)\omega\big(\phi(X_{\sigma(1)},\dots,X_{\sigma(k)}),X_{\sigma(k+1)},\dots,X_{\sigma(k+l-1)}\big) \tag{34.63}$$

for all $X_1,\dots,X_{k+l-1} \in \mathfrak{X}(M)$. This defines a graded derivation on $\Omega^\bullet(M)$ of degree $k-1$. Mimicking Cartan's magic formula 34.4.18, one defines the **Nijenhuis–Lie derivative** as

$$\mathcal{L}_\phi := [\iota_\phi, \mathrm{d}]\,. \tag{34.64}$$

The Frölicher–Nijenhuis bracket is the unique operation

$$[\cdot,\cdot]_{\mathrm{FN}} : \Omega^k(M;TM)\otimes\Omega^l(M;TM) \to \Omega^{k+l}(M;TM) \tag{34.65}$$

such that

$$\mathcal{L}_{[\phi,\psi]_{\mathrm{FN}}} = [\mathcal{L}_\phi, \mathcal{L}_\psi]\,. \tag{34.66}$$

A related object, the **Nijenhuis–Richardson bracket** is obtained by setting

$$[\phi,\psi]_{\mathrm{NR}} := \iota_\phi\circ\psi - (-1)^{\deg(\phi)\deg(\psi)}\iota_\psi\phi\,. \tag{34.67}$$

34.5 Grassmann bundle

Looking at Property 22.8.2 and noting that $\mathsf{GL}(n,\mathbb{R})$ is a Lie group, it is clear that one can endow the Grassmannian $\mathsf{Gr}(k,\mathbb{R}^n)$ from Definition 22.8.1 with a differentiable structure, thereby turning it into a smooth manifold. This allows for the construction of a new bundle. However, because the Grassmannian is not a vector space, the resulting bundle will only be a general fibre bundle and not a vector bundle.

Construction 34.5.1 (Grassmann bundle). Define the set of transition functions

$$\psi_{ji} : (\varphi_i(p), V) \mapsto \big(\varphi_j(p), t_{ji}(p)\cdot V\big)\,, \tag{34.68}$$

where $\{t_{ji}\}_{i,j\le n}$ is the tangent bundle cocycle. These transition functions can be used to create a new fibre bundle with typical fibre $\mathsf{Gr}(k,\mathbb{R}^n)$. The fibre over a point $p \in M$ is the Grassmannian $\mathsf{Gr}(k,T_pM)$ associated to the tangent space over p.

By replacing the tangent bundle TM by an arbitrary vector bundle E (and accordingly replacing the cocycle t with the cocycle of E), one can define the Grassmann bundle of a general vector bundle.

Notation 34.5.2. The Grassmann k-plane bundle of a vector bundle E is denoted by $\mathsf{Gr}(k, E)$.

Definition 34.5.3 (Tautological bundle). Consider the Grassmannian $\mathsf{Gr}(n, V)$ of an $(n+k)$-dimensional vector space V. The total space of the tautological k-bundle $\gamma_{n,k}$ is defined as the set of points (W, w) where $W \in \mathsf{Gr}(n, V)$ and $w \in W$. Local trivializations are constructed as follows:

$$\varphi_Z : \pi^{-1}(U) \to \mathsf{Gr}(n, V) \times Z : (W, w) \mapsto (W, \mathsf{pr}_Z(w)), \tag{34.69}$$

where pr_Z is the orthogonal projection onto the subspace $Z \in \mathsf{Gr}(n, V)$. This bundle inherits a natural vector bundle structure from V.

Definition 34.5.4 (Twisting sheaf). Consider the tautological line bundle J over a projective space $\mathbb{CP}^n \cong \mathsf{Gr}(1, \mathbb{C}^n)$. The dual line bundle $\mathsf{Hom}(J, \mathbb{C})$ is often denoted by $\mathcal{O}_{\mathbb{CP}^n}(1)$ or $\mathcal{O}(1)$, and is sometimes called **Serre's twisting sheaf**. Tensor powers of this bundle are accordingly denoted by $\mathcal{O}_{\mathbb{CP}^n}(k)$. To also allow for factors of J, one can extend the notation to negative indices: $\mathcal{O}_{\mathbb{CP}^n}(-k)$. The tautological bundle is then denoted by $\mathcal{O}_{\mathbb{CP}^n}(-1)$.

Property 34.5.5 (Euler sequence). The twisting sheaf fits in a short exact sequence:

$$0 \longrightarrow \mathcal{O}_{\mathbb{CP}^n} \longrightarrow \mathcal{O}_{\mathbb{CP}^n}(1)^{\oplus(n+1)} \longrightarrow T\mathbb{CP}^n \longrightarrow 0\,. \tag{34.70}$$

By dualizing, one can obtain an exact sequence in terms of the de Rham complex.

Proof. Note that vector fields on $\mathbb{CP}^n$ are, locally, obtained by pairing (coordinate-induced) basis vectors on $\mathbb{C}^{n+1}$ with linear functions on $\mathbb{C}^{n+1}$, i.e. with sections of $\mathcal{O}_{\mathbb{CP}^n}(1)$. This leads to the central term $\mathcal{O}_{\mathbb{CP}^n}(1) \otimes \mathbb{C}^{n+1} \cong \mathcal{O}_{\mathbb{CP}^n}(1)^{\oplus(n+1)}$. When projecting these vector fields down to $T\mathbb{CP}^n$, the kernel is generated by the Euler vector field (Example 34.3.10). Not only scalar multiples are sent to 0, all multiples by functions that descend to well-defined functions on $\mathbb{CP}^n$ will vanish. Hence, the kernel is isomorphic to the structure sheaf $\mathcal{O}_{\mathbb{CP}^n}$.

A more invariant proof is given as follows. For every one-dimensional space $L \in \mathbb{CP}^n$, the tangent space $T_L\mathbb{CP}^n$ is isomorphic to $\mathsf{Hom}(L, \mathbb{C}^{n+1}/L)$ since, by linearity, a variation of the space is determined by the image of a single point and moving this point along the space L itself does not change anything. By basic linear algebra, one has $\mathsf{Hom}(L, \mathbb{C}^{n+1}/L) \cong L^* \otimes \mathbb{C}^{n+1}/L$ or $\mathbb{C}^{n+1}/L \cong \mathsf{Hom}(L, \mathbb{C}^{n+1}/L) \otimes L$. This gives the exact sequence

$$0 \longrightarrow L \longrightarrow \mathbb{C}^{n+1} \longrightarrow T_L\mathbb{CP}^n \otimes L \longrightarrow 0\,.$$

After globalizing one obtains

$$0 \longrightarrow \mathcal{O}_{\mathbb{CP}^n}(-1) \longrightarrow \mathcal{O}_{\mathbb{CP}^n}^{\oplus(n+1)} \longrightarrow T\mathbb{CP}^n \otimes \mathcal{O}_{\mathbb{CP}^n}(-1) \longrightarrow 0\,.$$

After tensoring with the twisting sheaf, one obtains the Euler sequence. □

34.6 Frobenius' theorem

Definition 34.6.1 (Distribution). A section of the Grassmann k-plane bundle (Construction 34.5.1).

Definition 34.6.2 (Foliation). Let M be an n-dimensional manifold. A (regular) foliation of M is a decomposition

$$M = \bigsqcup_{j \in I} V_j \tag{34.71}$$

by connected (immersed) subsets $V_j \subset M$, called the **leaves** of the foliation, such that there exists a cover $\{(U_i, \varphi_i)\}_{i \in I}$ of M and an integer $p \leq n$, called the **dimension**[4] of the foliation, such that for every $i \in I$ and $j \in J$ the intersection $V_j \cap U_i$ is given by the following system of equations:

$$\begin{cases} \phi_i^{p+1}(x) & = c_1 \,, \\ & \vdots \\ \phi_i^{n}(x) & = c_{n-p} \,, \end{cases} \tag{34.72}$$

where $c_1, \dots, c_{n-p} \in \mathbb{R}$ are constants.

Definition 34.6.3 (Integrable distribution). Let M be a manifold and consider a distribution of k-planes $W \in \Gamma(\mathrm{Gr}(k, TM))$. A submanifold $N \subseteq M$ is said to integrate W (or to be **integral**) with initial condition $p_0 \in M$ if $p_0 \in N$ and if $\forall p \in N : W(p) = T_p N$. W is said to be integrable if there exists such a submanifold N for every $p_0 \in M$.

Property 34.6.4. If a distribution on M is integrable, the integral submanifolds form a (regular) foliation. Moreover, every (regular) foliation gives rise to an integrable distribution by taking $W(p) := T_p N$ for N the integral submanifold passing through the point $p \in M$.

Definition 34.6.5 (Frobenius' integrability condition). A distribution W on a manifold M is said to satisfy Frobenius' integrability condition on an open set $U \subseteq M$ if it is **involutive**, i.e., for every two vector fields X, Y defined on U, such that $X(p) \in W(p)$ and $Y(p) \in W(p)$ for all $p \in U$, the Lie bracket $[X, Y](p)$ is also an element of $W(p)$ for all $p \in U$.

Theorem 34.6.6 (Frobenius' integrability theorem). *Let W be a distribution over a manifold M. W is integrable if and only if it satisfies Frobenius' integrability condition.*

This theorem also admits a formulation in terms of differential forms.

[4] If this integer depends on the leaf, the foliation is said to be **singular**.

Property 34.6.7 (Differential formulation). Consider a rank-r distribution D. The annihilator $I(D)$ is the ideal containing all differential forms satisfying

$$\omega(X_1, \dots, X_k) = 0 \tag{34.73}$$

whenever all $X_i \in D$. Frobenius' theorem says that the following conditions are equivalent:

- D is integrable.
- $I(D)$ is a differential ideal (Definition 5.1.14).

Property 34.6.8 (Pfaffian system). The fact that D is a distribution implies that, locally, there exists a set $\{\theta^\alpha\}$ of $\dim(M) - r$ linearly independent, annihilating one-forms that generate $I(D)$. These one-forms are said to form a Pfaffian system. Furthermore, every ideal that is locally generated by linearly independent one-forms defines a smooth distribution. The Pfaffian system is **integrable** if the differential ideal spanned by these one-forms is generated by exact one-forms:

$$\theta^\alpha = \sum_{i=1}^{\dim(M)-r} f_i(x) \mathrm{d}g_i^\alpha(x) \tag{34.74}$$

for (locally defined) smooth functions $f_i, g_i \in C^\infty$. The functions f_i are called **integrating factors**.

Given a Pfaffian system $\{\theta^\alpha\}_{\alpha \leq \dim(M)-r}$, a last equivalent condition for (**complete**) integrability is that

$$\mathrm{d}\theta^i \wedge \theta^2 \wedge \cdots = 0 \tag{34.75}$$

for all $i \leq \dim(M) - r$.

Property 34.6.9 (Integral manifold). Consider a Pfaffian system $\mathcal{P} \equiv \{\theta^\alpha\}_{\alpha \leq \dim(M)-r}$. An integral manifold of $\mathcal{P}$ is a smooth function $f : N \to M$ such that

$$f^* \theta^\alpha = 0 \tag{34.76}$$

for all $1 \leq \alpha \leq \dim(M) - r$. Often, f will be a submanifold inclusion.

If the system is integrable, there exist $\dim(M) - r$ independent functions y^α such that $\{\mathrm{d}y^\alpha\}_{\alpha \leq \dim(M)-r}$ generates the same differential ideal as the Pfaffian system. The integral submanifolds of the distribution are given by the systems of equations

$$\begin{cases} y^1 = c^1, \\ y^2 = c^2, \\ \quad \vdots \end{cases} \tag{34.77}$$

for constants c^α.

Definition 34.6.10 (Characteristic system). Consider a system $\mathcal{P} \equiv \{\omega^\alpha = 0\}_{\alpha\in I}$ of exterior differential equations, where the degree of all forms is nonzero. As for a smooth distribution, one can locally find a set of one-forms that generates the same ideal. This is again called a **Pfaffian system**. The characteristic system of $\mathcal{P}$ is defined as the Pfaffian system of the differential closure of $\mathcal{P}$.

If $\mathcal{P}$ also contains functions, i.e. degree-zero forms, the characteristic system is that of $\{\omega^\alpha, \mathsf{d}f^\beta\}_{\alpha,\beta\in I}$ extended by the functions f^β themselves.

Definition 34.6.11 (Integral invariant). Consider a Pfaffian system $\mathcal{P} \equiv \{\theta^\alpha\}_{\alpha\in I}$. A differential form ω is called an invariant of $\mathcal{P}$ if

$$\mathcal{L}_X\omega = 0 \tag{34.78}$$

whenever $\theta(X) = 0$ for al $\theta \in \mathcal{P}$. These forms are also called **absolute integral invariants** because the integral of ω over every chain is invariant under the transformations generated by these X. If $\mathsf{d}\varphi$ is an (absolute integral) invariant of $\mathcal{P}$, then φ is called a **relative integral invariant**.

34.7 Linear connections

34.7.1 Koszul connections

Definition 34.7.1 (Koszul connection). Let $\pi : E \to M$ be a vector bundle over a manifold M. A Koszul connection or **linear connection** on E is a (smooth) linear map $\nabla : \Gamma(E) \to \Gamma(T^*M \otimes E)$ satisfying the Leibniz property

$$\nabla(f\sigma) = f\nabla\sigma + \mathsf{d}f \otimes \sigma \tag{34.79}$$

for all $f \in C^\infty(M)$. The evaluation ∇_X on a vector field X is often called a **covariant derivative** (with respect to X).

Property 34.7.2 (Linearity). It is easily seen that the connection acts linearly with respect to vector fields:

$$\nabla_{fX+Y}\sigma = f\nabla_X\sigma + \nabla_Y\sigma\,. \tag{34.80}$$

Property 34.7.3 (Local behaviour). Consider a vector $v \in T_pM$. If two sections $\sigma, \rho \in \Gamma(E)$ coincide on some neighbourhood U of p, then $\nabla_v\sigma = \nabla_v\rho$ at p. Furthermore, given a curve $\gamma : [0,1] \to M$ and two sections $\sigma, \rho \in \Gamma(E)$ such that $\sigma\circ\gamma = \sigma\circ\gamma$, one finds that $\nabla_{\dot\gamma}\sigma = \nabla_{\dot\gamma}\rho$. This implies that a Koszul connection only depends on the local behaviour of the given sections.

Remark 34.7.4. The above property shows the major difference between the Lie derivative and the covariant derivative when acting on sections σ of a tensor bundle. Lie

derivatives depend on the local behaviour of both X and σ. The covariant derivative on the other hand only depends on the value of X at $p \in M$ and on the local behaviour of σ.

Definition 34.7.5 (Parallel tensor field). A tensor field T is said to be parallel with respect to a connection ∇ if it satisfies $\nabla T = 0$. It is said to be parallel with respect to a vector field X if $\nabla_X T = 0$.

Example 34.7.6 (Riemannian geometry). Important examples in the case of the *Levi-Civita connection* on a *Riemannian manifold* are the volume form Vol and the *metric* g (see Chapter 36).

Property 34.7.7 (Affinity). Consider two Koszul connections $\nabla, \overline{\nabla}$ on a vector bundle $E \to M$. The operator $\nabla - \overline{\nabla}$ is an endomorphism of E, i.e. $\nabla - \overline{\nabla} \in \Omega^1(M; \mathsf{End}(E))$. It follows that the set of Koszul connections is an affine space. (This property gives its name to affine connections.)

Definition 34.7.8 (Connection coefficients). Let E be a smooth rank-k vector bundle. Consider a Koszul connection ∇, a (local) frame $\{e_i\}_{1\leq i\leq k}$ and a (local) coframe $\{f^i\}_{1\leq i\leq k}$ on E. For every vector field e_i one can (locally) write

$$\nabla e_i = \Gamma^k{}_{ji} e_k \otimes f^j \,. \tag{34.81}$$

The quantities $\Gamma^k{}_{ji}$ are called the connection coefficients or **Christoffel symbols** of ∇. For a general vector field $\sigma \equiv \sigma^i e_i$ one obtains (if $\{e_i, f^i\}_{1\leq i\leq k}$ are coordinate-induced):

$$\begin{aligned} \nabla\sigma &= (\nabla\sigma^i) \otimes e_i + \sigma^i(\nabla e_i) \\ &= (\partial_j \sigma^i) e_i \otimes f^j + \sigma^i(\Gamma^k{}_{ji} e_k \otimes f^j) \\ &= (\partial_j \sigma^k + \Gamma^k{}_{ji}\sigma^i) e_k \otimes f^j \,. \end{aligned} \tag{34.82}$$

Definition 34.7.9 (Exterior covariant derivative). Consider a Koszul connection on a vector bundle $\pi : E \to M$. The exterior covariant derivative of an E-valued k-forms $\theta \in \Omega^k(M; E)$ is defined as follows:

$$\begin{aligned} \mathsf{D}\theta(X_0, \dots, X_k) = &\sum_i^k (-1)^i \nabla_{X_i}\big(\theta(X_0, \dots, \hat{X}_i, \dots, X_k)\big) \\ &+ \sum_{i<j}^k (-1)^{i+j} \theta([X_i, X_j], X_0, \dots, \hat{X}_i, \dots, \hat{X}_j, \dots, X_k) \,, \end{aligned} \tag{34.83}$$

where, as usual, $\hat{X}_i$ indicates that this vector is omitted. This formula should remind the reader of the analogous formula for the ordinary exterior derivative (Formula 34.4.11).

34.7.2 Induced connections

Example 34.7.10 (Affine connection). A Koszul connection on the tangent bundle.

The following formula can be seen as a generalization of the Leibniz rule:

Formula 34.7.11 (Tensor products). Let E, E' be two vector bundles over a smooth manifold M. Koszul connections on E and E' induce a connection on the tensor product bundle $E \otimes E'$ as follows:

$$\nabla(\sigma \otimes \rho) := \nabla\sigma \otimes \rho + \sigma \otimes \nabla\rho \tag{34.84}$$

for $\sigma \in \Gamma(E), \rho \in \Gamma(E')$.

Formula 34.7.12 (Connection on tensors). Applying the Leibniz property of a Koszul connection to tensor contractions gives the following form of the induced connection on the (k, l)-tensor bundle:

$$\begin{aligned}\nabla_Y T(\omega^1, \dots, \omega^k, X_1, \dots, X_l) := Y\Big(T(\omega^1, \dots, \omega^k, X_1, \dots, X_l)\Big) \\ - \sum_{i=1}^{k} T(\omega^1, \dots, \nabla_Y \omega^i, \dots, \omega^k, X_1, \dots, X_l) \\ - \sum_{i=1}^{l} T(\omega^1, \dots, \omega^l, X_1, \dots, \nabla_Y X_i, \dots, X_l)\,,\end{aligned} \tag{34.85}$$

where $Y, X_1, \dots, X_l \in \mathfrak{X}(M)$ and $\omega^1, \dots, \omega^k \in \Omega^1(M)$.

Corollary 34.7.13 (Iterated derivatives). By noting that the covariant derivative of a vector field is a vector-valued differential form, one can use the previous formula to compute the covariant derivative of the covariant derivative:

$$(\nabla_X \nabla)_Y Z = \nabla_X(\nabla_Y Z) - \nabla_{\nabla_X Y} Z - \nabla_Y(\nabla_X Z)\,. \tag{34.86}$$

Parentheses were added to make it clear that the outer covariant derivatives act on the result of the inner derivatives. If these parentheses would not have been added, these terms could have been confused with the second covariant derivative (whose definition also follows from a Leibniz-type argument):

$$\nabla^2_{X,Y} S := \nabla_X(\nabla_Y S) - \nabla_{\nabla_X Y} S\,. \tag{34.87}$$

As an example of the second-order covariant derivative, the definition of the Hessian on arbitrary smooth manifolds is given.

Definition 34.7.14 (Hessian). Consider a smooth manifold with affine connection ∇. The Hessian of a function $f \in C^\infty(M)$ is defined as the iterated covariant derivative:

$$\mathsf{Hess}(f) := \nabla^2 f\,, \tag{34.88}$$

where one should note that by the above definition the first covariant derivative also acts on the second one, i.e.

$$\nabla^2 f(X, Y) = \nabla_X(\nabla_Y f) - \nabla_{\nabla_X Y} f\,. \tag{34.89}$$

For scalar functions and covector fields, it is known that $\nabla f = \mathrm{d}f$ and (in local coordinates)

$$\nabla_i \sigma_j = \partial_i \sigma_j - \Gamma^k{}_{ij} \sigma_k \,,$$

where $\Gamma^k{}_{ij}$ are the connection coefficients. Combining these facts, one obtains the following local expression for the Hessian of f:

$$\mathsf{Hess}(f) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} - \Gamma^k{}_{ij} \frac{\partial f}{\partial x_k} \right) \mathrm{d}x^i \otimes \mathrm{d}x^j \,. \tag{34.90}$$

Definition 34.7.15 (Pullback connection). Let $E \to N$ be a vector bundle with Koszul connection ∇ and let $f : M \to N$ be a smooth function. On the pullback bundle (Construction 33.2.5), there exists a unique Koszul connection ∇' satisfying

$$\nabla'(f^* \chi) = f^*(\nabla \chi) \tag{34.91}$$

for any section χ of E. This connection is often denoted by $f^*\nabla$.

Definition 34.7.16 (Invariant connection). Let G be a Lie group acting on a smooth manifold M. A Koszul connection ∇ on $E \to M$ is said to be invariant with respect to the G-action if it satisfies:

$$g^* \nabla = \nabla \tag{34.92}$$

for all $g \in G$. (In fact, invariance can be considered under the action of any set of endomorphisms of M.)

34.7.3 Curvature & torsion

Definition 34.7.17 (Curvature). Let ∇ be a Koszul connection on a vector bundle $E \to M$. The associated curvature 2-form $F_\nabla \in \Omega^2(M; \mathsf{End}(E))$ is defined as follows:

$$F_\nabla(X, Y) := [\nabla_X, \nabla_Y] - \nabla_{[X,Y]} \,. \tag{34.93}$$

Definition 34.7.18 (Torsion). Let ∇ be an affine connection on a smooth manifold M. The associated torsion 2-form $T \in \Omega^2(M; TM)$ is defined as follows:

$$T(X, Y) := \nabla_X Y - \nabla_Y X - [X, Y] \,. \tag{34.94}$$

Property 34.7.19 (Bianchi identities). Let ∇ be an affine connection on a smooth manifold M. The associated curvature and torsion forms are related as follows:

$$\begin{aligned} R(X,Y)Z - T\big(T(X,Y), Z\big) + \text{cyclic} &= \nabla_X T(Y,Z) \,, \\ \nabla_X R(Y,Z) + R\big(T(X,Y), Z\big) + \text{cyclic} &= 0 \,, \end{aligned} \tag{34.95}$$

where 'cyclic' indicates terms obtained by cyclic permutations of the vector fields.

34.8 Integration Theory

For the theory of measure spaces and Lebesgue integration, see Chapter 18. Further below, integration theory will be generalized from orientable manifolds to nonorientable manifolds. A good introduction for this is Tiee (2006).

34.8.1 Orientation and densities

One can define an orientation on manifolds by generalizing the situation for vector spaces (Definition 23.4.25).

Definition 34.8.1 (Orientable manifold). First, the definition of the volume element needs to be slightly modified. A **volume form** on M is a nowhere-vanishing top-dimensional differential form $\mathsf{Vol} \in \Omega^{\dim(M)}(M)$. The definition of an orientation is now virtually the same as for vector spaces.

An **oriented atlas** is given by all charts of M for which the pullback of the Euclidean volume form is a positive multiple of Vol. This also implies that the transition functions have a positive Jacobian determinant. A differentiable manifold that can equipped with such a volume form (and corresponding orientable atlas) is called an orientable manifold. If such a choice has been made, it is said to be **oriented**. Moreover, on an oriented manifold, a basis $\{v_1, \dots, v_n\}$ for a tangent space T_pM is said to be **positively oriented** if

$$\mathsf{Vol}_p(v_1, \dots, v_n) > 0\,. \tag{34.96}$$

Example 34.8.2 (Determinant). Let $M = \mathbb{R}^n$. The canonical Euclidean volume form is given by the determinant map

$$\det : (v_1, \dots, v_n) \mapsto \det(v_1, \dots, v_n)\,, \tag{34.97}$$

where the v_n's are expanded in the standard basis $\{e_1, \dots, e_n\}$ to calculate the determinant. The terminology of 'volume forms' is justified by noting that the determinant map gives the signed volume of the n-dimensional parallelotope spanned by the vectors $\{v_1, \dots, v_n\}$.

Property 34.8.3. Let ω_1, ω_2 be two volume forms on M. Because the space of top-degree forms is one-dimensional, there exists a smooth function f such that

$$\omega_1 = f\omega_2\,.$$

Furthermore, the sign of this function is constant on every connected component of M.

One can also rephrase orientability of manifolds in terms of bundles.

Definition 34.8.4 (Orientation bundle). Consider a manifold M. The transition function A of TM is given by the Jacobian of the transitions functions on M. The associated line bundle with transition function $\operatorname{sgn}\det(A)$ is called the orientation bundle $o(M)$.

In general, one can define the orientation bundle $o(E)$ for any vector bundle E, where one replaces the Jacobian in the above construction by the transition maps of E. From this, it is clear that the orientation bundle $o(M)$ is the same as $o(TM)$.

Alternative Definition 34.8.5 (Orientable manifold). A smooth manifold for which the orientation bundle is trivial.

Remark 34.8.6. By definition of the orientation bundle, the transition functions are those that have a positive determinant. This gives the equivalence with Definition 34.8.1. In the next chapter on principal bundles, yet another (equivalent) definition of orientability in terms of *G-structures* will be given (see Example 35.5.5).

A last definition, strongly related to the previous one, is the following.

Alternative Definition 34.8.7. Consider a manifold M and choose a triangulation $\mathcal{K}$ (Definition 7.2.8). M is orientable if the pullback of the tangent bundle TM to the 1-skeleton $\mathcal{K}^1$ is trivial.

Definition 34.8.8 (Tensor density). Consider a vector bundle $E \to M$ defined by transition functions A. The associated bundle of (tensor) s-densities is obtained by using the representation

$$\rho : A \mapsto \det(A)^{-s} . \tag{34.98}$$

The number $s \in \mathbb{R}$ is called the **weight** of the density. For $E \equiv TM$, one obtains the (tensor) s-densities on M, which, in the case of $s = 1$, are equivalent to top-dimensional forms on M. When twisting a vector bundle by an s-density bundle, the prefix 's-weighted' is often added.

Example 34.8.9 (Pseudoscalars). Let G be a Lie group and consider a group morphism $\phi : G \to \mathrm{O}(p,q)$ for some $p, q \in \mathbb{N}$. The pseudoscalar representation of G, induced by ϕ, is defined as the one-dimensional representation given by

$$\mathbf{1}_{\mathrm{sgn}} : g \mapsto \det\big(\phi(g)\big) . \tag{34.99}$$

The notation $\mathbf{1}_{\mathrm{sgn}}$ refers to the fact that this is a generalization of the alternating representation (Example 24.1.4) of the permutation groups S_n.

Sections of a vector bundle with transition functions defined by $\mathbf{1}_{\mathrm{sgn}}$ are generally called **pseudoscalar fields**. When using the pseudoscalar representation of the transition functions of the tangent bundle TM to construct an associated bundle, one obtains the pseudoscalar bundle Ψ_M. A vector bundle twisted by the pseudoscalar bundle Ψ often receives the prefix 'pseudo', e.g. the Ψ-twisted k-form bundle is called the bundle of k-**pseudoforms**.

Example 34.8.10 (Pseudovectors). The representation

$$\rho : A \mapsto \det(A)A \tag{34.100}$$

gives rise to a bundle similar to the tangent bundle, where the sign of the cocycles now has an influence on the fibres. Sections of such bundles are called **pseudovector fields**. This construction is equivalent to twisting the tangent bundle by the pseudoscalar bundle, hence its name.

Remark 34.8.11 (Honest densities). One should pay incredible attention to the definition of a **density** (i.e. without the prefix 'tensor'). A density is defined as an n-pseudoform, i.e. a section of the **density bundle** $|\Omega|(M) := \Omega^n(M) \otimes o(M)$. Here, the transition function is $|\det(A)|$, where A is the transition function of T^*M.[5] These are the objects one can integrate over any manifold, even the nonorientable ones. They are essentially maps $\Gamma(\det(T^*M)) \to C^\infty(M)$. A naive way to construct a density on a manifold M is by choosing a volume form $\mathrm{Vol}(M)$ and taking the absolute value $|\mathrm{Vol}(M)|$.

One can also define **honest s-densities** $|\Omega|^s(M)$ by tensoring tensor densities with the orientation bundle to obtain transition maps $|\det(A)|^s$, where A is again the cotangent transition function. This is also the only possible way to generalize the (tensor) s-densities to real s.

Property 34.8.12 (Orientability). A smooth manifold is orientable if and only if its canonical line bundle (Example 34.2.25) is trivial. Furthermore, for orientable manifolds, there exists an isomorphism $\Gamma(\det(T^*M)) \cong \Gamma(|\Omega|(M))$.

34.8.2 Orientation in homology

In this section, a characterization of orientability in terms of homology is given. As such, smoothness is not required. See Section 7.2 and Section 7.3 for an introduction to homology.

First, consider the canonical example $\mathbb{R}^n$. Intuitively one would expect an orientation on an Euclidean space to be a property that is preserved under rotations and reversed by reflections. On the sphere these operations have degree 1 and -1 respectively, so the perfect choice for an orientation would be the generator of $H_n(S^n) \cong \mathbb{Z}$. Luckily, there exists an isomorphism $H_n(S^n) \cong H_n(\mathbb{R}^n, \mathbb{R}^n \setminus \{*\})$. So, for every point $x \in \mathbb{R}^n$, one can define a local orientation as a choice of generator of the local homology group $H_n(\mathbb{R}^n, \mathbb{R}^n \setminus \{x\})$.

Alternative Definition 34.8.13 (Orientation). An orientation on a manifold M is a choice of local orientation for every point $p \in M$ such that every two points admitting a common covering chart have consistent local orientations.

[5] Note that, in the literature, one often uses the transition functions of TM, which are the inverses A^{-1}. This leads to expressions involving negative exponents.

Property 34.8.14 (Orientability). If a closed, connected manifold is ($\mathbb{Z}$-)orientable, there exists an isomorphism

$$H_n(M) \cong H_n(M, M\backslash\{p\}) \cong \mathbb{Z} \tag{34.101}$$

for all points $p \in M$. A choice of class in $H_n(M)$ that maps to a generator of $H_n(M, M\backslash\{p\})$ for all $p \in M$ is called a **fundamental class** or **orientation class**.

When M is not connected, the fundamental class equals the direct sum of the generators of the connected components (following the idea of the additivity axiom 7.5.1).

The above definition and property can be generalized to arbitrary unital rings R.

Definition 34.8.15 (R-orientability). A manifold is R-orientable if a consistent choice of local R-orientation exists or, equivalently, if $H_n(M; R) \cong R$.

Property 34.8.16 (Nonorientable manifolds). If M is not R-orientable, the map

$$H_n(M; R) \to H_n(M, M\backslash\{p\}; R)$$

is still injective with image $\{r \in R \mid 2r = 0\}$. In particular, every closed manifold is $\mathbb{Z}_2$-orientable.

Property 34.8.17 (Orientability implies R-orientability). By the *universal coefficient theorem*, it follows that a $\mathbb{Z}$-orientable manifold is also R-orientable for all unital rings R. Conversely, a manifold is $\mathbb{Z}$-orientable if it is R-orientable for all unital rings R.

34.8.3 Integration

Definition 34.8.18 (Measure zero). A subset $U \subseteq M$ of an orientable manifold is said to be of measure zero (or **null**) if it is the countable union of inverse images (with respect to the chart maps on M) of null sets in $\mathbb{R}^n$.

Definition 34.8.19 (Integrable form). A differential form for which its components with respect to any basis of $\Omega^{\dim(M)}(M)$ are Lebesgue integrable on $\mathbb{R}^{\dim(M)}$.

Formula 34.8.20 (Integration with compact support). Consider a top-dimensional form $\omega \in \Omega^{\dim(M)}$ on M with compact support on a coordinate patch $U \subseteq M$.

$$\int_M \omega = \int_U \omega := \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \omega_{12\ldots n}(x)\, dx^1\, dx^2 \cdots dx^n\,. \tag{34.102}$$

This integral is well defined because, under an orientation-preserving coordinate transformation, the component $\omega_{1\ldots n}$ transforms as $\omega'_{1\ldots n} = \det(J)\omega_{1\ldots n}$, where J is the Jacobian of the coordinate transformation. Inserting this in the integral and replacing dx_i by dx'_i gives the well-known change-of-variables formula from (Lebesgue) integration theory.

If one requires the manifold M to be paracompact, such that every open cover $\{U_i \subseteq M\}_{i\in I}$ admits a subordinate partition of unity $\{\phi_i\}_{i\in I}$, one can define the integral of a general compactly supported form $\omega \in \Omega^n(M)$ as follows:

$$\int_M \omega := \sum_{i\in I} \int_{U_i} \rho_i \omega \,. \tag{34.103}$$

Remark 34.8.21. Although integration was only defined for compactly supported forms, the general formula can also be applied to general forms. It is well defined whenever the forms $\rho_i\omega$ are integrable and the sum in the definition converges.

Property 34.8.22 (Compact manifolds). Let M be a smooth compact manifold. Because every form on M is automatically compactly supported, all forms are integrable on M.

Property 34.8.23 (Invariance under pullbacks). Consider an orientation-preserving diffeomorphism $f : M \to N$.

$$\int_M f^*\omega = \int_N \omega \tag{34.104}$$

Notation 34.8.24. Because the integral of differential forms is linear in the integrand and additive over disjoint unions, it can be interpreted as a bilinear pairing. This motivates the following notation:

$$\langle M, \omega\rangle := \int_M \omega \,. \tag{34.105}$$

34.8.4 Stokes' theorem

Theorem 34.8.25 (Stokes' theorem). *Let M be an orientable manifold with boundary ∂M and let ω be a differential k-form on M.*

$$\int_{\partial M} \omega = \int_M \mathrm{d}\omega \tag{34.106}$$

Corollary 34.8.26. The Kelvin–Stokes theorem 23.2.5, the divergence theorem 23.2.6 and Green's identity 23.2.7 are immediate results of Stokes' theorem.

Definition 34.8.27 (Calibration). A degree-p calibration on a smooth manifold M with volume form Vol is a differential form $\omega \in \Omega^p(M)$ satisfying the following conditions:

1. **Closedness**: $\mathrm{d}\omega = 0$.
2. **Volume**: Over any dimension-p submanifold, the integral of ω is smaller than its volume, with at least one submanifold saturating the inequality.

A submanifold is said to be **calibrated** if the restriction of the calibration to this submanifold coincides with the induced volume form.

Property 34.8.28. Calibrated submanifolds minimize the volume within their homology class.

34.9 Cohomology

34.9.1 de Rham complex

Definition 34.9.1 (Exact form). A differential form $\omega \in \Omega^k(M)$ such that $\omega = \mathrm{d}\chi$ for some $\chi \in \Omega^{k-1}(M)$.

Definition 34.9.2 (Closed form). A differential form $\omega \in \Omega^k(M)$ such that $\mathrm{d}\omega = 0$.

Definition 34.9.3 (de Rham complex). The sequence

$$0 \longrightarrow \Omega^0(M) \longrightarrow \Omega^1(M) \longrightarrow \cdots \longrightarrow \Omega^{\dim(M)}(M) \longrightarrow 0 \tag{34.107}$$

together with the sequence of exterior derivatives d_k forms a cochain complex by the nilpotency of the exterior derivative. This complex is called the de Rham complex $\Omega^\bullet_{\mathrm{dR}}(M)$. It encodes the information that every exact form is closed. The converse, however, is not true in general (see Theorem 34.9.8 below for more information).

Definition 34.9.4 (de Rham cohomology). Following Definition 5.1.3, the kth de Rham cohomology group on M is defined as the kth (co)homology group of the de Rham complex:

$$H^k_{\mathrm{dR}}(M) := \frac{\ker(\mathrm{d}_k)}{\mathrm{im}(\mathrm{d}_{k-1})}\,. \tag{34.108}$$

This quotient space is a vector space. Two elements of the same equivalence class in $H^k_{\mathrm{dR}}(M)$ are said to be **cohomologous**.

Definition 34.9.5 (Integral form). A closed k-form ω that lies in the image of the inclusion $H^k_{\mathrm{dR}}(M,\mathbb{Z}) \hookrightarrow H^k_{\mathrm{dR}}(M,\mathbb{R})$. Equivalently, a closed k-form is integral if integrating it over any integral k-cycle gives an integer.

Formula 34.9.6 (Cup product). Let $[\nu] \in H^k_{\mathrm{dR}}(M)$ and $[\omega] \in H^l_{\mathrm{dR}}(M)$. The cup product on de Rham cohomology is given by

$$[\nu] \smile [\omega] := [\nu \wedge \omega]\,. \tag{34.109}$$

The following theorem allows to drop the subscript 'dR' when using de Rham cohomology.

Theorem 34.9.7 (de Rham). *The de Rham cohomology over a smooth manifold is isomorphic to its singular cohomology (Section 7.3).*

Theorem 34.9.8 (Poincaré lemma[6]). *For every point $p \in M$, there exists a neighbourhood $U \subseteq M$ on which the de Rham cohomology is trivial. More generally, this lemma says that the following isomorphism exists for every smooth manifold M:*

$$H^\bullet(M \times \mathbb{R}^n) \cong H^\bullet(M)\,. \tag{34.110}$$

[6]The original theorem states that, on a contractible space (Definition 7.1.7), every closed form is exact.

In fact, this can even be further generalized due to the homotopy axiom of de Rham cohomology:

$$H^\bullet(E) \cong H^\bullet(M) \tag{34.111}$$

for every vector bundle E over M.

Corollary 34.9.9. Every closed form is locally exact, i.e. if $\mathrm{d}\omega = 0$ at the point $p \in M$, there exist a neighbourhood $U \subseteq M$ of p and a differential form λ such that

$$\omega = \mathrm{d}\lambda \tag{34.112}$$

at all points of U.

Definition 34.9.10 (Relative cohomology). Consider the submanifold inclusion $\iota : S \hookrightarrow M$. The relative de Rham complex is defined as follows:

$$\Omega^n(M, S) := \Omega^n(M) \oplus \Omega^{n-1}(S)\,, \tag{34.113}$$

where the coboundary operator d is defined by

$$\mathrm{d}(\omega, \lambda) := (\mathrm{d}\omega, \iota^*\omega - \mathrm{d}\lambda)\,. \tag{34.114}$$

The relative de Rham cohomology $H^\bullet(M, S)$ is defined as the cohomology of this complex. Classes are represented by closed forms on M that restrict to exact forms on S.

This definition can be generalized to any smooth map $f : M \to N$ by replacing ι^* in the coboundary by f^*. This cohomology ring is denoted by $H^\bullet(f)$. For all smooth functions f, the following long exact sequence exists:

$$\cdots \longrightarrow H^k(f) \longrightarrow H^k(M) \longrightarrow H^k(N) \longrightarrow H^{k+1}(f) \longrightarrow \cdots\,. \tag{34.115}$$

Definition 34.9.11 (Twisted de Rham complex). Consider the usual de Rham complex $\Omega^\bullet(M)$ on a smooth manifold M. For every degree-3 class $\alpha \in H^3(M)$, one can define the twisted de Rham differential

$$\mathrm{d}_\alpha := \mathrm{d} + \alpha \wedge\,. \tag{34.116}$$

Nilpotency follows from that of d and the degree of α. The cohomology of this complex is called the (α-)twisted de Rham cohomology of M.

The de Rham theorem above can be generalized to the setting of equivariant cohomology (Definition 7.6.1).

Property 34.9.12 (Equivariant de Rham theorem ♣). Consider a smooth manifold M with a smooth G-action and let $W(\mathfrak{g})$ be the Weil algebra (Definition 32.6.6) of G. Construct the dgca $\Omega^\bullet(M) \otimes W(\mathfrak{g})$ with differential $\mathrm{d}_{\mathrm{dR}} + \mathrm{d}_{\mathrm{W}}$. The infinitesimal G-action gives a map $\mathfrak{g} \to \mathfrak{X}(M)$, so Cartan calculus can be extended to all of $\Omega^\bullet(M) \otimes W(\mathfrak{g})$. The **basic**

differential forms are defined as the kernel of the Cartan operators $\iota_{\tilde{\xi}}, \mathcal{L}_{\tilde{\xi}}$ for all $\tilde{\xi} \in \mathfrak{g}$. This subcomplex, with the induced differential $\mathsf{d}_{\mathsf{dR}} + \mathsf{d}_{\mathsf{W}}$, is called the **Weil model** of equivariant de Rham cohomology.

If G is compact and connected, the cohomology of the Weil model is isomorphic to the G-equivariant cohomology $H^\bullet_G(M)$ from Definition 7.6.1.

Property 34.9.13 (Cohomological models). The intersection of the basic subcomplex with $\Omega^\bullet(M)$ can be identified with the complex of tensorial (basic) differential forms (Definition 35.3.13), i.e. the pullback $\pi^*\Omega^\bullet(M) \subset \Omega^\bullet(P)$. Moreover, the intersection with the Weil algebra gives a model for the classifying space BG and the Weil algebra gives a model for the total space EG.

34.9.2 Integration

At this point, a little side note can be given about why the de Rham cohomology groups (Definition 34.9.4) really constitute a cohomology theory. Some concepts from homology are needed that can be found in Section 7.2.

Let M be a compact manifold. Suppose that one wants to integrate a form over a singular k-chain $C \equiv \sum_{i=0}^k a_i\lambda_i$ on M. Through integration, one can pair the k-form ω and the singular chain C as if they are dual objects (hence p-forms are also called p-**cochains**) to produce a real number:

$$\langle \cdot, C\rangle : \Omega^k(M) \to \mathbb{R} : \omega \mapsto \int_C \omega := \sum_{i=0}^k a_i \int_{\Delta_k} \lambda_i^*\omega\,, \tag{34.117}$$

where λ_i^* pulls ω back to Δ^k, which is a subset of $\mathbb{R}^k$ as required. Stokes' theorem 34.8.25 then says that

$$\int_C \mathsf{d}\omega = \int_{\partial C} \omega\,. \tag{34.118}$$

Using the pairing $\langle\cdot,\cdot\rangle$, this can be rewritten more explicitly as

$$\langle \mathsf{d}\omega, C\rangle = \langle \omega, \partial C\rangle\,. \tag{34.119}$$

The operators d and ∂ can thus be interpreted as formal adjoints. By Stokes theorem, all chains C and cochains ω belonging to the same equivalence classes $[C] \in H_k(M;\mathbb{R})$ and $[\omega] \in H^k(M;\mathbb{R})$ give rise to the same number $\langle\omega, C\rangle$, so one can see that the singular homology groups and the de Rham cohomology groups on M are well defined dual groups. The name *co*homology is thus well chosen for Definition 34.9.4.

Remark 34.9.14 (L^2-cohomology). Given a *Riemannian metric* (see Chapter 36), one can define a notion of square-integrable differential forms. The cohomology theory of these forms is called L^2-cohomology.

34.9.3 Cohomology with compact support

Because integration is involved in all statements in this section, it will be assumed that all manifolds and bundles are orientable (unless stated otherwise).

The following definition characterizes cohomology with compact support directly through its relation to compact sets.

Definition 34.9.15 (Cohomology with compact support). Consider a manifold M (not necessarily orientable). The cohomology with compact support $H_c^\bullet(M)$ is defined as the following direct limit:

$$H_c^\bullet(M) := \varinjlim_{\text{compact } K \subseteq M} H^\bullet(M, M \backslash K) \,. \tag{34.120}$$

Property 34.9.16 (Relation to reduced cohomology). For a topological space X, the inclusion $U \hookrightarrow X$ for any open U induces a long exact sequence in compactly supported cohomology. Performing excision by $V := X \backslash U$ in the above definition of compact cohomology gives $H^\bullet\big(X, X \backslash (K \cup V)\big) \cong H^\bullet(U, U \backslash K)$ and thus $H_c^\bullet(X, V) \cong H_c^\bullet(U)$.

If X is chosen to be the one-point compactification $\widehat{M}$ of M and if $U = M$, the aforementioned long exact sequence implies that

$$H_c^\bullet(M) \cong H^\bullet(\widehat{M}, *) \cong \widetilde{H}^\bullet(\widehat{M}) \,, \tag{34.121}$$

where the fact that both $*$ and $\widehat{M}$ are compact, which is true for Hausdorff spaces, is used.

Theorem 34.9.17 (Poincaré duality). *Let M be an m-dimensional manifold. The pairing $\int : H^k(M) \otimes H_c^{m-k}(M) \to \mathbb{R}$ induces an isomorphism on cohomology:*

$$H^k(M) \cong \big(H_c^{m-k}(M)\big)^* \,. \tag{34.122}$$

If M is of finite type (Definition 31.1.16), the converse also holds:

$$H_c^k(M) \cong \big(H^{m-k}(M)\big)^* \,. \tag{34.123}$$

Corollary 34.9.18 (Poincaré lemma for compact cohomology). Let M be a (not necessarily orientable) manifold of finite type. For every rank-n vector bundle E over M, the following isomorphism exists:

$$H_c^\bullet(E) \cong H_c^{\bullet - n}(M) \,. \tag{34.124}$$

Definition 34.9.19 (Poincaré dual). Let M be an n-dimensional manifold and let $i : S \hookrightarrow M$ be a closed k-dimensional submanifold. The Poincaré dual of S in M is the unique cohomology class $[\eta_S] \in H^{n-k}(M)$ such that

$$\int_S i^* \omega = \int_M \omega \wedge \eta_S \tag{34.125}$$

for all compactly supported $\omega \in H_c^k(M)$. If S is compact in M, two Poincaré duals exist:

- **Closed dual**: The Poincaré dual obtained by using the fact that S is compact and hence closed in M.
- **Compact dual**: Because S is compact, all forms $\omega \in H^k(M)$ (not only the compactly supported ones) can be integrated over S and, assuming M is of finite type, Poincaré duality implies that there exists a unique cohomology class with compact support η'_S such that

$$\int_S i^*\omega = \int_M \omega \wedge \eta'_S \tag{34.126}$$

for all $\omega \in H^k(M)$.

Remark 34.9.20. Because the compact Poincaré dual induces a pairing on all closed forms ω, which include the compactly supported ones, the compact dual is equal to the closed Poincaré dual as a differential form. However, as elements in cohomology these can be quite different.

Property 34.9.21 (Localization principle). The support of the compact Poincaré dual of a compact submanifold S may be shrunk to any neighbourhood of S. More generally, the support of the (closed) Poincaré dual of a closed submanifold S can be shrunk to any tubular neighbourhood of S.

Formula 34.9.22 (Transversal intersections). The Poincaré dual of a transversal intersection is equal to the wedge product of the individual Poincaré duals:

$$\eta_{S \pitchfork T} = \eta_S \wedge \eta_T \,. \tag{34.127}$$

Definition 34.9.23 (Compact vertical cohomology). Let $\pi : E \to M$ be a smooth vector bundle over M. A differential form $\omega \in \Omega^\bullet(E)$ is said to be an element of $\Omega^\bullet_{\mathrm{cv}}(E)$ if $\mathsf{supp}(\omega) \cap \pi^{-1}(K)$ is compact for every compact subset $K \subseteq M$. The cohomology of this complex is called the **de Rham cohomology with compact support in the vertical direction**.

Remark. The above definition implies that $\omega \in \Omega^\bullet_{\mathrm{cv}}(E)$ is compactly supported on the fibre $\pi^{-1}(p)$ for all points $p \in M$. This observation explains the name of the cohomology theory.

Definition 34.9.24 (Fibre integration). Differential forms with vertically compact support on a rank-n vector bundle $\pi : E \to M$ can be divided into two classes:

- **Type 1**: Those locally of the form $f(x,u)\pi^*\phi \wedge (\mathrm{d}u_{i_1} \wedge \cdots \wedge \mathrm{d}u_{i_k})$, where ϕ is a form on the base manifold M, f has compact support and $k < n$.
- **Type 2**: Those locally of the form $f(x,u)\pi^*\phi \wedge (\mathrm{d}u_1 \wedge \cdots \wedge \mathrm{d}u_n)$ where ϕ is a form on the base manifold M and f has compact support.

The fibre integration map $\pi_* : \Omega^\bullet_{\mathrm{cv}}(E) \to \Omega^{\bullet-n}(M)$ is defined as follows. If ω is of type 1, then $\pi_*\omega := 0$. If ω is of type 2, then

$$\pi_*\omega := \left(\int \cdots \int f(x,u)\, du_1 \cdots du_n\right)\phi\,. \tag{34.128}$$

Formula 34.9.25 (Projection formula). Consider a rank-n vector bundle $\pi : E \to M$. For every pair of forms $(\phi, \omega) \in \Omega^\bullet(M) \times \Omega^\bullet_{\mathrm{cv}}(E)$, the following formula holds:

$$\pi_*(\pi^*\phi \wedge \omega) = \phi \wedge \pi_*\omega\,. \tag{34.129}$$

Furthermore, if $\phi \in \Omega^k_c(M)$ and $\omega \in \Omega^{m+n-k}_{\mathrm{cv}}(E)$,

$$\int_E \pi^*\phi \wedge \omega = \int_M \phi \wedge \pi_*\omega\,. \tag{34.130}$$

Theorem 34.9.26 (Thom isomorphism). *For every rank-n vector bundle $\pi : E \to M$, where M is (not necessarily orientable and) of finite type, fibre integration gives the following isomorphisms:*

$$\pi_* : H^\bullet_{cv}(E) \cong H^{\bullet-n}(M) : \mathcal{T}\,. \tag{34.131}$$

Corollary 34.9.27 (Poincaré lemma for vertically compact cohomology).

$$H^\bullet_{\mathrm{cv}}(M \times \mathbb{R}^n) \cong H^{\bullet-n}(M) \tag{34.132}$$

Formula 34.9.28 (Thom isomorphism). Define the **Thom class** of M as $\Phi := \mathcal{T}(1) \in H^0(M)$. Because $\mathcal{T}$ and π_* are mutual inverses and, hence, $\pi_*\Phi = 1$, the projection formula implies that

$$\mathcal{T}(\omega) = \pi^*\omega \wedge \Phi\,. \tag{34.133}$$

Property 34.9.29 (Orientation class). The Thom class Φ restricts to a generator of the cohomology of the typical fibre V:

$$H^n_c(V) \cong \widetilde{H}^n(S^n) \cong H^n(S^n)\,. \tag{34.134}$$

For compact orientable manifolds, such a generator gives rise to a generator of the homology group H_n, i.e. it gives rise to an orientation class (Property 34.8.14).

Property 34.9.30 (Poincaré dual). The Poincaré dual of a closed submanifold is equal to the Thom class of its normal bundle.

The construction of the Thom isomorphism involves some technicalities. For example, throughout the literature, the Thom isomorphism is stated in various forms using compactly supported cohomology, relative cohomology or reduced cohomology. The various approaches are related as follows.

Definition 34.9.31 (Thom space). Let $E \to M$ be a vector bundle. For every fibre in E, one can construct its one-point compactification (Definition 6.5.30) and, by gluing these together, the **sphere bundle** $\mathsf{Sph}(E)$ is obtained. The quotient space $\mathsf{Sph}(E)/B$, where all the adjoined points are identified, is called the Thom space $\mathsf{Th}(E)$.

By equipping E with a *metric* (see Chapter 36), one can give an alternative definition. Let V be the typical fibre of E. A new bundle, the unit sphere bundle $S(E)$, where the typical fibre is the unit sphere $S(V) := \{v \in V \mid \|v\| = 1\}$, can now be constructed. (It should be noted that this new bundle is not a vector bundle since the unit sphere is not a vector space.) A similar construction leads to the unit disk bundle $D(E)$, where the typical fibre is the unit disk $D(V) := \{v \in V \mid \|v\| \leq 1\}$. The Thom space $\mathsf{Th}(E)$ can be shown to be isomorphic to the quotient space $D(E)/S(E)$.

Property 34.9.32. If the base manifold M is compact, the Thom space is obtained as the one-point compactification of the total space E.

Property 34.9.33 (Different forms of Thom isomorphism). Let $E \to M$ be a vector bundle and denote the complement of the zero section by E_0 as in Definition 34.2.8. Homotopy invariance implies that

$$H^\bullet\big(D(E), S(E)\big) \cong H^\bullet(E, E_0)\,. \tag{34.135}$$

Then, using Corollary 34.2.12 together with the dual of Property 7.2.33, one can show that the reduced cohomology of the Thom space $\mathsf{Th}(E)$ is isomorphic to the relative cohomology of the pair (E, E_0):

$$\widetilde{H}^\bullet\big(\mathsf{Th}(E)\big) \cong H^\bullet(E, E_0)\,. \tag{34.136}$$

To relate this to vertically compact cohomology, Property 34.9.16 can be adapted. Compact support gave rise to the (reduced) cohomology of the compactified space. By analogy, vertically compact support corresponds to compactifications of the fibres, which is exactly how the Thom space is constructed.

The above arguments finally lead to the following triangle of isomorphisms:

$$\begin{tikzcd} & \widetilde{H}^\bullet\big(\mathsf{Th}(E)\big) \arrow[dl, "\cong"'] \arrow[dr, "\cong"] & \\ H^\bullet(E, E_0) \arrow[rr, "\cong"'] & & H^\bullet_{\mathrm{cv}}(E)\,. \end{tikzcd} \tag{34.137}$$

Definition 34.9.34 (Thom spectrum ♣). Let $E \to M$ be a vector bundle. The Thom spectrum of E is defined as the suspension spectrum of its Thom space:

$$(\Sigma^\infty \mathsf{Th}(E))_n \cong \mathsf{Th}(\mathbb{R}^n \oplus E)\,, \tag{34.138}$$

where $\mathsf{Th}(\mathbb{R}^n \oplus E) \cong \Sigma^n \mathsf{Th}(E)$ was used.

Now, consider the sequence $(\xi_n)_{n\in\mathbb{N}}$ of *universal vector bundles*. For every ξ_n, define the nth component space as follows:

$$MO_n := \mathsf{Th}(\xi_n)\,. \tag{34.139}$$

The Whitney sum $\xi_n \oplus \mathbb{R}$ can be obtained as a pullback of ξ_{n+1}. This map induces a morphism $\Sigma MO_n \to MO_{n+1}$, which gives the nth structure map of 'the' Thom spectrum MO.[7]

Definition 34.9.35 (Euler class). Consider a vector bundle $E \to M$ together with its Thom class Φ. The Euler class $e(E)$ is defined as the pullback $s_0^*\Phi$ of the Thom class along the zero section of E.

Property 34.9.36. If the orientation of E is reversed, the Euler class changes sign.

The following property distinguishes the Euler class among all characteristic classes of E.

Property 34.9.37 (Normalization). If the vector bundle admits a nowhere-vanishing section, its Euler class vanishes.

34.9.4 Čech–de Rham complex

Theorem 34.9.38 (Mayer–Vietoris sequence). *Consider a smooth manifold M with an open covering $U \cup V$. The cohomology of U, V is related to that of M by the following short exact sequence:*

$$0 \longrightarrow H^\bullet(M) \xrightarrow{\iota_U \oplus \iota_V} H^\bullet(U) \oplus H^\bullet(V) \xrightarrow{\pi_2 - \pi_1} H^\bullet(U \cap V) \longrightarrow 0\,. \tag{34.140}$$

Definition 34.9.39 (Čech–de Rham complex). The Čech complex (Definition 8.4.10) associated to the constant sheaf $\flat\mathbb{R}$, i.e. the sheaf of locally constant functions.

The Mayer–Vietoris sequence can be generalized to a statement about the Čech-de Rham complex.

Property 34.9.40 (Mayer–Vietoris sequence). The horizontal complex

$$0 \longrightarrow \Omega^\bullet(M) \longrightarrow \prod_{i_0} \Omega^\bullet(U_{i_0}) \longrightarrow \prod_{i_0,i_1} \Omega^\bullet(U_{i_0 i_1}) \longrightarrow \cdots \tag{34.141}$$

is acyclic, i.e. the δ-cohomology of the Čech–de Rham complex vanishes.

An important corollary is that one can compute the (de Rham) cohomology of M using the above double complex.

[7] Note that the Thom spectrum as defined here is not an Ω-spectrum (Definition 7.5.6), it is merely a sequential spectrum (prespectrum).

Property 34.9.41. The restriction map $\Omega^\bullet(M) \to C^\bullet(\mathcal{U};\Omega^\bullet)$ induces an isomorphism in cohomology.

One can also augment the Čech–de Rham complex in the other direction by the kernel of the de Rham differential in degree 1. These are the locally constant functions on the intersections $U_{i_0\dots i_p}$. The cohomology of this augmenting sequence $C^\bullet(\mathcal{U};\flat\mathbb{R})$ is called the **Čech cohomology** of M. By the same reason as for why the Mayer–Vietoris sequence implied the above theorem, the following theorem is obtained.

Theorem 34.9.42 (Čech = de Rham). *For a smooth manifold M, admitting a good cover $\mathcal{U}$, the Čech cohomology of $\mathcal{U}$ is isomorphic to the de Rham cohomology of M:*

$$H^\bullet(M) \cong \check{H}^\bullet(\mathcal{U};\flat\mathbb{R})\,. \tag{34.142}$$

By noting that good covers are cofinal in the set of open covers, one can pass to the full Čech cohomology:

$$H^\bullet(M) \cong \check{H}^\bullet(M;\flat\mathbb{R})\,. \tag{34.143}$$

Corollary 34.9.43. All compact manifolds admit a finite good cover and, hence, have finite-dimensional de Rham cohomology.

Property 34.9.44 (Exponential sequence). Consider the following exact sequence of topological groups:

$$0 \longrightarrow \mathbb{Z} \xrightarrow{2\pi} \mathbb{C} \xrightarrow{\exp} \mathrm{U}(1) \longrightarrow 0\,. \tag{34.144}$$

Let $(M,\mathcal{O}_M)$ be a *complex manifold* (see Chapter 39) or a smooth manifold and restrict the above sequence to the real numbers. The exact sequence induces an exact sequence of structure sheaves:

$$0 \longrightarrow \flat\mathbb{Z} \longrightarrow \mathcal{O}_M \longrightarrow \mathcal{O}_M^\times \longrightarrow 0\,. \tag{34.145}$$

This, in turn, induces a long exact sequence in cohomology and, by Example 8.4.6 (if M is paracompact), the connecting homomorphism leads to an isomorphism:

$$H^{\bullet+1}(M;\mathbb{Z}) \cong \check{H}^\bullet\big(M;\mathrm{U}(1)\big)\,. \tag{34.146}$$

Note that the cohomology on the right-hand side is not singular cohomology. Singular U(1)-valued cohomology could also be related to integral cohomology through the *universal coefficient theorem*, but extra terms involving Ext-functors would appear.

34.9.5 Nonorientable manifolds

This section gives a differential-geometric incarnation of Section 7.4 on local coefficients.

Definition 34.9.45 (Twisted cohomology). Let $E \to M$ be a flat vector bundle over M. By Construction 34.4.26, the algebra $\Omega^\bullet(M) \otimes E$ can be given the structure of a differential graded algebra (Definition 5.1.9) and, hence, gives rise to a cohomology theory $H^\bullet(M; E)$. This is called the E-twisted de Rham cohomology of M.

Remark 34.9.46. According to the remark following Construction 34.4.26, attention should be payed to which trivialization was used in the construction of $H^\bullet(M; E)$. However, it can be shown that two trivializations give rise to the same E-twisted cohomology if they admit a common refinement for which the induced sections differ by a locally constant matrix in $\mathsf{GL}(n, \mathbb{R})$.

If one takes $E = o(M)$ to be the orientation line bundle over M, the (honest) densities of Remark 34.8.11 are obtained. The cohomology of this complex is simply called the **twisted de Rham cohomology**.

Property 34.9.47 (Isomorphism). The twisted cohomologies defined by two trivializations induced from atlases on M are isomorphic.[8]

Property 34.9.48 (Trivial twisting). If M is orientable, its twisted cohomology is isomorphic to its ordinary (de Rham) cohomology. More generally, the E-twisted de Rham cohomology is isomorphic to the ordinary de Rham cohomology whenever E is trivial.

Poincaré duality (Theorem 34.9.17) can be generalized almost verbatim to the twisted case.

Theorem 34.9.49 (Poincaré duality). *Integration of densities induces the following isomorphism:*

$$H^k(M) \cong \left(H_c^{m-k}\big(M; o(M)\big)\right)^* . \tag{34.147}$$

If M is of finite type, the converse also holds:

$$H_c^k(M) \cong \left(H^{m-k}\big(M; o(M)\big)\right)^* . \tag{34.148}$$

The Thom isomorphism also holds for nonorientable bundles.

Theorem 34.9.50 (Thom isomorphism). *Let $E \to M$ be a rank-n vector bundle. Fibre integration gives the following isomorphism:*

$$H_{cv}^{\bullet+n}(E) \cong H^\bullet\big(M; o(E)\big) . \tag{34.149}$$

[8] Although one almost always works with a natural trivialization, i.e. the open subsets of M are obtained from charts on M, this is technically not necessary. For more 'exotic' cases, the isomorphisms not always exist.

34.9.6 Generalized cohomology ♣

In this section, some statements from singular/de Rham cohomology on vector bundles are generalized to statements about the generalized (Eilenberg–Steenrod) cohomology theories from Section 7.5. In the remainder of this section, $E^\bullet$ will denote a multiplicative generalized cohomology theory.[9]

Definition 34.9.51 (Orientation). Consider a rank-n vector bundle with typical fibre V. An E-orientation or E-**Thom class** is a cohomology class $u \in \widetilde{E}^n(\mathsf{Th}(V))$ that restricts to a generator on every fibre of V.

@@ COMPLETE @@

34.10 Morse theory

34.10.1 Morse functions

Definition 34.10.1 (Nondegeneracy). At a critical point $p \in M$, the Hessian of a smooth function $f : M \to \mathbb{R}$, in local coordinates, gives a well-defined quadratic form. A critical point is said to be nondegenerate if the Hessian is nonsingular there.

Definition 34.10.2 (Morse function). Let M be a smooth manifold. A smooth function is called a Morse function if it has no degenerate critical points.

Property 34.10.3 (Density). The set of Morse functions is open and dense in the C^2-topology (see Section 40.2 on *jet spaces*).

Definition 34.10.4 (Palais–Smale condition). A smooth function $f \in C^1(M)$ is said to satisfy the Palais–Smale condition if every sequence $(x_n)_{n\in\mathbb{N}} \subset M$ with

1. $|f(x_n)|$ bounded for all $n \in \mathbb{N}$, and
2. $\|Df(x_n)\| \longrightarrow 0$

contains a convergent subsequence. It is clear that every smooth function on a compact manifold and every proper function (Definition 6.5.15) satisfies this condition.

Corollary 34.10.5. If $f \in C^1(M)$ is Morse and satisfies the Palais–Smale condition, it has only finitely many critical points in every bounded subset or in any set where f is bounded.

Definition 34.10.6 (Morse index). Consider a Morse function $f \in C^\infty(M)$. The number of negative eigenvalues at a critical point $p \in M$ is called the (Morse) index of f at p. This is often denoted by $\lambda_p(f)$.

[9] 'Multiplicative' indicates that there exists a cup product such that every group $E^\bullet(M)$ becomes a graded ring.

To any Morse function one can associate a series called the **Morse counting-series**:

$$M_t(f) := \sum_{p \in \mathrm{crit}(f)} t^{\lambda_p(f)} . \tag{34.150}$$

If M is compact, the nondegeneracy condition implies that the above sum only has a finite number of terms.

Property 34.10.7 (Morse lemma). Consider a Morse function $f : M \to \mathbb{R}$ and let $p \in M$ be a nondegenerate critical point of f. There exists a chart $(U, x_1, \dots, x_n)$ around p such that $x_i(p) = 0$ and

$$f|_U(x) = f(p) - x_1^2 - \cdots + x_k^2 + \cdots , \tag{34.151}$$

where $k = \lambda_p(f)$ is the Morse index of f.

Corollary 34.10.8. The critical points of a Morse function are isolated.

Remark 34.10.9 (Morse–Palais lemma). The Morse lemma can be generalized to open subsets of Banach spaces (and thus to infinite-dimensional manifolds).

Definition 34.10.10 (Self-indexing function). A Morse function whose value at every critical points is equal to its index.

34.10.2 Morse–Bott functions

By the Morse lemma, the critical points of a Morse function are isolated. When this condition is relaxed, a more general class of functions is obtained. Here, it is assumed that M comes equipped with a covariant derivative (Section 34.7).

Definition 34.10.11 (Morse–Bott function). A smooth function $f : M \to \mathbb{R}$ for which the critical set $\mathrm{Crit}(f)$ is a submanifold of M and at every point $p \in \mathrm{Crit}(f)$ the tangent space is the kernel of the Hessian of f, i.e. its Hessian is nondegenerate in the normal directions at every critical point.

34.10.3 Morse homology

Definition 34.10.12 (Gradientlike vector field). Consider a Morse function $f \in C^\infty(M)$. A vector field X is said to be gradientlike with respect to f if is satisfies the following conditions:

1. For all $p \notin \mathrm{Crit}(f) : X|_p(f) > 0$.
2. For all $p \in \mathrm{Crit}(f)$, there exists a Morse chart containing p such that

$$X = -2 \sum_{i=1}^{\lambda_p(f)} x^i \partial_i + 2 \sum_{i=\lambda_p(f)+1}^{\dim(M)} x^i \partial_i . \tag{34.152}$$

Its flow lines have the same orientation away from critical points and it coincides with the gradient at critical points. Furthermore, such vector fields always exist.

Property 34.10.13. Let $f \in C^\infty(M)$ be a Morse function on a compact[10] manifold and consider a gradientlike vector field X with respect to f. The integral curves of X are complete and the limits are critical points of f.

Definition 34.10.14 (Stable and unstable manifold). Let $f \in C^\infty(M)$ be a Morse function and consider a gradientlike vector field X (with respect to f). For every critical point p of f, the stable and unstable manifold of X are defined as follows:

$$W_p^\pm(X) := \left\{ x \in M \mid \lim_{t \to \pm\infty} \Phi_t(x) = p \right\}, \tag{34.153}$$

where Φ_t denotes the flow of $-X$. These sets carry the structure of smooth manifolds that are locally diffeomorphic to $\mathbb{R}^{\dim(M)-\lambda_p(f)}$ and $\mathbb{R}^{\lambda_p(f)}$, respectively.

Definition 34.10.15 (Morse–Smale pair). Let $f \in C^\infty(M)$ be a Morse function and consider a gradientlike vector field X with respect to f. If, for all critical points $p, q \in \mathrm{Crit}(f)$, one has that

$$W_p^+(X) \pitchfork W_q^-(X), \tag{34.154}$$

the pair (f, X) is called a Morse–Smale pair.

Property 34.10.16. If M is compact, there exists a self-indexing Morse–Smale pair.

Property 34.10.17. For every Morse function on a compact manifold, there exists a generic metric such that $(f, \nabla f)$ is Morse–Smale.

From here on, it will be assumed that, given a Morse function $f \in C^\infty(M)$, the pair $(f, \nabla f)$ is Morse–Smale. Let $\mathcal{M}(p, q)$ denote the set of integral curves of $-\nabla f$ that start at p and end at q, i.e. the integral curves γ that satisfy

$$\gamma([0,1]) \subset W_p^-(\nabla f) \cap W_q^+(\nabla f). \tag{34.155}$$

By the structure of the stable and unstable manifolds, this solution space has dimension $\lambda_p(f) - \lambda_q(f)$. Integral curves can be arbitrarily reparametrized. To obtain a well-defined moduli space $\overline{\mathcal{M}}(p, q)$, this $\mathbb{R}$-action is quotiented out (it is free and proper, so the resulting space is again a smooth manifold).

Definition 34.10.18 (Morse homology). The chain groups are defined as follows

$$CM_k(M, f) := \bigoplus_{\substack{p \in \mathrm{Crit}(f) \\ \lambda_p(f) = k}} \mathbb{Z}\langle p \rangle. \tag{34.156}$$

[10] This property also holds more generally, but then one has to restrict to integral curves γ whose image $f(\gamma)$ is bounded.

For critical points $p, q \in \mathrm{Crit}(f)$ such that $\lambda_p(f) = \lambda_q(f)+1$, the moduli space is a discrete, compact set. This allows to define the boundary operator as follows:

$$\partial p := \sum_{\substack{q \in \mathrm{Crit}(f) \\ \lambda_q(f) = \lambda_p(f)-1}} \left|\overline{\mathcal{M}}(p,q)\right| \langle q \rangle \,. \tag{34.157}$$

One can show that $\partial^2 = 0$. Morse homology is defined as the homology of this complex:

$$HM_\bullet(M, f) := \frac{\ker(\partial)}{\mathrm{im}(\partial)} \,. \tag{34.158}$$

34.11 Differential operators

In this section, the study of PDEs is generalized to vector bundles. The case of PDEs on $\mathbb{R}^n$ was treated in Chapter 21.

Definition 34.11.1 (Differential operator). A (linear) differential operators between two vector bundels E, F over the same base manifold M is a linear map $D : \Gamma(E) \to \Gamma(F)$ that can locally be expressed as a system of partial differential equations. The principal symbols on different charts glue together to give a globally defined morphism

$$\sigma_D : T^*M \to \mathrm{Hom}(\pi^*E, \pi^*F) \,, \tag{34.159}$$

where $\pi : T^*M \to M$ is the cotangent bundle projection.

By extension of the ordinary theory of PDEs, one says that the differential operator is **elliptic** (hyperbolic, ...) if its associated PDE is elliptic (hyperbolic, ...). E.g. a differential operator is said to be hyperbolic if the the above morphism is invertible for all cotangent vectors.

Definition 34.11.2 (Normally hyperbolic operators). Consider a *(pseudo)Riemannian* vector bundle E (see Chapter 36). A linear differential operator on E is said to be normally hyperbolic if its principal symbol is proportional to the given metric.

34.11.1 Elliptic complexes

Definition 34.11.3 (Elliptic complex). Consider a sequence of differential operators $\left(D_n : C^\infty(E_n) \to C^\infty(E_{n+1})\right)_{n\in\mathbb{N}}$ on a collection of vector bundles $\{\pi_n : E_n \to M\}_{n\in\mathbb{N}}$. This system is called an elliptic complex if it is a cochain complex, i.e. $D_{n+1} \circ D_n = 0$, and if the induced sequence $\left(\sigma_p(\xi)(D_n)\right)_{n\in\mathbb{N}}$ is exact for all $x \in M$ and $\xi \neq 0$.

@@ COMPLETE @@

Chapter 35

Principal Bundles

The main reference for this chapter is Sontz (2015a). Ths chapter uses the language of (Lie) group theory quite heavily. For all things related to group theory, the reader is referred to Section 2.2 and Section 2.3. For more information on Lie groups and their associated Lie algebras, the reader is referred to Chapter 32.

Chapter Contents

35.1 Principal bundles **653**
35.1.1 Associated bundles 654
35.1.2 Sections 656
35.2 Universal bundle **657**
35.3 Connections **658**
35.3.1 Vertical vectors 658
35.3.2 Ehresmann connections 660
35.3.3 Connection forms 661
35.3.4 Local representations 663
35.3.5 Parallel transport 664
35.3.6 Holonomy group 667
35.4 Covariant derivatives **668**
35.4.1 Koszul connections 668
35.4.2 Exterior covariant derivative 669
35.4.3 Curvature 671
35.4.4 Torsion 672
35.5 Reduction of the structure group **673**
35.6 Characteristic classes **675**

35.6.1 Chern–Weil theory 676
35.6.2 Complex bundles 677
35.6.3 Real bundles 679
35.6.4 Lie group cohomology 680
35.6.5 Chern–Simons forms 680
35.7 Differential cohomology ♣ **681**
35.7.1 Differential characters 681
35.7.2 Combining singular and de Rham cohomology 682
35.7.3 Deligne cohomology 683
35.8 Cartan geometry **685**
35.8.1 Klein geometry 685
35.8.2 Cartan geometry 686
35.9 $\mathcal{D}$-geometry ♣ **688**

35.1 Principal bundles

Definition 35.1.1 (Principal bundle). A fibre bundle $\pi : P \to B$ equipped with a right action $\rho : P \times G \to B$ that satisfies two properties:

1. **Free action**: ρ is free. This implies that the orbits are isomorphic to the structure group.
2. **Fibrewise transitivity**: The action preserves fibres, i.e. $y \cdot g \in F_b$ for all $y \in F_b, g \in G$. In turn this implies that the fibres over B are exactly the orbits of ρ.

Together these properties imply that the typical fibre F and structure group G can be identified. The right action of G on P will often be denoted by R_g (unless this would give conflicts with the same notation for the action of G on itself).

Remark 35.1.2 (G-torsor). Although the fibres are homeomorphic to G, they do not carry a group structure due to the lack of a distinct identity element. This turns them into G-torsors (Definition 2.3.16). However, it is possible to locally endow the fibres with a group structure by choosing an element of every fibre to be the identity element.

Property 35.1.3. A corollary of the definition is that the bundle $\pi : P \to B$ is isomorphic to the bundle $\xi : P \to P/G$, where P/G denotes the orbit space of P with respect to the G-action (which can be proven to be proper) and ξ is the quotient projection.

In fact this property can be used to give an alternative characterization of smooth principal bundles.

Property 35.1.4 (Quotient manifold theorem). Consider a smooth manifold P equipped with a free and proper (right) action of a Lie group G. The following statements hold:

- The orbit space P/G is a smooth manifold.
- The projection $P \to P/G$ is a submersion.
- P is principal G-bundle over P/G.

Property 35.1.5 (Dimension). The dimension of P is given by

$$\dim(P) = \dim(B) + \dim(G)\,. \tag{35.1}$$

Property 35.1.6. Every local trivialization φ_i is G-equivariant:

$$\varphi_i(z \cdot g) = \varphi_i(z) \cdot g\,. \tag{35.2}$$

Definition 35.1.7 (Principal bundle map). A bundle map between principal G-bundles is a pair of morphisms (f_B, f_P) such that:

1. (f_B, f_P) is an ordinary bundle map (Definition 33.1.10).
2. f_P is G-equivariant.

The following property proves that the equivariance condition on principal bundle maps is, in fact, a very strong condition.

Property 35.1.8. Every principal bundle map covering the identity is an isomorphism.

Definition 35.1.9 (Vertical automorphism). Consider a principal G-bundle $\pi : P \to B$. An automorphism f of this bundle is said to be vertical if it covers the identity, i.e. $\pi \circ f = \pi$. It is the subgroup $\mathrm{Aut}_V(P) \subset \mathrm{Aut}(P)$ of vertical automorphisms that is known as the **group of gauge transformations** or **gauge group**[1] in physics.

Remark. It should be clear that the above definition can easily be generalized to arbitrary fibre bundles.

35.1.1 Associated bundles

Construction 35.1.10 (Associated principal bundle). For every fibre bundle, one can construct an associated principal G-bundle by replacing the fibre F by G itself using the fibre bundle construction theorem 33.2.1, where the left action of G is given by left multiplication in G.

Property 35.1.11 (Triviality). A fibre bundle ξ is trivial if and only if its associated principal bundle is trivial. More generally, two fibre bundles are isomorphic if and only if their associated principal bundles are isomorphic.

[1] This should not be confused with the structure group G, which is also sometimes called the gauge group in physics.

Example 35.1.12 (Frame bundle). Let V be an n-dimensional vector space and denote the set of frames (Definition 22.1.6) of V by FV. It follows from the fact that every basis transformation is given by the action of an element of the general linear group that FV is isomorphic to $\mathsf{GL}(V) \cong \mathsf{GL}(\mathbb{R}^n)$.

Given a rank-n vector bundle E, one can construct an associated principal bundle by replacing every fibre $\pi^{-1}(b)$ by $F\big(\pi^{-1}(b)\big) \cong \mathsf{GL}(\mathbb{R}^n)$. The right action on this bundle by $g \in \mathsf{GL}(\mathbb{R}^n)$ is given by the basis transformation $\tilde{e}_j = g^i_j e_i$. This bundle, the frame bundle, is denoted by FE or FM in the case of the tangent bundle $E = TM$.

Construction 35.1.13 (Associated bundle to a principal bundle). Consider a principal G-bundle $\pi : P \to B$ and let F be a space equipped with a left G-action $\triangleright$. One can construct an associated bundle $P_F \equiv P \times_\triangleright F$ in the following way:

1. Define an equivalence relation $\sim_G$ on the product space $P \times F$ by
$$(p, f) \sim_G (p', f') \iff \exists g \in G : (p', f') = (p \cdot g, g^{-1} \triangleright f)\,. \tag{35.3}$$

2. Define the total space of the associated bundle as the following quotient space:
$$P_F := (P \times F)/\sim_G\,. \tag{35.4}$$

3. Define the projection $\pi_F : P_F \to B$ as follows:
$$\pi_F : [p, f] \mapsto \pi(p)\,, \tag{35.5}$$
where $[p, f]$ is the equivalence class of $(p, f) \in P \times F$ in P_F.

Example 35.1.14 (Tangent bundle). Starting from the frame bundle FM over a manifold M, one can reconstruct, up to a bundle isomorphism, the tangent bundle TM in the following way. Consider the left G-action $\triangleright$ of a matrix group given by
$$\triangleright : G \times \mathbb{R}^n \to \mathbb{R}^n : (g \triangleright f)^i \mapsto g^i{}_j f^j\,. \tag{35.6}$$
The tangent bundle is isomorphic to the associated bundle $FM \times_\triangleright \mathbb{R}^n$, where the bundle map is defined as $[e, v] \mapsto v^i e_i \in TM$.

Example 35.1.15 (Adjoint bundle). Consider a principal G-bundle P. G acts on itself by conjugation through the adjoint action
$$\mathsf{Ad} : G \times G \to G : (g, h) \mapsto g^{-1} h g\,. \tag{35.7}$$
This action induces an associated bundle $\mathsf{Ad}(P) := P \times_G G$, suitable named the adjoint bundle of P.

Property 35.1.16 (Vertical automorphisms). There exists an isomorphism between the vertical automorphism group $\mathsf{Aut}_V(P)$ and the group of sections of the adjoint bundle $\mathsf{Ad}(P)$.

Construction 35.1.17 (Associated bundle map). Given a principal bundle map (f_B, f_P) between two principal bundles one can construct an associated bundle map between any two of their associated bundles with the same typical fibre in the following way:

- The total space map $\tilde{f}_P : P \times_G F \to P' \times_{G'} F$ is given by
$$\tilde{f}_P([p, f]) := [f_P(p), f]\,. \tag{35.8}$$
- The base space map is simply given by f_B itself:
$$\tilde{f}_B(b) = f_B(b)\,. \tag{35.9}$$

35.1.2 Sections

Although every vector bundle has at least one global section, the zero section, a general principal bundle does not necessarily have a global section. This is made clear by the following property.

Property 35.1.18 (Trivial bundles). A principal G-bundle P is trivial if and only if there exists a global section of P. Furthermore, there exists a bijection between the set of global sections $\Gamma(P)$ and the set of trivializations $\mathsf{Triv}(P)$.

Corollary 35.1.19. Every local section $\sigma : U \to P$ induces a local trivialization φ by
$$\varphi^{-1} : (m, g) \mapsto \sigma(m) \cdot g\,. \tag{35.10}$$
The converse is also true. Consider a local trivialization $\psi^{-1} : U \times G \to \pi^{-1}(U)$. A local section can be obtained by taking $\sigma(u) := \psi^{-1}(u, e)$.

Property 34.2.6 can now be reformulated as follows.

Property 35.1.20 (Trivial vector bundles). A vector bundle is trivial if and only if its associated frame bundle admits a global section. This can easily be interpreted as follows. If one can for every fibre choose a basis in a smooth way, one can also express the restriction of any vector field to a fibre in terms of this basis in a smooth way.

Property 35.1.21 (Higgs fields). Let $\pi : P \to M$ be a principal G-bundle and let P_F be an associated bundle. There exists a bijection between the sections of P_F and the G-equivariant maps $\phi : P \to F$, i.e. maps satisfying $\phi(p \cdot g) = g^{-1} \cdot \phi(p)$.

An explicit correspondence is given by
$$\sigma_\phi : M \to P_F : m \mapsto [p, \phi(p)]\,, \tag{35.11}$$
where p is any point in $\pi^{-1}(\{m\})$. This is well-defined due to Eq. (35.3). In the other direction, one finds
$$\phi_\sigma : P \to F : p \mapsto j_p^{-1} \circ \sigma(\pi(p))\,, \tag{35.12}$$
where $j_p : F \to P_F : f \mapsto [p, f]$. Either of these maps is sometimes called a **Higgs field** in the physics literature.

35.2 Universal bundle

Definition 35.2.1 (Universal bundle). Consider a topological group G. A universal bundle of G is any principal bundle of the form

$$G \hookrightarrow EG \to BG$$

such that EG is weakly contractible. The space BG is called the **classifying space** of G.

Definition 35.2.2 (n-universal bundle). A principal bundle with an $(n-1)$-connected total space.

Property 35.2.3 (Delooping). For every topological group G, one can prove that the loop space of BG is (weakly) homotopy equivalent to G itself, i.e. $\Omega BG \cong G$. As such, it also deserves the name of delooping.

Property 35.2.4 (Groups). Let G be a group regarded as a discrete topological space. Because the fundamental group of a topological group is Abelian by Property 7.1.16, the classifying space BG is a group if and only if G is Abelian.

This also has an abstract nonsense generalization. The classifying space functor B : **TopGrp** $\to$ **Top** is product-preserving and, hence, it maps group objects to group objects. So, Abelian groups are mapped to topological groups and, even better, to Abelian groups. An important consequence is that all Abelian topological groups are in particular infinite loop spaces.

Property 35.2.5 (Classification). The collection of principal G-bundles over a paracompact Hausdorff space X is in bijection with $[X, BG]$, the set of homotopy classes of continuous functions $f : X \to BG$. This bijection is given by the pullback-construction $f \mapsto f^*EG$.

Due to the homotopical nature of this classification one can also replace G by any homotopy equivalent space. For Lie groups the natural choice is a *maximal compact subgroup* since these are deformation retracts and, hence, homotopy equivalent.

Corollary 35.2.6 (Vector bundles). Since every vector bundle is uniquely related to its frame bundle, there exists a bijection between principal GL-bundles and vector bundles. This implies that rank-k vector bundles are classified by the homotopy mapping space $[X, B\mathsf{GL}(k)]$. Because $\mathsf{O}(k)$ is the maximal compact subgroup of $\mathsf{GL}(k)$, one also obtains the result that any real vector bundle over a paracompact space admits a *Riemannian structure* (see Chapter 36).

Property 34.2.4 now follows from Eckmann–Hilton duality (Property 7.1.42) together with the above delooping property.

Remark 35.2.7. There also exists a slightly different notion of universal bundles and their associated classifying property. When one requires the total space of the universal

bundle to be contractible instead of weakly contractible, the mapping space only classifies numerable principal bundles (Definition 33.1.8), but now over arbitrary bases.

An explicit construction of the numerable universal bundle for any topological group G was given by *Milnor*.

Construction 35.2.8 (Milnor ♣). First, consider the infinite join E_∞ equipped with the strong topology. This space is constructed as the direct limit of finite joins (Construction 6.3.10):

$$E_n = \underbrace{G \circ \cdots \circ G}_{n \text{ times}},$$

where E_n is embedded in E_{n+1} using the identity element, i.e. every element of E_∞ corresponds to an element of some finite join. Then, construct the quotient of E_n (resp. E_∞) by the canonical right action of G on E_n (resp. E_∞). The bundle $p_n : E_n \to B_n$ (resp. $p : E_\infty \to B_\infty$) is an n-universal bundle (resp. ∞-universal bundle). It follows from the above property that $p : E_\infty \to B_\infty$ is a universal bundle for G.

Construction 35.2.9 (Abstract nonsense ♣). Let G be a topological group and consider the delooping (groupoid) $\mathbf{B}G$ from Example 3.5.11. This groupoid can also be obtained as the action groupoid (Example 3.5.10) associated to the trivial action of G on $\{*\}$. The regular action of G on itself also induces an action groupoid $\mathbf{E}G := G//G$. The map $G \to \{*\}$ in turn induces a map of groupoids $\mathbf{E}G \to \mathbf{B}G$, which, under geometric realization as in Property 13.1.14, gives a universal bundle map.

35.3 Connections

35.3.1 Vertical vectors

Because smooth fibre bundles are also smooth manifolds, one can define traditional concepts such as the tangent bundle. Due to the composite nature of these geometric objects, one can decompose the tangent bundle in horizontal and vertical (sub)bundles.

Definition 35.3.1 (Vertical vector). Let $\pi : E \to M$ be a smooth fibre bundle. The subbundle

$$VE := \ker(\pi_*) \tag{35.13}$$

of TE is called the **vertical (sub)bundle** over E. The sections of this bundle are called vertical vector fields.

For principal G-bundles an alternative definition exists.

Alternative Definition 35.3.2. Consider a smooth principal G-bundle $\pi : P \to M$. First, construct a map ι_p for every element $p \in P$:

$$\iota_p : G \to P : g \mapsto p \cdot g\,. \tag{35.14}$$

Then, define a tangent vector $v \in T_pP$ to be vertical if it lies in the image of $\iota_{p,*}$:

$$V_pP := \mathrm{im}(\iota_{p,*})\,. \tag{35.15}$$

This definition is equivalent to the previous one because of the following short exact sequence:

$$0 \longrightarrow \mathfrak{g} \xrightarrow{\iota_{p,*}} T_pP \xrightarrow{\pi_*} T_pM \longrightarrow 0\,. \tag{35.16}$$

Property 35.3.3 (Dimension). It follows from the second definition that the vertical vectors of a principal G-bundle are nothing but the pushforward of the Lie algebra $\mathfrak{g}$ under the right action of G on P. Furthermore, the exactness of the sequence implies that $\iota_{p,*} : \mathfrak{g} \to V_pP$ is an isomorphism of vector spaces. In particular, it implies that

$$\dim(V_pP) = \dim(\mathfrak{g}) = \dim(G)\,. \tag{35.17}$$

Definition 35.3.4 (Fundamental vector field). Let P be a principal G-bundle and consider $A \in \mathfrak{g}$, where $\mathfrak{g}$ is the Lie algebra corresponding to G. The vertical vector field $A^\# : P \to TP$, given by

$$A^\#(p) := \iota_{p,*}A \in VpP\,, \tag{35.18}$$

is called the fundamental vector field associated to A. Its action on any $f \in C^\infty(P)$ is given by

$$A^\#_p(f) = \frac{\mathrm{d}}{\mathrm{d}t} f\big(p \cdot \exp(tA)\big)\Big|_{t=0}\,. \tag{35.19}$$

Property 35.3.5. The construction of fundamental vector fields gives a Lie algebra morphism $\mathfrak{g} \to \mathfrak{X}(P)$:

$$[A, B]^\# = [A^\#, B^\#]\,, \tag{35.20}$$

where the Lie bracket on the left is the one in $\mathfrak{g}$ and the Lie bracket on the right is the one in $\mathfrak{X}(P)$ given by Property 34.3.20.

Property 35.3.6. The vertical bundle satisfies the following equivariance condition:

$$R_{g,*}(VpP) = V_{p\cdot g}P\,. \tag{35.21}$$

By differentiating the equality

$$R_g \circ \iota_p = \iota_{p\cdot g} \circ \mathrm{ad}_{g^{-1}} \tag{35.22}$$

and using Example 32.3.2 and Definition 35.3.4, one can obtain the following algebraic reformulation:

$$R_{g,*}\big(A^\#(p)\big) = \big(\mathrm{Ad}_{g^{-1}}A\big)^\#(p \cdot g)\,. \tag{35.23}$$

35.3.2 Ehresmann connections

The definition of vertical vector fields was quite natural. The next step would be to define the horizontal subspace as a complementary subspace to VP. However, the exact sequence (35.16) does not even split canonically, so one should make a choice of splitting.

Definition 35.3.7 (Ehresmann connection). Consider a smooth fibre bundle E. An (Ehresmann) connection on E is the selection of a subspace $H_eE \leq T_eE$ for every $e \in E$ such that:

1. The horizontal and vertical bundles are complementary: $V_eE \oplus H_eE = T_eE$.
2. The choice of subspace depends smoothly on $e \in E$ in the distributional sense (Definition 34.6.1).

The vectors in H_eE are said to be **horizontal** (with respect to the chosen connection).

Definition 35.3.8 (Horizontal bundle). The horizontal (sub)bundle HE is defined as $\bigsqcup_{e\in E} H_eE$ with the bundle structure induced from TE.

Definition 35.3.9 (Principal connection). A principal connection on a smooth principal G-bundle P is a G-equivariant Ehresmann connection, i.e. an Ehresmann connection for which the horizontal subspaces satisfy the following G-equivariance condition:

$$R_{g,*}(H_pP) = H_{p\cdot g}P\,. \tag{35.24}$$

Remark. Note that this condition was automatically satisfied for vertical bundles by Property 35.3.6.

Property 35.3.10 (Dimension). Properties 35.1.5 and 35.3.3, together with the direct sum decomposition of TP, imply the following relation for all $p \in P$:

$$\dim(H_pP) = \dim(M)\,. \tag{35.25}$$

All dimensional relations between the data of a principal bundle are now summarized:

$$\begin{aligned} \dim(P) &= \dim(M) + \dim(G) \\ \dim(M) &= \dim(H_pP) \\ \dim(G) &= \dim(V_pP) \end{aligned} \tag{35.26}$$

for all $p \in P$.

Definition 35.3.11 (Dual connection). First, define the dual of the horizontal bundle:

$$H_p^*P := \left\{h \in T_p^*P \mid \forall v \in V_pP : h(v) = 0\right\}. \tag{35.27}$$

It is the space of 1-forms that vanish on the vertical subspace. A dual connection can then be defined as the selection of a vertical covector bundle V_p^*P satisfying the conditions of Definitions 35.3.7 and 35.3.9, where V and H are to be interchanged. Note

that, here, the horizontal subbundle is canonically defined while the vertical subbundle depends on a choice of complement.

This definition implies the following ones.

Definition 35.3.12 (Horizontal and vertical forms). Let $\theta \in \Omega^k(P)$ be a differential k-form.

- θ is said to be horizontal if

$$\theta_p(v_1, \dots, v_k) = 0 \tag{35.28}$$

  whenever at least one of the v_i is in V_pP.

- θ is said to be vertical if

$$\theta_p(v_1, \dots, v_k) = 0 \tag{35.29}$$

  whenever all of the v_i are in H_pP.

For functions $f \in C^\infty(P)$, it is vacuously true that they are both vertical and horizontal.

Definition 35.3.13 (Tensorial form). Consider a vector space V equipped with a representation $\rho : G \to V$. A V-valued differential form $\theta \in \Omega^\bullet(P; V)$ on a principal G-bundle $\pi : P \to M$ is said to be **tensorial** or **basic of type** (V, ρ) if

1. it is horizontal, and
2. if it is equivariant:

$$R_g^* \theta = \rho(g^{-1}) \circ \theta\,. \tag{35.30}$$

The space of tensorial forms of type (V, ρ) is sometimes denoted by $\Omega^\bullet_{\rho,\mathrm{hor}}(P; V)$.

Forms satisfying this definition admit an important reinterpretation. Let $E := P \times_\rho V$ be the associated vector bundle of (V, ρ). Tensorial k-forms of type (V, ρ) are naturally isomorphic to E-valued k-forms on M. The isomorphism is given fibrewise by

$$\tau|_u : \Omega^\bullet|_{\pi(u)}(M; P \times_\rho V) \to \Omega^\bullet_{\rho,\mathrm{hor}}|_u(P; V) : \phi \mapsto f_u^{-1}(\pi^*\phi)\,, \tag{35.31}$$

where $f_u : V \to E_{\pi(u)} \cong (\pi^*E)_u : v \mapsto [u, v]$. This is a generalization of correspondence 35.1.21.

35.3.3 Connection forms

Definition 35.3.14 (Connection 1-form). Let P be a principal G-bundle. A connection 1-form, associated to a given principal connection, is a $\mathfrak{g}$-valued 1-form $\omega \in \Omega^1(P; \mathfrak{g})$ that satisfies the following two conditions:

1. **Cancellation**:

$$\omega(A^{\#}) = A \tag{35.32}$$

for all $A \in \mathfrak{g}$.

2. **Equivariance**:

$$\omega \circ R_{g,*} = \mathsf{Ad}_{g^{-1}} \circ \omega \tag{35.33}$$

for all $g \in G$.

The horizontal subspaces are recovered as the kernel of the connection 1-form:

$$H_p P = \ker(\omega|_p)\,. \tag{35.34}$$

Formula 35.3.15. Given a principal connection on a principal G-bundle P, the associated connection 1-form is given by the following map:

$$\omega := (\iota_{p,*})^{-1} \circ \mathsf{pr}_{\mathrm{Vert}}\,, \tag{35.35}$$

where $\mathsf{pr}_{\mathrm{Vert}}$ is the projection $TP \to VP$.

The following property mirrors Property 34.7.7 (see further below for the relation between principal connections and Koszul connections).

Property 35.3.16 (Affinity). The set of connection 1-forms is affine over $\Omega^1_{\mathsf{Ad},\mathrm{hor}}(P;\mathfrak{g}) \cong \pi^*\Omega^1(M;\mathsf{ad}(P))$.

Property 35.3.17 (Pullback connection). Consider two principal G-bundles P_1, P_2. Let ω be a connection 1-form on P_1 and let $F : P_1 \to P_2$ be a bundle map. The pullback $F^*\omega$ gives a principal connection on P_2.

Definition 35.3.18 (Maurer–Cartan form). For every $g \in G$, the tangent space T_gG is isomorphic to $T_eG \equiv \mathfrak{g}$. A canonical isomorphism $T_gG \to \mathfrak{g}$ is given by the Maurer–Cartan form

$$\Omega|_g := L_{g^{-1},*}\,. \tag{35.36}$$

Construction 35.3.19. Consider the one-point manifold $M = \{*\}$. When constructing a principal G-bundle over M, one can see that the total space $P = \{*\} \times G$ can be identified with the structure group G. From the relations in Property 35.3.10, it follows that the horizontal spaces are null-spaces, which trivially defines a smooth distribution and a connection in the sense of Ehresmann (Definition 35.3.7), and that the vertical spaces are equal to the tangent spaces, i.e. $V_gG = T_gG$, where the identification $P \cong G$ (as manifolds) is used.

The simplest way to define a connection form on this bundle would be the trivial projection $\mathbb{1}_{TP} : TP \to TP = VP$. However, the image of this map would be $T_g G$ and not $\mathfrak{g}$ as required. This can be solved by using the Maurer–Cartan form:

$$\omega = \Omega\,. \tag{35.37}$$

Property 35.3.20 (Uniqueness). The Maurer–Cartan form is the unique principal connection on the bundle $G \hookrightarrow G \to \{*\}$.

Definition 35.3.21 (Darboux derivative). Consider a smooth function $f : M \to G$ between a manifold and a Lie group. The Darboux derivative of f is defined as follows:

$$\omega_f := f^*\Omega\,. \tag{35.38}$$

The function f is called an **integral** or **primitive** of ω_f.

Property 35.3.22. Let M be a connected manifold. If two functions $f, \tilde{f} : M \to G$ have the same Darboux derivative, there exists an element $g \in G$ such that $f(p) = g \cdot \tilde{f}(p)$ for all $p \in M$.

Example 35.3.23 (Real line). In the case of $M = G = \mathbb{R}$, one recovers the ordinary behaviour of derivatives. When two functions $f, \tilde{f} : \mathbb{R} \to \mathbb{R}$ have the same derivative, they are equal up to an additive constant.

Theorem 35.3.24 (Fundamental theorem of calculus). *Consider a smooth manifold M and a Lie group G with Lie algebra $\mathfrak{g}$. If $\omega : TM \to \mathfrak{g}$ satisfies the Maurer–Cartan equation*

$$\mathrm{d}\omega + \frac{1}{2}[\omega \wedge \omega] = 0\,, \tag{35.39}$$

there exists (locally) a smooth function $f : M \to G$ such that $\omega = f^*\Omega$.

35.3.4 Local representations

Definition 35.3.25 (Yang–Mills field). Consider a principal G-bundle $\pi : P \to M$ and an open subset $U \subseteq M$. Given a principal connection ω on P and a local section $\sigma : U \to P$, the Yang–Mills field $\omega^U \in \Omega^1(U; \mathfrak{g})$ is defined as follows:

$$\omega^U := \sigma^*\omega\,. \tag{35.40}$$

Definition 35.3.26 (Local representation). Consider a principal bundle P and let (U, φ) be a bundle chart on P. The local representation of a principal connection ω on P with respect to the chart (U, φ) is defined as $(\varphi^{-1})^*\omega$.

Formula 35.3.27. Consider a principal connection ω on a principal G-bundle $\pi : P \to M$. Because of Corollary 35.1.19, every local section $\sigma : U \to P$ induces both a Yang–Mills field ω^U and a local representation of ω. These two forms are related by the following equation:

$$\omega|_{(m,g)}(v, A) = \mathrm{Ad}_{g^{-1}}\left(\omega^U_m(v)\right) + \Omega_g(A)\,, \tag{35.41}$$

where $v \in T_m U$ and $A \in \mathfrak{g}$.

Formula 35.3.28 (Compatibility condition). Consider a principal G-bundle $\pi : P \to M$ and two open subsets $U, V \subseteq M$. Given two local sections $\sigma_U : U \to P$, $\sigma_V : V \to P$ and a principal connection ω on P, one can define two Yang–Mills field ω^U and ω^V.

On the intersection $U \cap V \subseteq M$ there exists a (unique) gauge transformation $\xi : U \cap V \to G$ such that $\sigma_V(m) = \sigma_U(m) \cdot \xi(m)$. Using this gauge transformation, one can relate ω^U and ω^V as follows:

$$\omega^V = \mathsf{Ad}_{\xi^{-1}} \circ \omega^U + \xi^* \Omega \,. \tag{35.42}$$

This formula holds more generally to (locally) relate the connection 1-forms ω and $\xi^*\omega$ for any gauge transformation $\xi \in \mathsf{Aut}_V(P)$.

Example 35.3.29 (General linear group). Consider $G = \mathsf{GL}(n, \mathbb{R})$. The second term in Eq. (35.42) can be written as follows:[2]

$$(\xi^*\Omega)^i{}_j = (\xi(m)^{-1})^i{}_k \frac{\partial}{\partial x^\mu} \xi(p)^k{}_j \mathrm{d}x^\mu \tag{35.43}$$

at every point $m \in M$. Formally, this can be written coordinate-independently as

$$\xi^*\Omega \equiv \xi^{-1} \mathrm{d}\xi \,. \tag{35.44}$$

This form is often called the **Maurer–Cartan connection**.

Example 35.3.30 (Christoffel symbols). Let $\Gamma^i{}_{j\mu}, \overline{\Gamma}^k{}_{l\nu}$ be the Yang–Mills fields corresponding to a connection on the frame bundle of a manifold M induced by the choices of coordinates x^i and y^i, respectively. In this case, the expansion coefficients of the Yang–Mills field are called the **Christoffel symbols** (compare this to Definition 34.7.8). Using Eqs. (35.42) and (35.44), one obtains:

$$\overline{\Gamma}^i{}_{j\mu} = \frac{\partial y^\nu}{\partial x^\mu} \left(\frac{\partial x^i}{\partial y^k} \Gamma^k{}_{l\nu} \frac{\partial y^l}{\partial x^j} + \frac{\partial x^i}{\partial y^k} \frac{\partial^2 y^k}{\partial x^j \partial x^\nu} \right) . \tag{35.45}$$

35.3.5 Parallel transport

Definition 35.3.31 (Horizontal lift). Consider a principal bundle $\pi : P \to M$ and a curve $\gamma : [0,1] \to M$. Given an Ehresmann connection, for every point $p_0 \in \pi^{-1}(\gamma(0))$ there exists a unique curve $\widetilde{\gamma}_{p_0} : [0,1] \to P$ satisfying the following conditions:

1. **Initial condition**: $\widetilde{\gamma}_{p_0}(0) = p_0$,
2. **Lifting**: $\pi \circ \widetilde{\gamma}_{p_0} = \gamma$, and
3. **Horizontal**: $\widetilde{\gamma}'_{p_0}(t) \in \mathsf{Hor}\left(T_{\widetilde{\gamma}_{p_0}(t)} P\right)$ for all $t \in [0,1]$.

[2] A derivation can be found in Lecture 22 of Schuller (2016).

The curve $\widetilde{\gamma}_{p_0}$ is called the horizontal lift of γ starting at p_0. When it is clear from the context what the basepoint p_0 is, the subscript is often omitted and one writes $\widetilde{\gamma}$ instead of $\widetilde{\gamma}_{p_0}$.

Remark 35.3.32 (Horizontal curve). Curves satisfying the last condition in the above property are said to be horizontal.

Method 35.3.33. Consider a principal G bundle $\pi : P \to M$. Let γ be a curve in M and let ω be a principal connection 1-form on P. The horizontal lift of γ can be found as follows. Let δ be a curve in P that projects onto γ, such that

$$\widetilde{\gamma}_{p_0}(t) = \delta(t) \cdot g(t) \tag{35.46}$$

for some curve g in G. The curve g can be found as the unique solution of the following first-order ODE:

$$\mathsf{Ad}_{g(t)^{-1}} \omega_{\delta(t)}(X_{\delta,\delta(t)}) + \Omega_{g(t)}(Y_{g,g(t)}) = 0\,, \tag{35.47}$$

where X_δ, Y_g are tangent vectors to the curves δ and g, respectively. The solution is uniquely determined by the initial value condition $\delta(0) \cdot g(0) = p_0$.

Remark 35.3.34. Given a local section $\sigma : U \to P$, one can rewrite the above ODE in a more explicit form. First, remark that the section induces a curve $\delta := \sigma \circ \gamma$. Taking the derivative yields $X_\delta = \sigma_* X_\gamma$. With this one can rewrite the ODE as

$$\mathsf{Ad}_{g(t)^{-1}} \omega_{\delta(t)}(\sigma_* X_{\gamma,\gamma(t)}) + \Omega_{g(t)}(Y_{g,g(t)}) = 0\,. \tag{35.48}$$

After using the equality $f^*\omega = \omega \circ f_*$ and introducing the Yang–Mills field $A = \sigma^*\omega$, this becomes

$$\mathsf{Ad}_{g(t)^{-1}} A(X_{\gamma,\gamma(t)}) + \Omega_{g(t)}(Y_{g,g(t)}) = 0\,. \tag{35.49}$$

Example 35.3.35. For matrix Lie groups this ODE can be reformulated as follows. Given the trivial section $s : U \to U \times G : x \mapsto (x, e)$, where U is an open subset of M, the horizontal lift of γ can locally be parametrized as

$$\widetilde{\gamma}(t) = \underbrace{(s \circ \gamma)(t)}_{\delta(t)} \cdot g(t) = \big(\gamma(t), g(t)\big)\,, \tag{35.50}$$

where g is a curve in G. To determine $\widetilde{\gamma}$ it is thus sufficient to find g. The ODE (35.47) then becomes

$$g'(t) = -\omega\big(\gamma(t), e, \gamma'(t), 0\big) g(t)\,. \tag{35.51}$$

Using the trivial section s, one can further rewrite this formula. Consider the action of the Yang–Mills field $s^*\omega$ on the derivative $\gamma_* = \big(\gamma(t), \gamma'(t)\big)$. Using the fact that it is linear in the second argument, it can be rewritten as

$$s^*\omega\big(\gamma(t), \gamma'(t)\big) = A\big(\gamma(t)\big)\gamma'(t)\,, \tag{35.52}$$

where $A : M \to \mathsf{Hom}(\mathbb{R}^{\dim(M)}, \mathfrak{g})$ gives a linear map at each point $\gamma(t) \in M$. The action can also be rewritten as

$$s^*\omega\big(\gamma(t), \gamma'(t)\big) = \omega\big(s_*(\gamma(t), \gamma'(t))\big) = \omega\big(\gamma(t), e, \gamma'(t), 0\big)\,. \tag{35.53}$$

Combining these relations with the ODE (35.51) gives

$$\left(\frac{\mathrm{d}}{\mathrm{d}t} + A(\gamma(t))\gamma'(t)\right)g(t) = 0\,, \tag{35.54}$$

where $\frac{\mathrm{d}}{\mathrm{d}t}$ is the matrix given by element-wise multiplication of the derivative $\frac{\mathrm{d}}{\mathrm{d}t}$ and the identity matrix I.

The ODE (35.47) can now be solved. Direct integration and iteration gives

$$g(t) = \left[\mathbb{1} - \int_0^t A(\gamma'(t_1))\,dt_1 + \int_0^t \int_0^{t_1} A(\gamma'(t_1))A(\gamma'(t_2))\,dt_2\,dt_1 - \cdots\right] g(0)\,, \tag{35.55}$$

where A is the Yang–Mills field associated to the local section σ. This can be rewritten using the standard *Dyson trick* (see Formula 57.2.5):

$$g(t) = \left[\mathbb{1} - \int_0^t A(\gamma'(t_1))\,dt_1 + \frac{1}{2!}\int_0^t \int_0^t \mathcal{T}\Big(A(\gamma'(t_1))A(\gamma'(t_2))\Big)\,dt_2\,dt_1 - \cdots\right] g(0). \tag{35.56}$$

By noting that this formula is equal to the path-ordered exponential, one finds

$$g(t) = \mathcal{T}\exp\left(-\int_0^t A(\gamma'(t'))\,dt'\right) g(0)\,. \tag{35.57}$$

Definition 35.3.36 (Parallel transport). The parallel transport map along the curve γ is defined as follows:

$$\mathsf{Par}_t^\gamma : \pi^{-1}(\gamma(0)) \to \pi^{-1}(\gamma(t)) : p_0 \mapsto \widetilde{\gamma}_{p_0}(t)\,. \tag{35.58}$$

This map is G-equivariant and it restricts to an isomorphism on the fibres. The group element given by the path-ordered exponential in Eq. (35.57) is called the **holonomy** along the curve γ.

Using the above constructions that assign Lie group elements to paths, one can give an alternative definition of principal connections.

Alternative Definition 35.3.37 (Principal connection ♣). Let M be a smooth manifold and consider its path groupoid[3] $\mathcal{P}_1(M)$ which has the points of M as objects and homotopy classes of smooth paths in M as morphisms. Consider a principal G-bundle

[3] See Definition 44.7.14 for a rigorous exposition.

$\pi : P \to M$ and denote the delooping (Example 3.5.11) of G by $\mathbf{B}G$. The assignment of holonomies to smooth paths defines a functor

$$\mathsf{hol}_i : \mathcal{P}_1(U_i) \to \mathbf{B}G \tag{35.59}$$

for every chart $U_i \subseteq M$. These can be glued together using the transition cocycles g_{ij} (in their incarnation as natural isomorphisms) to obtain a functor

$$\mathsf{hol} : \mathcal{P}_1(M) \to \mathbf{Trans}_1(P) \subset G\mathbf{Torsor}\,, \tag{35.60}$$

where $\mathbf{Trans}_1(P)$ is the full subcategory of the category of G-torsors on the fibres of P (Remark 35.1.2).

It can be shown that any functor of this type gives rise to a principal connection on P and, conversely, every principal connection gives rise to a holonomy functor through the parallel transport construction.

35.3.6 Holonomy group

Definition 35.3.38 (Holonomy group). Consider a principal G-bundle $\pi : P \to M$ with principal connection ω and choose a point $m \in M$. Let $\Omega_m^{\mathrm{ps}} M \subset \Omega_m M$ denote the subset of the based loop space consisting of piecewise smooth loops with base-point $m \in M$. The holonomy group $\mathsf{Hol}_p(\omega)$ based at $p \in \pi^{-1}(m)$ with respect to ω is given by

$$\mathsf{Hol}_p(\omega) := \{g \in G \mid p \sim p \cdot g\}\,, \tag{35.61}$$

where two points $p, q \in P$ are identified if there exists a loop $\gamma \in \Omega_m^{\mathrm{ps}} M$ such that the horizontal lift $\widetilde{\gamma}$ connects p and q.

Definition 35.3.39 (Reduced holonomy group). The subgroup of the holonomy group generated by contractible loops.

Definition 35.3.40 (Holonomy bundle). Let M be a path-connected manifold and consider a principal bundle P over M with principal connection ω. One can equip P with an equivalence relation $\sim$ such that $p \sim q$ if and only if there exists a horizontal curve connecting p and q. For every point $p \in P$ one can then construct the following set:

$$H(p) := \{q \in P \mid p \sim q\}\,. \tag{35.62}$$

Path-connectedness of the base manifold implies that $H(p)$ and $H(q)$ are isomorphic for all $p, q \in P$. Using this fact one can show that $\sqcup_{p \in P} H(p)$ is in fact a principal bundle itself. Its structure group is $\mathsf{Hol}_p(\omega)$ for any $p \in P$.

35.4 Covariant derivatives

35.4.1 Koszul connections

Definition 35.4.1 (Horizontal lifts on associated bundles). Let $P_F \equiv P \times_G F$ be an associated bundle of a principal G-bundle $\pi : P \to M$ and let γ be a curve in M with horizontal lift $\widetilde{\gamma}_p$ in P. The horizontal lift of γ to P_F through the point $[p, f] \in P_F$ is defined as follows:

$$\widetilde{\gamma}^{P_F}_{[p,f]}(t) := [\widetilde{\gamma}_p(t), f]\,. \tag{35.63}$$

Note that, although the element f seems to stay constant along the horizontal lift, it in fact changes according to Eq. (35.3).

Definition 35.4.2 (Parallel transport). Similar to the case of principal bundles P, the parallel transport map on an associated bundle P_F is defined as

$$\mathsf{Par}^\gamma_t : \pi_F^{-1}(\gamma(0)) \to \pi_F^{-1}(\gamma(t)) : [p, f] \mapsto \widetilde{\gamma}^{P_F}_{[p,f]}(t)\,. \tag{35.64}$$

Example 35.4.3 (Vector bundles). Consider a vector space V with a representation $\rho : G \to \mathsf{GL}(V)$, a principal G-bundle $\pi : P \to M$ and the associated vector bundle $\pi_V : P \times_{\mathsf{GL}(V)} V \to M$. By working over a chart (U, φ), one can locally treat P and P_V as product bundles. Parallel transport on this vector bundle is then defined as follows. Let γ be a curve in M such that $\gamma(0) = x_0$ and $\gamma(1) = x_1$. Furthermore, let the horizontal lift $\widetilde{\gamma}(t) = \big(\gamma(t), g(t)\big)$ satisfy $\widetilde{\gamma}(0) = (x_0, h)$ as initial condition. Parallel transport of the point $(x_0, v_0) \in U \times V$ along γ is given by the following map:

$$\mathsf{Par}^\gamma_t : \pi_V^{-1}(x_0) \to \pi_V^{-1}(\gamma(t)) : (x_0, v_0) \mapsto \big(\gamma(t), \rho\big(g(t)h^{-1}\big)v_0\big)\,. \tag{35.65}$$

It should be noted that this map is independent of the initial element $h \in G$ despite the presence of the factor h^{-1}. Moreover, Par^γ_t is an isomorphism of vector spaces and can thus be used to identify distant fibres (as long as they lie in the same path component).

Remark 35.4.4. For every vector bundle, one can construct the frame bundle and use the parallel transport map on this bundle to define parallel transport of vectors. Therefore, the previous construction is applicable to any vector bundle.

Definition 35.4.5 (Covariant derivative). Consider a vector bundle $\pi : E \to M$ with typical fibre V and a connection 1-form ω on its associated principal $\mathsf{GL}(V)$-bundle. Let $\sigma : M \to E$ be a section of E and let X be a vector field on M. The covariant derivative of σ with respect to X is defined as follows:

$$\nabla_X \sigma|_{x_0} := \lim_{t \to 0} \frac{(\mathsf{Par}^\gamma_t)^{-1}\sigma\big(\gamma(t)\big) - \sigma(x_0)}{t}\,, \tag{35.66}$$

where γ is any curve satisfying $\gamma(0) = x_0$ and $\gamma'(0) = X(x_0)$. Let $\widetilde{\gamma}$ and X^H be the horizontal lifts of γ and X, respectively. An equivalent expression is the following one:

$$\nabla_X \sigma = \pi_{TE}(\sigma_* X - X^H \circ \sigma)\,. \tag{35.67}$$

One can also rephrase this in terms of the horizontal vector field associated to the lift $\widetilde{\gamma}$ (akin to Definition 34.3.17). By Property 35.1.21, every section σ of an associated bundle corresponds to a G-equivariant map $\phi(\sigma) : P \to V$. In terms of this map, one obtains

$$\phi(\nabla_X \sigma) = X^H\big(\phi(\sigma)\big)\,, \tag{35.68}$$

where X^H acts componentwise on V.

Property 35.4.6 (Koszul connection). The map

$$\Gamma(TM) \times \Gamma(E) \to \Gamma(E) : (X, \sigma) \mapsto \nabla_X \sigma \tag{35.69}$$

defines a Koszul connection (Definition 34.7.1). It follows, that every principal connection on a principal bundle induces a Koszul connection on all of its associated vector bundles.

35.4.2 Exterior covariant derivative

Definition 35.4.7 (Exterior covariant derivative). Let P be a principal bundle equipped with a connection 1-form ω and let $\theta \in \Omega^k(P)$ be a differential k-form. The exterior covariant derivative $\mathsf{D}\theta$ is defined as follows:

$$\mathsf{D}\theta(X_0, \dots, X_k) := \mathsf{d}\theta(X_0^H, \dots, X_k^H)\,, \tag{35.70}$$

where d is the exterior derivative (Definition 34.4.7) and X_i^H is the projection of X_i on the horizontal subspace $\mathsf{Hor}(T_pP)$. From this definition, it follows that the exterior covariant derivative $\mathsf{D}\theta$ is a horizontal form (Definition 35.3.12).

Remark 35.4.8. The exterior covariant derivative can also be defined for general vector-valued k-forms. This can be done by defining it component-wise with respect to a given basis. Afterwards one can prove that the choice of basis plays no role.

Property 35.4.9 (Equivariant forms). D preserves equivariance and, hence, if Φ is an equivariant form, then $\mathsf{D}\Phi$ is tensorial.

Formula 35.4.10 (Tensorial forms). An explicit expression for the action of D on tensorial forms of type (V, ρ) is given by:

$$\mathsf{D}\theta = \mathsf{d}\theta + \omega \barwedge \theta\,, \tag{35.71}$$

where $\barwedge$, sometimes also written as $\wedge_\rho$ is defined analogously to the wedge products (Construction 34.4.24 and Formula 34.4.29) using the induced representation of $\mathfrak{g}$ on V:

$$\omega \barwedge \theta(X, Y) := \omega(X) \triangleright \theta(Y) - \omega(Y) \triangleright \theta(X)\,. \tag{35.72}$$

Property 35.4.11. The action of the exterior covariant derivative (Definition 34.7.9) on $P\times_\rho V$-valued forms on M and that of the exterior covariant derivative (Definition 35.4.7) on V-valued forms on P is intertwined by the isomorphism (35.31):

$$D_\nabla \circ \tau = \tau \circ D_\omega \,. \tag{35.73}$$

This recovers expression (35.68).

The compatibility condition (35.42) for connection 1-forms can be restated in terms of the covariant derivative.

Property 35.4.12 (Gauge transformation). Consider a principal bundle P with a connection 1-form ω. For every gauge transformation $\xi \in \mathsf{Aut}_V(P)$, one (locally) has the following expression:

$$\xi^*\omega = \omega + \xi^{-1}\mathsf{D}\xi \,. \tag{35.74}$$

Because of Property 35.1.21, one can use the following construction to find an explicit expression for the covariant derivative on an associated vector bundle.

Formula 35.4.13 (Associated bundle). Let $\pi : P \to M$ be a principal G-bundle and let $P_V := P \times_G V$ be an associated vector bundle. Given a section $\sigma : M \to P_V$, one can construct a G-equivariant map $\phi : P \to V$ using Eq. (35.12). By Property 35.4.11 and Eq. (35.71), the exterior covariant derivative of ϕ and, accordingly, the covariant derivative of σ, is given by:

$$\nabla_{\pi_* X_i}\sigma = \mathsf{D}\phi(X) = \mathsf{d}\phi(X) + \omega(X) \triangleright \phi \,, \tag{35.75}$$

where $X \in \mathfrak{X}(P)$.

Formula 35.4.14 (Local expression). Given a local section $\varphi : U \to P$, one can pull back the above expression to the base manifold M. This gives

$$(\varphi^*\mathsf{D}\phi)(X) = \mathsf{d}(\varphi^*\phi)(X) + \varphi^*\omega \triangleright \varphi^*\phi(X) \,, \tag{35.76}$$

where $X \in \mathfrak{X}(M)$. After introducing the notations $S := \varphi^*\phi$ and $\nabla_X S := (\varphi^*\mathsf{D}\phi)(X)$ and remembering the definition of the Yang–Mills field 35.3.25, this becomes

$$\nabla_X S = \mathsf{d}S(X) + \omega^U(X) \triangleright S \,. \tag{35.77}$$

Example 35.4.15. Let $G = \mathsf{GL}(n,\mathbb{R})$. In local coordinates, Eq. (35.77) can be rewritten as follows:

$$(\nabla_X S)^i = \frac{\partial S^i}{\partial x^k} X^k + \Gamma^i{}_{jk} S^j X^k \,. \tag{35.78}$$

This is exactly the formula known from classical differential geometry and general relativity.

35.4.3 Curvature

Definition 35.4.16 (Curvature). Let ω be a principal connection 1-form. The curvature Ω of ω is defined as the exterior covariant derivative $\mathsf{D}\omega$.

Formula 35.4.17 (Cartan structure equation). Let ω be a connection 1-form and let Ω be its curvature form. The curvature can be expressed in terms of the connection as follows:

$$\Omega = \mathsf{d}\omega + \frac{1}{2}[\omega \wedge \omega]\,. \tag{35.79}$$

Recall the (geometric) fundamental theorem of calculus (Theorem 35.3.24). The Maurer–Cartan equation is equivalent to the vanishing of the (algebraic) curvature of a general $\mathfrak{g}$-valued 1-form.

Remark 35.4.18. Note the apparent similarity with Eq. (35.71). The extra factor 1/2 should, however indicate that the similarity is only superficial. The connection 1-form is not tensorial and, hence, Eq. (35.71) is not valid.

Property 35.4.9 implies the following important statement.

Property 35.4.19 (Tensorial). In contrast to a connection 1-form, the associated curvature is a tensorial $\mathfrak{g}$-valued two-form or, equivalently, an $\mathsf{End}(P)$-valued two-form on the base manifold M.

Definition 35.4.20 (Flat connection). A principal connection is said to be flat if its curvature vanishes everywhere. A bundle is said to be flat if it admits a flat connection.

Property 35.4.21 (Local systems ♣). If the connection on a vector bundle is flat, the flat sections constitute a (linear) local system (Definition 8.4.2). Moreover, this sheaf characterizes the bundle and connection up to isomorphism, i.e. there exists an equivalence of categories of flat vector bundles and linear local systems.

Example 35.4.22 (Maurer–Cartan connection). Let ω_G be the Maurer–Cartan form on a Lie group G. Because the only horizontal vector field on the bundle $G \hookrightarrow G \to \{*\}$ is the zero vector, the curvature of ω_G is 0. It follows, that the Maurer–Cartan form is a flat connection.

The following property is an immediate consequence of Frobenius' integrability theorem 34.6.6 and the fact that a connection vanishes on the horizontal subbundle.

Property 35.4.23 (Integrability). For any principal connection, the associated horizontal distribution

$$p \mapsto H_p P \tag{35.80}$$

is integrable if and only if the connection is flat. By contrast, the vertical distribution is always integrable.

Property 35.4.24 (Second Bianchi identity). Let ω be a principal connection 1-form with curvature Ω. The curvature is covariantly constant:

$$\mathsf{D}\Omega = 0\,. \tag{35.81}$$

One should pay attention to the fact that this result does not generalize to arbitrary differential forms. The following property gives a formula for the double covariant derivative of tensorial forms.

Formula 35.4.25 (Curvature on associated bundles). The above definition of the curvature, together with Eq. (35.71) or, equivalently, Formula 35.4.13, implies that one can express the action of the curvature on sections of associated bundles as follows:

$$\mathsf{D}^2\phi = \Omega \triangleright \phi\,. \tag{35.82}$$

This curvature form Ω coincides with the one from Definition 34.7.17.

Similar to Definition 35.3.25, one can also define the Yang–Mills field strength.

Definition 35.4.26 (Field strength). Let P be a principal bundle equipped with a principal connection 1-form ω and associated curvature Ω. Given a local section $\sigma : U \to P$, one defines the (Yang–Mills) field strength F as the pullback $\sigma^*\Omega$.

Theorem 35.4.27 (Ambrose–Singer). *The Lie algebra of the holonomy group* $\mathsf{Hol}_p(\omega)$ *is spanned by the elements of the form* $\Omega_q(X, Y)$*, where* q *ranges over the holonomy bundle* $H(p)$ *and* X, Y *are horizontal.*

35.4.4 Torsion

Definition 35.4.28 (Solder form). Let $\pi : P \to M$ be a principal G-bundle and let V be a $\dim(M)$-dimensional vector space equipped with a representation[4] $\rho : G \to \mathsf{GL}(V)$ such that $TM \cong P \times_G V$ as associated bundles. A solder(ing) form on P is a tensorial 1-form (Definition 35.3.13) of type (V, ρ).

Definition 35.4.29 (Torsion). Let $\pi : P \to M$ be a principal G-bundle equipped with a connection 1-form ω and a solder form θ. The torsion Θ is defined as the exterior covariant derivative $\mathsf{D}\theta$. This is the content of the **Cartan structure equation**:

$$\Theta = \mathsf{d}\theta + \omega \barwedge \theta\,. \tag{35.83}$$

Property 35.4.30 (First Bianchi identity). Let ω be a connection 1-form, Ω its associated curvature, θ a solder form and Θ its associated torsion.

$$\mathsf{D}\Theta = \Omega \barwedge \theta \tag{35.84}$$

[4] In general, this will be $V = \mathbb{R}^{\dim(M)}$ and $G = \mathsf{GL}(n, \mathbb{R})$.

35.5 Reduction of the structure group

Definition 35.5.1 (Reduction). Consider a principal G-bundle $\pi : P \to M$ and let H be a subgroup of G. If the transition functions of P can be chosen to take values in H, it is said that the structure group G can be reduced to H.

More generally, a principal bundle $H \hookrightarrow \widetilde{P} \to M$ with structure group H is called an H-reduction of P if there exists a bundle isomorphism $\widetilde{P} \times_H G \cong P$. This allows for morphisms besides inclusions, such as covering maps $\lambda : H \to G$, making the term 'reduction' rather misleading. (See, for example, the definition of *spinor bundles* in Section 36.3.) For covering maps, the term **lift(ing)** is sometimes used.

Definition 35.5.2 (G-structure). Consider an n-dimensional manifold M and a topological group G. A G-structure on M is a reduction of the structure group $\mathsf{GL}(n)$ of the frame bundle FM along a group morphism $\iota : G \to \mathsf{GL}(n)$.

Definition 35.5.3 (Integrability). A G-structure P on M is said to be integrable if, for every point $x \in M$, there exists a chart $U \ni x$ such that the associated holonomic frame $\{\partial_i\}_{i \leq \dim(M)}$ induces a local section of P.

Property 35.5.4. An integrable G-structure admits a torsion-free connection.

Example 35.5.5 (Orientable manifold). An n-dimensional manifold is orientable if and only if the structure group can be reduced to $\mathsf{GL}^+(n)$, the group of invertible matrices with positive determinant (or even $\mathsf{SL}(n)$, since $\mathsf{GL}^+(n)$ deformation retracts onto $\mathsf{SL}(n)$). Furthermore, if it exists, it is integrable.

Example 35.5.6 (Riemannian manifold). An $\mathsf{O}(n)$-structure turns M into a *Riemannian manifold* (see Definition 36.1.2). Because the cotangent bundle T^*M transforms under the contragredient representation of the transition maps of the tangent bundle TM, which coincides with the regular representation in the case of $\mathsf{O}(n)$, these two bundles are equivalent. The isomorphism is given by the *musical isomorphism(s)* (see Definition 36.1.3). Moreover, Riemannian structures are always integrable.

The following property gives a classification of bundle reductions.

Property 35.5.7 (Equivariant morphisms). Consider a principal G-bundle P and let F be a set that admits a transitive action $\varphi : G \to \mathsf{Aut}(F)$. For every $f \in F$ and every equivariant morphism $\psi : P \to F$, there exists a reduction of G to the isotropy subgroup G_f defined by

$$P_f := \{p \in P \mid \psi(p) = f\}\,. \tag{35.85}$$

One can generalize this definition to arbitrary Lie group actions by restricting to the equivariant morphisms that take value in a single orbit.[5]

[5] Since transitive actions have a unique orbit, this is a well-defined generalization.

Consider a subgroup inclusion $\iota : H \hookrightarrow G$. If H is closed, the action of G on G/H is transitive and one can specialize the above construction to the coset space G/H. It follows, that reductions are classified by equivariant maps into the coset space G/H or, according to Property 35.1.21, by the (global) sections of the associated coset bundle $P \times_G G/H$.

Corollary 35.5.8. If G is connected, every principal G-bundle is reducible to a maximal compact subgroup of G.

Definition 35.5.9 (Reducible connection). Consider a principal G-bundle P equipped with a connection 1-form ω. If a bundle map F induces an H-reduction of P, the connection ω is said to be reducible (and to be compactible with the given reduction) if $F^*\omega$ takes values in $\mathfrak{h}$.

Property 35.5.10. Consider a principal bundle P together with a reduction P_f induced by an equivariant morphism $\psi : P \to F$ with $f \in F$. A principal connection on P is reducible to P_f if and only if ψ is parallel with respect to this connection, i.e. $\mathrm{D}\psi = 0$.

The following two properties characterize bundle reductions in terms of holonomy bundles.

Property 35.5.11 (Holonomy bundles and reductions). The holonomy bundle $H(p)$ is a reduction of P for every $p \in P$. Furthermore, any connection ω is reducible to $H(p)$ and it can be proven that this reduction is minimal, i.e. there exists no further reduction.

Corollary 35.5.12. A principal bundle (and any associated connection) is irreducible to a subgroup of the structure group[6] if and only if it is equivalent to its holonomy bundle.

The following property is less well known in the literature.

Property 35.5.13 (Flat connections ♣). A principal bundle is flat if and only if its structure group G can be lifted to the discrete group G^δ, i.e. the same group but with the discrete topology. An equivalent condition is that the structure group can be lifted to the fundamental group of the base space $\pi_1(M)$. This latter condition is related to the fact that, for flat connections, parallel transport is path independent and, hence, is fully characterized by the loops in M. Note that once such a lift is chosen or, equivalently, if the structure group of the bundle is discrete, a unique flat connection exists.

More abstractly, one obtains the following bijection:

$$\mathsf{hol} : H^1_{\text{flat conn}}(M; G) \cong \mathbf{Grp}\big(\pi_1(M), G\big)/G\,, \tag{35.86}$$

where $H^1_{\text{flat conn}}(M; G) := \pi_0 \mathbf{H}(M, \flat \mathbf{B}G)$ is the cohomology set (Definition 14.5.1) of *differentiable stacks* and $\flat$ is the flat modality (Property 44.3.8). By the adjunctions of cohesive

[6] Lifts as in the case of Spin-structures do not fall under the holonomy classification.

modalities, this classification is equivalent to giving a morphism

$$\mathsf{transport} : \mathbf{\Pi}(M) \to \mathbf{B}G \tag{35.87}$$

from the fundamental (∞-)groupoid (Definition 7.1.17) to the delooping groupoid.

Remark 35.5.14. The above condition can also be applied to define flatness for topological bundles, where the notion of connections does not make sense.

35.6 Characteristic classes

Definition 35.6.1 (Characteristic class). Let M be a manifold. A characteristic class is a map from isomorphism classes of vector bundles or principal bundles $E \to M$ to cohomology classes $c(E) \in H^\bullet(M; R)$ that is stable under pullback. The coefficient ring R is often assumed to be the base field ($\mathbb{R}$ or $\mathbb{C}$), but this is not always the case (e.g. the *Stiefel–Whitney classes* from Section 36.3.2).

Using the classification property 35.2.5, one can give a concise construction of characteristic classes in the case of principal bundles.

Construction 35.6.2. Consider a principal G-bundle over M with classifying map $\varphi \in [M, BG]$. For every $c \in H^\bullet(BG)$, one defines the characteristic class

$$c(P) := \varphi^* c \in H^\bullet(M)\,. \tag{35.88}$$

As the definition implies, both vector bundles and principal bundles admit a theory of characteristic classes. However, in the literature, most authors focus on either one of them and, hence, it is not always easy to see which theorems can be translated and how to do this whenever possible. The relation between the two theories is given by the associated bundle construction 35.1.10 (see Sorensen (2017) for more information). The characteristic classes of a vector bundle are defined as those of its frame bundle. Because of this duality, one can freely switch between the language of vector bundles and principal bundles, depending on where the results will be applied.

Because the statement of the 'splitting principle' is quite different when given in the language of principal bundles or that of vector bundles, it will be stated for both cases. First, an additional construction is needed.

Definition 35.6.3 (Flag bundle). Let $\pi : E \to M$ be a vector bundle. Using the definition of the flag manifold (Definition 22.8.5), one can construct, for every fibre E_p, a space $\mathsf{Fl}(E_p)$ that has the complete flags of E_p as points, expressed as a sequence of one-dimensional subspaces. From this, one can then construct the flag bundle $\pi_{\mathrm{Fl}} : \mathsf{Fl}(E) \to M$ with the flag manifolds as fibres.

Theorem 35.6.4 (Splitting principle). *Consider a vector bundle $\pi : E \to M$. Its flag bundle has the following properties:*

- *The pullback bundle $\pi_{Fl}^* E$ can be decomposed as a Whitney sum of line bundles.*
- *The induced morphism on cohomology $\pi_{Fl}^* : H^\bullet(M) \to H^\bullet(\mathsf{Fl}(E))$ is injective.*

For the following form of the splitting principle, see Debray (2018); May (2005).

Theorem 35.6.5 (Splitting principle). *Consider a principal bundle $\pi : P \to M$ for which the structure group G is compact. Every compact Lie group contains a maximal torus $T \cong \mathbb{T}^n$, where $\mathbb{T}$ is the standard 1-torus $S^1 \cong \mathsf{U}(1)$. The inclusion $\iota : T \hookrightarrow G$ induces a G-bundle $B\iota : BT \to BG$ with fibre G/T and total space EG. The pullback of $B\iota$ along the classifying map $p \in [M, BG]$ of P defines another G-bundle $\rho : p^* B\iota \to M$, again with fibre G/T. This fibre bundle has the following properties:*

- *$\rho^* p$ admits a reduction of the structure group to T.*
- *The induced morphism on cohomology $\rho^* : H^\bullet(M) \to H^\bullet(\rho^* P)$ is injective.*

Because $B\mathbb{T}^n \cong (B\mathbb{T})^n$, one can use the fibration $B\iota$ to pull back any class $c \in H^\bullet(BG)$ to a tuple of classes in $H^\bullet(B\mathsf{U}(1))$. Therefore, every characteristic class of $\rho^* P$ is a tuple of characteristic classes of circle bundles. The injectivity of ρ^* implies that every characteristic class of P can be characterized by such a tuple.

35.6.1 Chern–Weil theory

The characteristic classes of a vector bundle can be constructed from a connection and curvature form on the vector bundle. The resulting expressions are polynomial in the curvature forms. This result is established by Chern–Weil theory.

Definition 35.6.6 (Chern–Weil morphism). Let $\pi : P \to M$ be a principle G-bundle and choose a connection 1-form ω with curvature two-form Ω. There exists a morphism of algebras

$$\mathfrak{K}[\mathfrak{g}]^G \to \Omega^\bullet(P) : f \mapsto f(\Omega)\,, \tag{35.89}$$

where $\mathfrak{K}$ is the base field, satisfying:

- $f(\Omega)$ is closed.
- $f(\Omega)$ pulls back uniquely to a (closed) form $\overline{f}(\Omega) := \pi^* f(\Omega)$ on M.
- $\overline{f}(\Omega)$ does not depend on the chosen connection, i.e. the difference $\overline{f}(\Omega) - \overline{f}(\Omega')$ for any two connection 1-forms ω, ω' is exact.

In the remainder of this section, this approach will be used to find explicit descriptions of characteristic classes of vector bundles and principal bundles.

35.6.2 Complex bundles

On *complex bundles*, one can always choose $\mathfrak{u}(n)$-valued connection 1-forms, since the structure group can be reduced to $\mathsf{U}(n)$. See Chapter 39 for more information.

Definition 35.6.7 (Chern class). Consider a rank-n vector bundle $\pi : E \to M$ with curvature two-form Ω. Using Chern–Weil theory, one defines the Chern classes $c_k(E) \in H^{2k}(M)$ as follows:

$$\det\left(\mathbb{1} + \frac{it}{2\pi}\Omega\right) =: \sum_{k=1}^{n} c_k(E)t^k \,. \tag{35.90}$$

This series is sometimes called the **Chern polynomial** $c_t(E)$. For $t = 1$, one obtains the **total Chern class**.

Definition 35.6.8 (Chern character). Consider a rank-n vector bundle $\pi : E \to M$ with curvature two-form Ω. Using Chern–Weil theory, one defines the Chern character as follows:

$$\mathrm{ch}(E) := \mathrm{tr}\left(\exp\left(\frac{i\Omega}{2\pi}\right)\right). \tag{35.91}$$

If $c_i := c_i(E)$ denotes the ith Chern class of E, the Chern character can also be expressed as

$$\mathrm{ch}(E) = \sum_{k=0}^{n} \frac{c_1^k + \cdots + c_n^k}{k!} \,. \tag{35.92}$$

The term with prefactor $1/k!$ is a homogeneous polynomial of degree $k \in \mathbb{N}$. One sometimes calls this term the kth Chern character. By Chern–Weil theory, this form is proportional to $\mathrm{tr}(\Omega^k)$.

Formula 35.6.9 (Whitney product formula[7]). The following equality holds for all bundles E_1, E_2:

$$c_t(E_1 \oplus E_2) = c_t(E_1)c_t(E_2) \,. \tag{35.93}$$

Corollary 35.6.10 (Chern root). The product formula and the splitting principle imply that the Chern polynomial of any rank-n vector bundle can be decomposed as follows:

$$c_t(E) = \prod_{i=1}^{n}(1 + x_i t) \,, \tag{35.94}$$

[7] This formula is also called the **Whitney sum formula**.

where, in the case of decomposable vector bundles $E = \oplus_{i=1}^{n} L_i$, the x_i are the first Chern classes $c_1(L_i)$. The factors x_i are called the **Chern roots**.

By working out the above formula, one can see that the coefficient in degree $k \in \mathbb{N}$, i.e. the kth Chern class, is given by the kth elementary symmetric polynomial:

$$c_k(E) = \sum_{i_1 < \cdots < i_k} x_{i_1} \cdots x_{i_k} \,. \tag{35.95}$$

Definition 35.6.11 (Canonical class). The first Chern class of the canonical bundle of a smooth manifold M. It is often denoted by K_M.

Definition 35.6.12 (Theta characteristic). Consider a smooth manifold M with its canonical class K_M. The theta characteristic, if it exists, is a characteristic class Θ such that $\Theta \cup \Theta = K_M$, where $\cup$ is the cup-product (Formula 34.9.6).

After finding the Chern roots of a vector bundle, they can be used to define various other classes.

Construction 35.6.13 (Genus). Let $f \in K[[t]]$ be a formal power series with constant term 1. For any $k \in \mathbb{N}$, one can easily see that $f(x_1) \cdots f(x_k)$ is a symmetric power series (also with constant term 1). For every such f, define the f-genus by the formula[8]

$$G_f(E) := \det f\left(\frac{it}{2\pi}\Omega\right). \tag{35.96}$$

The coefficients of this power series define characteristic classes of E.

Example 35.6.14 (Chern class). The total Chern class is recovered as the genus of $f = 1 + x$.

The following genus is very important, especially in the context of the *Atiyah–Singer index theorem* (see further below).

Example 35.6.15 (Todd class). Consider the function

$$Q(x) := \frac{x}{1 - e^{-x}} = 1 + \frac{x}{2} + \sum_{i=1}^{+\infty} \frac{(-1)^{i-1} B_i}{(2i)!} x^{2i} \,, \tag{35.97}$$

where B_i is the ith *Bernoulli number*. Let $\pi : E \to M$ be a rank-n vector bundle. If x_i are the Chern roots of E, the Todd class is defined as

$$\mathrm{td}(E) := \prod_{i=1}^{n} Q(x_i) \,. \tag{35.98}$$

The characteristic function of the Todd genus is the unique power series with constant term 1 that has the property that, for all $n \in \mathbb{N}$, the nth degree term in $f(x)^{n+1}$ has coefficient 1.

[8] In the case that E splits as a sum for line bundles, one simply obtains the product $f(x_1) \cdots f(x_k)$.

Another genus that is used in the context of the index theorems is the following one.

Example 35.6.16 ($\hat{A}$-genus[9]). The $\hat{A}$-genus is defined through the following function:

$$Q(x) := \frac{\sqrt{x}/2}{\sinh(\sqrt{x}/2)} = 1 - \frac{x}{24} + \frac{7x^2}{5760} - \cdots . \tag{35.99}$$

35.6.3 Real bundles

In the case of real vector bundles, which will be assumed to come equipped with a *fibre metric* as to allow for $\mathfrak{o}(n)$-valued connection 1-forms, one can also define a set of characteristic classes.

Definition 35.6.17 (Pontryagin class). Consider a real vector bundle $\pi : E \to M$. The Pontryagin classes of E are defined as follows:

$$p_k(E) := (-1)^k c_{2k}(E^{\mathbb{C}}) \in H^{4k}(M) , \tag{35.100}$$

where $E^{\mathbb{C}}$ is the *complexification* of E. If E has the structure of a complex vector bundle, one can use the relation $E^{\mathbb{C}} \cong E \oplus \overline{E}$ to express the Pontryagin classes purely in terms of the Chern classes of E, e.g.

$$p_1(E) = c_1^2(E) - 2c_2(E) = \mathsf{ch}_2(E) . \tag{35.101}$$

When the vector bundles in question are orientable, the structure group can further be reduced to $\mathsf{SO}(n)$. If the rank is even, one can define the following characteristic class.

Definition 35.6.18 (Euler class). Let $\pi : E \to M$ be an orientable vector bundle of rank $2k$. The Euler class of E is defined as follows:

$$e(E) := p_k(E) \cup p_k(E) . \tag{35.102}$$

Property 35.6.19. Using the fact that one can write the total Pontryagin class using Chern–Weil theory as

$$p(E) = \det\left(1 - \frac{1}{2\pi}\Omega\right) \tag{35.103}$$

and that the determinant is the square of the *Pfaffian*, one can equivalently define the Euler class as follows:

$$e(E) := \mathsf{Pf}\left(-\frac{1}{2\pi}\Omega\right) . \tag{35.104}$$

[9]This is pronounced as *A-roof genus*.

35.6.4 Lie group cohomology

Using the language of characteristic classes, one can find a concise description of the (continuous) group cohomology of Lie groups. First of all, there is the isomorphism between continuous group cohomology and cohomology of classifying spaces:

$$H^\bullet(BG;\mathbb{Z}) \cong H_c^\bullet(G;\mathbb{Z})\,. \tag{35.105}$$

@@ COMPLETE @@

35.6.5 Chern–Simons forms

The image of invariant polynomials under the Chern–Weil morphism is closed. This not only allows to interpret them as cohomology classes as done above, but it also implies that one can find a (local) trivialization:

$$\langle \Omega_A, \cdots \rangle_n = \mathsf{dCS}^n(A)\,, \tag{35.106}$$

where $\langle \cdots \rangle_n$ denotes an invariant polynomial of degree $2n \in \mathbb{N}$. Such a form is called a **Chern–Simons form** or **secondary characteristic form**.

More generally, consider the *concordance* $P \times [0,1]$ for some principal bundle P with itself together with a connection $\widehat{A}$. This connection defines a path between two connections A, A' on P. The relative Chern–Simons form is defined through transgression:

$$\mathsf{CS}^n(A,A') := \int_{[0,1]} \langle \Omega_{\widehat{A}}, \cdots \rangle_n\,. \tag{35.107}$$

The differential of this form gives the difference of characteristic forms:

$$\mathsf{dCS}^n(A,A') = \langle \Omega_{A'}, \cdots \rangle_n - \langle \Omega_A, \cdots \rangle_n\,. \tag{35.108}$$

As usual for local trivializations, the Chern–Simons form is only defined up to an exact form.

Example 35.6.20 (Killing form). Consider the Killing transgression form, which, for $\mathfrak{su}(n)$, is induced by the trace functional. The related Chern–Simons form is given by

$$\langle \mathsf{d}A, A \rangle_2 + \frac{2}{3}\langle A, [A \wedge A] \rangle_2\,. \tag{35.109}$$

This form is the exterior derivative of the second Chern character, which, for $\mathsf{SU}(n)$-bundles, is equivalent to the second Chern class. A similar expression can be obtained for the Chern–Simons form associated to all other Chern characters.

35.7 Differential cohomology ♣

In the foregoing sections, a multitude of objects were introduced that are related to principal fibre bundles. For example, connections and their associated curvature forms could be used to construct differential quantities, while characteristic classes contained data about the topology of the bundle. However, even in the simple case of U(1)-bundles, neither the (first) Chern class, nor the curvature form are able to uniquely characterize the bundle.

35.7.1 Differential characters

In this section, all (co)chains, (co)cycles and (co)boundaries are assumed to be smooth. By doing so, no generality is lost since every continuous chain is homotopic to a smooth one.

Definition 35.7.1 (Differential character). Let M be a smooth manifold. A (**Cheeger–Simons**) differential character of **degree** $k \in \mathbb{N}_0$ is a group morphism $\chi : Z_{k-1}(M) \to \mathsf{U}(1)$ given by integration over boundaries:[10]

$$\chi(\partial\gamma) := \exp\left(2\pi i \int_\gamma \omega(\chi)\right) \tag{35.110}$$

for some $\omega(\chi) \in \Omega^k(M)$. The group of differential characters of degree k is denoted by $\widehat{H}^k(M;\mathbb{Z})$. For $k = 0$, the convention $\widehat{H}^0(M;\mathbb{Z}) := H^0(M;\mathbb{Z})$ is used.

Property 35.7.2 (Thin invariance). Differential characters vanish on boundaries of thin chains, i.e. for chains $\gamma \in C_k(M)$ such that $\int_\gamma \omega = 0$ for all $\omega \in \Omega^k(M)$, one has $\chi(\partial\gamma) = 1$.

Property 35.7.3 (Curvature). Every differential character is represented by a unique, closed and integral k-form. The map

$$\mathsf{curv} : \widehat{H}^k(M;\mathbb{Z}) \to \Omega^k_{\text{int}}(M) : \chi \mapsto \omega(\chi) \tag{35.111}$$

is called the curvature map. If $\mathsf{curv}(\chi) = 0$, the character χ is said to be **flat**.

Property 35.7.4 (Characteristic class). Every differential character gives rise to a characteristic class as follows. The group of cocycles is free and the quotient map $\mathbb{R} \to \mathsf{U}(1)$ is onto, so every differential character lifts to a group homomorphism $\widetilde{\chi} : Z_{k-1}(M) \to \mathbb{R}$ such that $\chi(z) = \exp\big(2\pi i\widetilde{\chi}(z)\big)$. The map

$$\mathsf{ch}(\chi) : C_k(M) \to \mathbb{Z} : \gamma \mapsto \int_\gamma \mathsf{curv}(\chi) - \widetilde{\chi}(\partial\gamma) \tag{35.112}$$

induces a well-defined map $\mathsf{ch} : \widehat{H}^k(M;\mathbb{Z}) \to H^k(M;\mathbb{Z})$. If $\mathsf{ch}(\chi) = 0$, the character χ is said to be **topologically trivial**. The characteristic class associated to a differential character is sometimes called the **Dixmier–Douady** class (see e.g. Brylinski (1993)).

[10] Some authors omit the exponential function by working modulo $\mathbb{Z}$. This just replaces the multiplicative group U(1) by the isomorphic additive group $\mathbb{R}/\mathbb{Z}$.

Example 35.7.5 (Circle bundles). Consider a U(1)-bundle $\pi : P \to M$ with connection 1-form ω. Holonomy around a closed curve γ gives a parallel transport map

$$P \to P : p \mapsto p \cdot g(p, \gamma) \tag{35.113}$$

for a smooth function $g : \Omega_p P \to \mathsf{U}(1)$. In fact, g only depends on the homology of γ and the projection $\pi(p)$, so one obtains a map $g \in \widehat{H}^2(M; \mathbb{Z})$ with curvature $\frac{-1}{2\pi i}\Omega$ and characteristic class $c_1(P)$. The converse also holds, every differential character of degree 2 determines a principal U(1)-bundle with connection (up to connection-preserving isomorphisms). This leads to the following equivalence:

$$\begin{aligned}\widehat{H}^2(M; \mathbb{Z}) \cong \{&\text{isomorphism classes of } (P, \nabla) \mid P \text{ a circle bundle and} \\ &\nabla \text{ a principal connection}\}\,.\end{aligned} \tag{35.114}$$

The curvature and characteristic class maps fit in some exact sequences.

Property 35.7.6 (Curvature exact sequence). The first sequence is induced by the curvature map. A vanishing curvature form says that the character vanishes identically on boundaries. This is exactly the property satisfied by cohomology classes:

$$0 \longrightarrow H^{k-1}(M; \mathsf{U}(1)) \longrightarrow \widehat{H}^k(M; \mathbb{Z}) \xrightarrow{\text{curv}} \Omega^k_{\text{int}}(M) \longrightarrow 0\,. \tag{35.115}$$

The first cohomology group classifies flat circle bundles by Property 35.5.13, so this sequence says that, by extending the above example to higher n-bundles (this can be formalized cf. *bundle gerbes*), two circle n-bundles with the same curvature differ by a flat circle $(n-1)$-bundle.

Property 35.7.7 (Characteristic class exact sequence).

$$0 \longrightarrow \Omega^{k-1}(M)/\Omega^{k-1}_{\text{int}}(M) \longrightarrow \widehat{H}^k(M; \mathbb{Z}) \xrightarrow{\text{ch}} H^k(M; \mathbb{Z}) \longrightarrow 0 \tag{35.116}$$

The first map is induced by the holonomy functional

$$\iota : \Omega^{k-1}(M) \to \widehat{H}^k(M; \mathbb{Z}) : \omega \to \exp\left(2\pi i \int_{_} \omega\right), \tag{35.117}$$

which has the closed integral forms as kernel. This exact sequence says that two connections on the same principal U(1)-bundle differ by a global connection form (up to an integral form).

35.7.2 Combining singular and de Rham cohomology

There is an alternative to the Cheeger–Simons approach. Let C^n and Z^n again denote the smooth cochain and cocycle groups.

Definition 35.7.8 (Differential cocycle). A tuple $(c, h, \omega) \in C^n(M;\mathbb{Z}) \times C^{n-1}(M) \times \Omega^n(M)$ such that

$$\begin{aligned} \delta c &= 0 \\ \mathrm{d}\omega &= 0 \\ \delta h &= \omega - c\,. \end{aligned} \tag{35.118}$$

A differential cocycle thus consists of a singular cocycle (topological information) and a de Rham cocycle (differential information), that are equal up to a (singular) coboundary.

The cochain complex $C^n(M;\mathbb{Z}) \times C^{n-1}(M) \times \Omega^n(M)$ with differential

$$\mathrm{d} : (c, h, \omega) \mapsto (\delta c, \omega - c - \delta h, \mathrm{d}\omega) \tag{35.119}$$

defines a cohomology theory $\widehat{H}(n)^\bullet(M)$.

Property 35.7.9 (Relation to differential characters). Differential characters and differential cocycles are related as follows:

$$\widehat{H}^k(M;\mathbb{Z}) \cong \widehat{H}(k)^k(M)\,. \tag{35.120}$$

Given a differential cocycle (c, h, ω), the curvature and characteristic class of the associated differential character are ω and c, respectively. The function $e^{2\pi i h}$ is called the **monodromy** of the cocycle. It can be checked that Property 35.7.4 is exactly the third relation in the definition of cocycles above. The $\mathrm{mod}\ \mathbb{Z}$-reduction of h gives the differential character associated to the cocycle.

Example 35.7.10. The first ordinary differential cohomology group $\widehat{H}^1(M;\mathbb{Z})$ is isomorphic to the group of smooth functions $C^\infty(M;\mathrm{U}(1))$.

35.7.3 Deligne cohomology

The following theorem states that the differential characters are essentially the unique objects with these properties and that they define a generalized cohomology theory.

Theorem 35.7.11 (Simons–Sullivan). *There is an essentially unique functor*

$$\widehat{H}^\bullet(-;\mathbb{Z}) : \mathbf{Diff} \to \mathbf{Ab}^{\mathbb{Z}}$$

such that there exist four natural transformations

1. **Flat class**: $j : H^{\bullet-1}(-;\mathrm{U}(1)) \to \widehat{H}^\bullet(-;\mathbb{Z})$,
2. **Topological trivialization**: $\iota : \Omega^{\bullet-1}/\Omega^{\bullet-1}_{\mathrm{int}} \to \widehat{H}^\bullet(-;\mathbb{Z})$,
3. **Characteristic class**: $\mathrm{ch} : \widehat{H}^\bullet(-;\mathbb{Z}) \to H^\bullet(-;\mathbb{Z})$, *and*

4. **Curvature**: $\mathsf{curv} : \widehat{H}^\bullet(-;\mathbb{Z}) \to \Omega^\bullet_{\text{int}}$

that fit in the following commutative diagram, where the diagonal sequences are exact:

$$\begin{tikzcd}
0 \arrow[dr] & & & & & & 0 \\
& H^{\bullet-1}(-;\mathsf{U}(1)) \arrow[rr, "\text{Bockstein}"] \arrow[dr, "j"] & & & H^\bullet(-;\mathbb{Z}) \arrow[ur] \arrow[dr, "\text{de Rham}"] & & \\
H^{\bullet-1}_{\text{dR}} \arrow[ur] \arrow[dr, two heads] & & \widehat{H}^\bullet(-;\mathbb{Z}) \arrow[ur, "\text{ch}"] \arrow[dr, "\text{curv}"] & & & H^\bullet_{\text{dR}} & \\
& \Omega^{\bullet-1}/\,\Omega^{\bullet-1}_{\text{int}} \arrow[ur, "\iota"] \arrow[rr, "\mathsf{d}"'] & & & \Omega^\bullet_{\text{int}} \arrow[ur, hook] \arrow[dr] & & \\
0 \arrow[ur] & & & & & & 0
\end{tikzcd}$$

Functors satisfying the above properties are said to define **ordinary differential cohomology theories**. Another approach to differential cohomology is given by the Deligne complex.

Definition 35.7.12 (Deligne complex). Let $\mathbf{B}^k\mathsf{U}(1)_{\text{conn}}$ denote the cochain complex

$$\mathcal{O}^\times_M \xrightarrow{\mathsf{d}\log} \Omega^1 \xrightarrow{\mathsf{d}} \cdots \xrightarrow{\mathsf{d}} \Omega^k. \tag{35.121}$$

(Smooth) Deligne cohomology is defined as follows:

$$H^{k+1}_D(M;\mathbb{Z}) := \check{H}^0\big(M;\mathbf{B}^k\mathsf{U}(1)_{\text{conn}}\big)\,, \tag{35.122}$$

where $\check{H}^\bullet$ denotes Čech cohomology (Definition 8.4.10). (Note that $\mathbf{B}^k\mathsf{U}(1)_{\text{conn}}$ is turned into a chain complex by inverting the degrees.)

Property 35.7.13 (Deligne–Beilinson product). Consider the Deligne complex for two integers $k, l \in \mathbb{N}$. There exists a cup product

$$\begin{aligned}
\cup : \mathbf{B}^k\mathsf{U}(1)_{\text{conn}} \otimes \mathbf{B}^l\mathsf{U}(1)_{\text{conn}} &\to \mathbf{B}^{k+l+1}\mathsf{U}(1)_{\text{conn}} \\
: x \otimes y \mapsto x \cup y &:= \begin{cases} x \wedge \mathsf{d}y & \text{if } \deg(y) = l\,, \\ 0 & \text{otherwise}\,. \end{cases}
\end{aligned} \tag{35.123}$$

Example 35.7.14 (Circle bundles). A (Čech–)Deligne cocycle in degree 2 consists of data (A_i, g_{ij}) such that

$$\begin{tikzcd}
A_i \arrow[r, hook] & A_i - A_j = \mathsf{d}\log g_{ij} = g_{ij}^{-1}\mathsf{d}g_{ij} \\
g_{ij} \arrow[u, "\mathsf{d}\log"] \arrow[r, hook] & g_{jk}g_{ki}^{-1}g_{ij} = 1\,,
\end{tikzcd} \tag{35.124}$$

where the inclusion arrows denote the restriction to intersections $U_{ij} := U_i \cap U_j$. Formula 35.3.28 and the subsequent example, specialized in the case of $\mathsf{U}(1)$-bundles, show that the above data are exactly the components of a principal circle bundle with connection.

Remark 35.7.15. As was the case for differential characters, higher Deligne cohomology classes classify higher $\mathsf{U}(1)$-bundles with connection. The main benefit of this approach is that one gets an 'explicit' description of the local data. See Hitchin (1999) for a good introduction.

Remark 35.7.16 (Trivial bundles and twisted bundles). From Čech–Deligne cohomology, one knows that a trivial k-bundle α is defined by a $(k-1)$-cochain β such that

$$(\delta\beta)_{i_0\dots i_k} = \alpha_{i_0\dots i_k}\,, \tag{35.125}$$

i.e. a trivial k-bundle is equivalent to a twisted $(k-1)$-bundle.

@@ COMPLETE (see book by Deligne) @@

35.8 Cartan geometry

In the first part of this section, a short overview of *Klein*'s **Erlangen program**, which unifies (and generalizes) Euclidean and non-Euclidean geometries, will be given. In the second part of this section, *Cartan*'s generalization in terms of bundle theory is explained. A reference for this section is Sharpe (2000).

35.8.1 Klein geometry

Definition 35.8.1 (Klein geometry). Consider a Lie group G together with a closed subgroup H. If it is connected, the orbit space G/H is called a Klein geometry with **principal group** G. If the principal group is also connected, the Klein geometry is said to be **geometrically oriented**.

If the associated Lie algebras are denoted by $\mathfrak{g}, \mathfrak{h}$, respectively, the pair $(\mathfrak{g}, \mathfrak{h})$ is called a **Klein pair**. In fact, any pair $(\mathfrak{g}, \mathfrak{h} \leq \mathfrak{g})$ can be called a Klein pair.

Property 35.8.2 (Bundle). It is clear that every Klein geometry gives a homogeneous space and, hence, a principal bundle of rank $\dim(G) - \dim(H)$.

Example 35.8.3 (Euclidean space). Consider the Euclidean group $\mathsf{Euc}(n) := \mathbb{R}^n \rtimes \mathsf{O}(n)$, i.e. the symmetry group of the Euclidean space $\mathbb{R}^n$. This group clearly acts transitively and the subgroup $\mathsf{O}(n)$ can be seen to leave the origin fixed. This implies that $\mathbb{R}^n$ is a homogenous space and even a Klein geometry of the form $\mathsf{Euc}(n)/\mathsf{O}(n)$.

Definition 35.8.4 (Effective Klein pair). The action of G on G/H is not necessarily effective, i.e. the kernel

$$\ker(\rho) = \{x \in G \mid \forall g \in G : g^{-1}xg \in H\} \tag{35.126}$$

is not necessarily trivial. If it is, the Klein geometry is said to be effective. In terms of the associated Klein pair, this means that $\mathfrak{h}$ contains no nontrivial ideals of $\mathfrak{g}$. A Klein geometry is said to be locally effective if the kernel is discrete.

Definition 35.8.5 (Reductive Klein pair). A Klein pair $(\mathfrak{g}, \mathfrak{h})$ for which $\mathfrak{g}$ admits a decomposition of the form

$$\mathfrak{g} = \mathfrak{h} + \mathfrak{m}\,, \tag{35.127}$$

where $\mathfrak{m}$ is an $\mathfrak{h}$-module.

Definition 35.8.6 (Model geometry). A model geometry consists of the following data:

1. an effective Klein pair $(\mathfrak{g}, \mathfrak{h})$,
2. a Lie group H such that $\mathsf{Lie}(H) = \mathfrak{h}$, and
3. a representation $\mathsf{Ad} : H \to \mathsf{Aut}(\mathfrak{g})$ that restricts to the adjoint representation $\mathsf{Ad}_H : H \to \mathsf{Aut}(\mathfrak{h})$.

Definition 35.8.7 (Local Klein geometry). A local Klein geometry consists of the following data:

1. a Lie group G,
2. a closed subgroup $H \subset G$, and
3. a subgroup $\Gamma \subset G$ acting by covering transformations on G/H such that the left coset space $\Gamma\backslash G/H$ is connected.

35.8.2 Cartan geometry

The definition of a Klein geometry can be rephrased in the language of bundle theory. First, an alternative characterization of Lie groups in terms of the Maurer–Cartan connection is required.

Alternative Definition 35.8.8 (Lie group). Let G be a smooth manifold and let $\mathfrak{g}$ be a Lie algebra. G is a Lie group, with Lie algebra $\mathfrak{g}$, if it comes equipped with a $\mathfrak{g}$-valued 1-form ω satisfying the following conditions:

1. **Maurer–Cartan equation**: $\mathrm{d}\omega + \frac{1}{2}[\omega \wedge \omega] = 0$,
2. **Soldering**: ω restricts to an isomorphism on every fibre, and

3. **Completeness**: ω is complete, i.e. every vector field that maps constantly to $\mathfrak{g}$ is complete.

@@ FIX/COMPLETE THIS PROPERTY @@

In a similar way, Klein geometries can be characterized as follows.

Property 35.8.9. The bundle $\pi : G \to G/H$ of a Klein geometry G/H admits a 1-form $\omega : TG \to \mathfrak{g}$ that satisfies the following conditions:

- ω satisfies the Maurer–Cartan equation,
- ω restricts to an isomorphism on every fibre,
- ω is complete,
- ω is H-equivariant: $R_h^* \omega = \mathrm{Ad}(h^{-1})\omega$, and
- ω cancels $\mathfrak{h}$-fundamental vector fields: $\omega(A^\#) = A$ for all $A \in \mathfrak{h}$.

The last two conditions show that ω defines a principal connection 1-form, while the first condition states that this connection is flat. In fact, this 1-form is exactly the Maurer–Cartan form on G, where the last conditions are obtained by restricting to the subgroup $H \subset G$.

By dropping the flatness and completeness conditions, one obtains the notion of Cartan geometries.

Definition 35.8.10 (Cartan geometry). Consider a principal H-bundle $\pi : P \to M$ and a Lie algebra $\mathfrak{g}$ such that $\mathfrak{h} \leq \mathfrak{g}$ (in general, it is assumed that these form a model geometry). A Cartan geometry is characterized by a 1-form $\omega : TP \to \mathfrak{g}$ satisfying the following conditions:

1. ω restricts to an isomorphism on every fibre,
2. ω is H-equivariant, and
3. ω cancels $\mathfrak{h}$-fundamental vector fields.

The 1-form ω is called the **Cartan connection**.

Definition 35.8.11 (Curvature). By analogy with the Maurer–Cartan condition and the Cartan structure equation (Formula 35.4.17), the curvature of a Cartan connection is defined as follows:

$$\Omega := \mathrm{d}\omega + \frac{1}{2}[\omega \wedge \omega]\,. \tag{35.128}$$

Definition 35.8.12 (Torsion). The torsion of a Cartan connection is defined as the image of its curvature under the projection $\mathfrak{g} \to \mathfrak{g}/\mathfrak{h}$.

By restricting to reductive model spaces, an important decomposition of the Cartan connection is obtained.

Property 35.8.13. Consider a Cartan geometry $\pi : P \to M$ with a reductive model space $(\mathfrak{g}, \mathfrak{h})$ and decompose the Cartan connection as $\omega = \omega_{\mathfrak{h}} + \omega_{\mathfrak{m}}$. This decomposition has the following important properties:

- The form $\omega_{\mathfrak{h}}$ defines a principal connection on the Cartan geometry P.
- The form $\omega_{\mathfrak{m}}$ defines a solder form on M.
- The decomposition of the associated curvature form Ω gives the curvature and torsion of the induced principal connection and solder forms respectively.

Furthermore, the Cartan geometry $\pi : P \to M$ gives a reduction of the frame bundle FM induced by the solder form $\omega_{\mathfrak{m}}$.

35.9 $\mathcal{D}$-geometry ♣

Definition 35.9.1 (Sheaf of differential operators). Let M be a smooth manifold (or scheme, see Chapter 10). Denote its structure sheaf by $\mathcal{O}_M$ and denote the sheaf of $\mathcal{O}_M$-algebras of vector fields (interpreted as derivations) on M by $\mathcal{D}_M$:

$$\mathcal{D}_M(U) := \left\{ v \in \mathsf{End}(\mathcal{O}_M(U)) \mid v(fg) = v(f)g + fv(g) \right\}. \tag{35.129}$$

Definition 35.9.2 ($\mathcal{C}_M$-module). An $\mathcal{O}_M$-module equipped with an action of $\mathcal{D}_M$. This is equivalent to a linear map

$$\nabla : \mathcal{D}_M \to \mathsf{End}(\mathcal{O}_M) \tag{35.130}$$

satisfying the following properties:

1. **$\mathcal{O}_M$-linearity**: $\nabla_{fv}\sigma = f\nabla_v\sigma$,
2. **Leibniz rule**: $\nabla_v(f\sigma) = v(f)\sigma + f\nabla_v\sigma$, and
3. **Flatness**: $\nabla_{[v,w]}\sigma = [\nabla_v, \nabla_w]\sigma$.

If M is locally free (Definition 8.6.5), i.e. corresponds to a locally trivial bundle, one obtains the algebraic reformulation of a vector bundle with flat connection.

Chapter 36

Riemannian Geometry

The main reference for this chapter is Petersen (2006).

Chapter Contents

36.1 Riemannian manifolds **689**
36.1.1 Metric 689
36.1.2 Riemannian manifolds 692
36.1.3 Levi-Civita connection 692
36.1.4 Killing vectors 694
36.2 Curvature **695**
36.3 Spinor bundles **696**
36.3.1 Spin structures 696
36.3.2 Stiefel–Whitney classes 698
36.3.3 Dirac operators ♣ 699
36.4 Conformal structures **702**
36.5 Distributions ♣ **704**

36.1 Riemannian manifolds

36.1.1 Metric

Definition 36.1.1 (Bundle metric). Consider the bundle $T^{(0,2)}M$. From Definition 23.3.1, it follows that every section g of this bundle defines a bilinear map

$$g_p : T_pM \times T_pM \to \mathbb{R} \tag{36.1}$$

for all $p \in M$. If this map is symmetric and nondegenerate, it is called a **Lorentzian** or **pseudo-Riemannian metric**. If the map is also positive, it is called a **Riemannian metric**.[1] The collection $\{\langle \cdot \mid \cdot \rangle_p : p \in M\}$ is called a **bundle metric** or **fibre metric**.

Definition 36.1.2 (Riemannian manifold). A smooth manifold equipped with a Riemannian metric.

A Riemannian metric induces a duality between TM and T^*M (akin to the Riesz representation theorem 25.2.7). This is given by the *flat* and *sharp* isomorphisms.

Definition 36.1.3 (Musical isomorphisms). Let $g : TM \times_M TM \to \mathbb{R}$ be a Riemannian metric on M. The **flat** isomorphism is defined as follows:

$$\flat : v \mapsto g(v, \cdot)\,. \tag{36.2}$$

The **sharp** isomorphism is defined as the inverse map. For an arbitrary covector field ω, it is implicitly given by

$$g(\omega^\sharp, v) = \omega(v)\,. \tag{36.3}$$

These **musical isomorphisms** can be used to raise and lower tensor indices. In index notation, they are given by contraction with metric tensor:

$$\begin{aligned} \flat &: v^\mu \mapsto v_\lambda := g_{\lambda\mu} v^\mu\,, \\ \sharp &: \omega_\mu \mapsto \omega^\lambda := g^{\lambda\mu}\omega_\mu\,. \end{aligned} \tag{36.4}$$

Definition 36.1.4 (Codifferential). Using the de Rham differential and the Hodge star operator (Definition 23.4.27), one can define a boundary operator $\delta : \Omega^k(M) \to \Omega^{k-1}(M)$:

$$\delta := (-1)^k *^{-1} \mathsf{d} * = (-1)^{n(k+1)+1} * \mathsf{d} * \,. \tag{36.5}$$

It is not hard to check that this is indeed a boundary operator in the sense of Definition 5.1.1.

Definition 36.1.5 (Hodge Laplacian[2]). Using the de Rham differential and codifferential, one can define a derivation on $\Omega^k(M)$:

$$\Delta := \mathsf{d}\delta + \delta\mathsf{d}\,. \tag{36.6}$$

It should be noted that, in contrast to the differential d, the Hodge Laplacian depends on the metric through the definition of the codifferential. The square root of this operator is the **Hodge–de Rham operator** $\mathsf{d} + \delta$.

Definition 36.1.6 (Harmonic form). An element of the kernel of the Hodge Laplacian Δ. The space of harmonic k-forms is often denoted by $\mathcal{H}^k(M)$.

[1] Recall Section 22.3 about Hermitian forms and metric forms.

[2] Sometimes called the **Hodge–de Rham** or **Laplace–de Rham** operator.

Theorem 36.1.7 (Hodge decomposition). *Let M be a closed Riemannian manifold. Every differential k-form admits a decomposition of the form*

$$\omega = \mathrm{d}\alpha + \delta\beta + h\,, \tag{36.7}$$

where $h \in \mathcal{H}^k(M)$.

Corollary 36.1.8 (Hodge theorem). The kth de Rham cohomology group is isomorphic (as an Abelian group/vector space) to the space of harmonic k-forms:

$$H^k(M) \cong \mathcal{H}^k(M)\,. \tag{36.8}$$

Remark 36.1.9. The gradient, rotor (curl) and divergence from standard vector calculus (Section 23.1) can be rewritten and generalized using exterior derivatives as follows. Consider a smooth function f and a smooth vector field X.

$$\begin{aligned} \nabla f &= \mathrm{d}f^\sharp \\ \nabla \times X &= \left(*\,\mathrm{d}X^\flat\right)^\sharp \\ \nabla \cdot X &= \left(*^{-1}\mathrm{d} * X^\flat\right)^\sharp, \end{aligned} \tag{36.9}$$

Property 23.1.12 now follows from the nilpotency condition $\mathrm{d}^2 = 0$.

Proof (Explanation). Looking at Eq. (34.41) for the exterior derivative of a smooth function and remembering the definition of the gradient 23.1.1 in $\mathbb{R}^3$, one can see that these two definitions look very similar. The major difference lies in the fact that ∇f is a vector and $\mathrm{d}f$ is a covector. However, because of the nondegeneracy of the metric, there exists an isomorphism between these spaces: the musical isomorphisms.

Similar relations hold for the divergence 23.1.7 and rotor 23.1.9. However, here, one has to use a different construction because the spaces Λ^1 and Λ^2 have to be used. The Hodge star (Definition 23.4.27) can be used to obtain the correct dimensions.

For simplicity, a derivation in $\mathbb{R}^3$ with the Euclidean metric is given. Consider a vector field $X = (f_1, f_2, f_3)$ where the f_i are smooth. Using these functions f_i, one can construct a one-form $X^\flat = f_1\mathrm{d}x_1 + f_2\mathrm{d}x_2 + f_3\mathrm{d}x_3$ and a two-form $*X^\flat = f_1\mathrm{d}x_2 \wedge \mathrm{d}x_3 + f_2\mathrm{d}x_3 \wedge \mathrm{d}x_1 + f_3\mathrm{d}x_1 \wedge \mathrm{d}x_2$. After applying the exterior derivative, one obtains

$$\begin{aligned} \mathrm{d}X^\flat &= \left(\frac{\partial f_3}{\partial x_2} - \frac{\partial f_2}{\partial x_3}\right)\mathrm{d}x_2 \wedge \mathrm{d}x_3 + \left(\frac{\partial f_1}{\partial x_3} - \frac{\partial f_3}{\partial x_1}\right)\mathrm{d}x_3 \wedge \mathrm{d}x_1 + \left(\frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2}\right)\mathrm{d}x_1 \wedge \mathrm{d}x_2\,, \\ \mathrm{d} * X^\flat &= \left(\frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} + \frac{\partial f_3}{\partial x_3}\right)\mathrm{d}x_1 \wedge \mathrm{d}x_2 \wedge \mathrm{d}x_3\,. \end{aligned}$$

□

36.1.2 Riemannian manifolds

Definition 36.1.10 (Riemannian isometry). Consider two pseudo-Riemannian manifolds (M, g_M) and (N, g_N). A smooth function $f : M \to N$ is called an isometry if $f^* g_N = g_M$, i.e. if it preserves the metric tensor.

Property 36.1.11. Let M be a pseudo-Riemannian manifold. For every $p \in M$, there exists a splitting $T_p M = P \oplus N$ where P is a subspace on which the metric is positive-definite and N is a subspace on which the metric is negative-definite. This splitting is not unique, only the dimensions of the two subspaces are well-defined invariants.

Due to the continuity of the metric, the dimensions of this splitting will be the same for all points in a connected neighbourhood. For connected manifolds, this amounts to a global invariant.

Definition 36.1.12 (Index and signature). Let M be a connected pseudo-Riemannian manifold. The dimension of the 'negative' subspace N in the above splitting $T_p P = P \oplus N$ is called the index of the manifold. The **signature** is defined as the pair $(\dim(P), \dim(N))$.

Definition 36.1.13 (Hyperbolic manifold). A pseudo-Riemannian manifold with index 1.

Theorem 36.1.14 (Whitney's embedding theorem). *Every smooth paracompact manifold M can be embedded in $\mathbb{R}^{2\dim(M)}$.*

Theorem 36.1.15 (Whitney's immersion theorem). *Every smooth paracompact manifold M can be immersed in $\mathbb{R}^{2\dim(M)-1}$.*

The following theorem by *Whitney* and *Cohen* is slightly stronger (although the name 'conjecture' remained, it is actually a proven statement).

Theorem 36.1.16 (Immersion conjecture). *Every smooth paracompact manifold M can be immersed in $\mathbb{R}^{2\dim(M)-a(\dim(M))}$, where $a(n)$ is the number of 1s in the binary expansion of n.*

Definition 36.1.17 (Riemannian cone). Let (M, g) be a Riemannian manifold and consider the product manifold $M \times]0, +\infty[$. This manifold can also be turned into a Riemannian manifold by equipping it with the metric $t^2 g + \mathrm{d}t \otimes \mathrm{d}t$. This manifold is called the Riemannian cone or **metric cone** of (M, g).

36.1.3 Levi-Civita connection

Definition 36.1.18 (Riemannian connection). An affine connection ∇ on (M, g) is said to be Riemannian if it satisfies the following two conditions:

1. **metric-compatibility**:

$$\nabla g = 0 \tag{36.10}$$

or, equivalently,

$$X(g(Y,Z)) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z) \tag{36.11}$$

for all vector fields X, Y and Z.

2. **torsion-freeness**:

$$\nabla_X Y - \nabla_Y X = [X, Y] \tag{36.12}$$

for all vector fields X, Y.

A Riemannian connection is also often called a **Levi-Civita connection**.

Because the existence of a metric is equivalent to the integrability of an $\mathsf{O}(n)$-structure by Example 35.5.6, Property 35.5.4 and Property 35.5.10 imply that there exists a torsion-free connection that preserves this structure. The following theorem gives an even sharper result.

Theorem 36.1.19 (Fundamental theorem). *Every pseudo-Riemannian manifold admits a unique Levi-Civita connection.*

Formula 36.1.20 (Koszul formula). The Levi-Civita connection ∇ on a pseudo-Riemannian manifold (M, g) is implicitly defined by the following formula:

$$\begin{aligned} 2g(\nabla_X Y, Z) &= \mathcal{L}_X g(Y,Z) + \mathsf{d}(\iota_X g)(Y,Z) && (36.13)\\ &= X(g(Y,Z)) + Y(g(Z,X)) - Z(g(X,Y)) \\ &\qquad + g([X,Y],Z) - g([Z,X],Y) - g([Y,Z],X)\,. && (36.14) \end{aligned}$$

The following (local) formula is often useful in calculations (especially in general relativity, see Chapter 67).

Formula 36.1.21 (Divergence). Consider the Riemannian volume form

$$\mathsf{Vol} = \sqrt{\det(g)}\mathsf{d}x^1 \wedge \cdots \wedge \mathsf{d}x^n\,. \tag{36.15}$$

The divergence of a vector field X is defined as follows:

$$\mathcal{L}_X \mathsf{Vol} =: \mathsf{div}(X)\,\mathsf{Vol}\,. \tag{36.16}$$

With respect to an orthonormal frame field this is equivalent to

$$\mathsf{div}(X) = \frac{1}{2}\sum_{i=1}^{n} (\mathcal{L}_X g)(e_i, e_i)\,. \tag{36.17}$$

Let ∇ be the Levi-Civita connection. Using the Koszul formula, one can show that the above formula implies the following equality

$$\mathsf{div}(X) = \mathsf{tr}(Y \mapsto \nabla_Y X) \equiv \nabla_\mu X^\mu\,, \tag{36.18}$$

where tr denotes the contraction (or trace) induced by g. This makes it clear that the covariant divergence is indeed a good generalization of the divergence 23.1.7 from vector calculus.

Using the metric determinant, one can locally write the divergence in terms of ordinary partial derivatives:

$$\nabla_\mu V^\mu = \frac{1}{\sqrt{\det(g)}} \partial_\mu (\sqrt{\det(g)} V^\mu) \,. \tag{36.19}$$

Definition 36.1.22 (Laplace–Beltrami operator). Consider a Riemannian manifold (M, g). The Laplace–Beltrami operator on M is defined as the Laplace operator 23.1.19, i.e. as the divergence of the gradient:

$$\Delta := \operatorname{tr} \nabla^2 \,. \tag{36.20}$$

This is, up to a sign[3], equal to the **Bochner Laplacian**:

$$\Delta = -\nabla^* \nabla \,. \tag{36.21}$$

The geodesic equation 30.2.47 can be generalized as follows.

Definition 36.1.23 (Geodesic). A curve γ on a Riemannian manifold (M, g) that is autoparallel with respect to the Levi-Civita connection:

$$\nabla_{\dot\gamma} \dot\gamma = 0 \,. \tag{36.22}$$

@@ COMPLETE (HESSIAN, ...) @@

36.1.4 Killing vectors

Definition 36.1.24 (Killing vector). Let (M, g) be a pseudo-Riemannian manifold. A vector field X satisfying

$$\mathcal{L}_X g = 0 \tag{36.23}$$

is called a Killing vector field.

Formula 36.1.25. Given a Levi-Civita connection ∇ on (M, g), one can rewrite the Killing condition as follows:

$$\nabla_{(\mu} X_{\nu)} = 0 \,. \tag{36.24}$$

The above condition can be generalized to tensor fields.

Definition 36.1.26 (Killing tensor). Let ∇ be the Levi-Civita connection on (M, g). A tensor T satisfying

$$\nabla_{(\mu_N} T_{\mu_1 \dots \mu_{N-1})} = 0 \tag{36.25}$$

is called a Killing tensor.

[3] Some authors already include a minus sign in the definition of the Laplace–Beltrami operator.

36.2 Curvature

Formula 36.2.1 (Riemann curvature tensor). The Riemann (curvature) tensor R is defined as the following $(1,3)$-tensor:

$$R(v,w)z := [\nabla_v, \nabla_w]z - \nabla_{[v,w]}z\,, \tag{36.26}$$

where ∇ is the Levi-Civita connection. Locally, it is given by

$$R^i_{jkl}e_i = R(e_k, e_l)e_j\,. \tag{36.27}$$

Property 36.2.2 (Bianchi identity). The first (or **algebraic**) Bianchi identity reads

$$R(X,Y)Z + R(Y,Z)X + R(Z,X)Y = 0\,. \tag{36.28}$$

The second (or **differential**) Bianchi identity is a similar identity for the covariant derivative:

$$(\nabla_Z R)_{X,Y}W + (\nabla_X R)_{Y,Z}W + (\nabla_Y R)_{Z,X}W = 0\,. \tag{36.29}$$

Formula 36.2.3 (Directional curvature operator[4]).

$$R_v(w) := R(w,v)v \tag{36.30}$$

Formula 36.2.4 (Sectional curvature).

$$\mathsf{sec}(v,w) := \frac{g(R(w,v)v,w)}{g(v,v)g(w,w) - g(v,w)^2} = \frac{g(R_v(w),w)}{g(v\wedge w, v\wedge w)} \tag{36.31}$$

An important result states that the sectional curvature only depends on the span of v, w.

Remark 36.2.5. For surfaces, the sectional curvature coincides with the Gaussian curvature K by Gauss's Theorema Egregium 30.2.51. In general, the sectional curvature gives the Gaussian curvature of the plane spanned by the vectors v, w.

Formula 36.2.6 (Ricci tensor). In coordinate-free notation, the Ricci tensor is defined as the trace of the Riemann tensor:

$$\mathsf{Ric}(v,w) := \mathsf{tr}\big(x \mapsto R(x,v)w\big) \tag{36.32}$$

Equivalently,

$$\mathsf{Ric}(v,w) = \sum_{i=1}^{n} g(R(e_i,v)w, e_i)\,. \tag{36.33}$$

Locally, this becomes:

$$R_{\mu\nu} \equiv \mathsf{Ric}_{\mu\nu} := R^{\lambda}{}_{\mu\nu\lambda}\,. \tag{36.34}$$

[4] Also called the **tidal force operator** (mostly in physics).

Property 36.2.7. The Ricci tensor can also be rewritten in terms of the sectional curvature whenever $\|v\| = 1$. Let $\{e_1, \dots, e_{n-1}\}$ be a set of orthonormal vectors such that $\{e_1, \dots, e_{n-1}, v\}$ forms an orthonormal basis.

$$\mathrm{Ric}(v, v) = \sum_{i=1}^{n-1} \mathrm{sec}(v, e_i) \tag{36.35}$$

It follows, that the Ricci tensor can be interpreted as an averaged (sectional) curvature.

Formula 36.2.8 (Ricci scalar).

$$R := R^{\mu}{}_{\mu} \tag{36.36}$$

This (scalar) quantity is also called the **scalar curvature**.

Formula 36.2.9 (Einstein tensor).

$$G_{\mu\nu} := R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \tag{36.37}$$

Property 36.2.10. For 4-manifolds, the Einstein tensor $G_{\mu\nu}$ is the only tensor containing at most second derivatives of the metric that satisfies

$$\nabla_{\mu} G^{\mu\nu} = 0\,. \tag{36.38}$$

Definition 36.2.11 (Einstein manifold). A Riemannian manifold for which the Ricci tensor is proportional to the metric.

Remark. This name is justified by the fact that Einstein manifolds are exactly the solutions of the *Einstein field equations* (67.2.1) (with a cosmological constant).

36.3 Spinor bundles

In this section, all (pseudo)Riemannian manifolds are assumed to be orientable since this assures that the structure group of TM can be reduced to the special orthogonal group. For the more general definition of Pin-bundles, see Choquet-Bruhat and DeWitt-Morette (2000).

36.3.1 Spin structures

From Definition 22.4.28, it is known that one can define a (complex) vector $v \in V$ as an equivalence class of couples $(c, \mathfrak{b})$, where $\lambda \in \mathbb{C}^{\dim(V)}$ is a (coordinate) vector and ψ is a (linear) frame of V. This will now be extended to Clifford algebras and spinors. Note that this section was not placed in the chapter on Clifford algebras because the natural setting for spinors is Riemannian geometry.

Definition 36.3.1 (Spinor). Let V be a vector space equipped with a metric g of signature (p,q). Consider the set $\mathbb{C}^{2^k} \times F_{\mathsf{SO}}V \times \mathsf{Spin}(p,q)$, where $k = \lfloor \frac{p+q}{2} \rfloor$ and $F_{\mathsf{SO}}V$ is the set of orthonormal frames in V. One can define an equivalence relation on this set as follows. Two triples $(c_1, \mathfrak{b}_1, \Lambda_1)$ and $(c_2, \mathfrak{b}_2, \Lambda_2)$ are identified if and only if

$$c_2 = \Lambda c_1 \qquad \mathfrak{b}_1 = L\mathfrak{b}_2 \qquad \Lambda = \Lambda_2\Lambda_1^{-1} \qquad \rho(\Lambda) = L\,, \tag{36.39}$$

where ρ is the 2-to-1 covering map $\mathsf{Spin}(p,q) \to \mathsf{SO}(p,q)$. An equivalence class of such triples (or a representative thereof) is called a **spinor**. The 2^k numbers in c_1 are called the **components** of the spinor in the **spin frame** $(\mathfrak{b}_1, \Lambda_1)$.

It should be noted that the two elements of a spin frame $(\mathfrak{b}, \Lambda)$ are not independent. Choose a 'fiducial frame' $(\mathfrak{b}_0, e)$, where e is the identity element of $\mathsf{Spin}(p,q)$. The couple $(\mathfrak{b}, \Lambda)$ is a well-defined spin frame if and only if $\rho(\Lambda) = L$ whenever $\mathfrak{b} = L\mathfrak{b}_0$. Note that different choices of fiducial frame give different, yet isomorphic, spinor spaces.

Definition 36.3.2 (Spinor field). Let (M,g) be a (pseudo)Riemannian manifold. Every tangent space T_pM is a vector space equipped with a nondegenerate bilinear form and, hence, one can use the above definition to construct a spinor space at p. If the orthonormal frame bundle $F_{\mathsf{SO}}M$ is trivial, one can choose a section $p \mapsto \mathfrak{b}(p)$ and define the fiducial frame field to be $p \mapsto (\mathfrak{b}(p), e)$.

However, if $F_{\mathsf{SO}}M$ is not trivial, this construction only works locally. To be able to extend it to the whole manifold, one needs to patch the different frame fields together. The required compatibility conditions reads as follows:

$$\rho(\Lambda_i(x)\Lambda_j^{-1}(x)) = L_{ij}(x) \tag{36.40}$$

whenever

$$\mathfrak{b}_i(x) = L_{ij}(x)\mathfrak{b}_j(x) \tag{36.41}$$

for all $x \in U_i \cap U_j$.

It can be shown that a manifold admits the definition of a global spin frame field if and only if it admits a spin structure.

Definition 36.3.3 (Spin structure). Consider the orthonormal frame bundle

$$\pi_{\mathsf{SO}} : F_{\mathsf{SO}}M \to M$$

that is obtained by reducing the structure group of the frame bundle FM from $\mathsf{GL}(n)$ to $\mathsf{SO}(n)$.

The smooth manifold M is said to have a spin structure (cf. Definition 35.5.2) if there exists a principal $\mathsf{Spin}(n)$-bundle $\pi_{\mathsf{Spin}} : F_{\mathsf{Spin}} \to M$ and an equivariant 2-fold lifting of F_{SO} to F_{Spin}, i.e. a morphism $\xi : F_{\mathsf{Spin}} \to F_{\mathsf{SO}}M$ together with the 2-fold covering map $\rho : \mathsf{Spin}(n) \to \mathsf{SO}(n)$ that satisfy

- $\pi_{\mathrm{SO}} \circ \xi = \pi_{\mathrm{Spin}}$, and
- $\xi(p \lhd g) = \xi(p) \cdot \rho(g)$

for all $g \in \mathsf{Spin}(n)$, where $\lhd$ and $\cdot$ denote the right actions of the respective structure groups. If M admits a spin structure, it is called a **spin manifold** and the principal $\mathsf{Spin}(n)$-bundle F_{Spin} is called the **spin frame bundle** (of M).

Definition 36.3.4 (Spinor bundle). The vector bundle associated to the spin frame bundle induced by the (fundamental) spinor representation (Definition 27.4.15). Sections of the spinor bundle are called **spinor fields**. If the dimension of the manifold is even, the spinor representation decomposes into two irreducible representations, the associated vector bundles are called bundles of **half-spinors**.

More generally, given a spin representation $\rho : \mathsf{Spin}(n) \to \mathsf{Aut}(V)$, any associated bundle $F_{\mathrm{Spin}} \times_\rho V$ is called a spinor bundle.

Example 36.3.5. A 3-manifold is spin if it is compact and orientable.

Example 36.3.6. Any (stably) parallelizable manifold is spin.

36.3.2 Stiefel–Whitney classes

Definition 36.3.7 (Stiefel–Whitney classes). Consider a rank-n vector bundle E over a paracompact space M (hence admitting an $\mathsf{O}(n)$-reduction). The Stiefel–Whitney classes of E are characteristic classes $w_n(E) \in H^\bullet(M; \mathbb{Z}_2)$ defined by the following conditions:

1. **Normalization**: The tautological line bundle over the circle, i.e. the Möbius strip, has a nontrivial w_1.
2. **Rank**: $w_0(E) = 1$ and $w_i(E) = 0$ for all $i > n$.
3. **Whitney sum formula**: $w_k(E \oplus E') = \bigoplus_{i=0}^k w_i(E) \smile w_{k-i}(E')$ for all finite-rank vector bundles $E' \to M$.
4. **Naturality**: Every morphism $f : M \to M'$ of base manifolds induces a pullback-isomorphism $w(f^*E) = f^*w(E)$.

Property 36.3.8 (Independent sections). Consider a rank-n vector bundle $\pi : E \to M$. If $w_i(E)$ vanishes, the restriction of E to the i-skeleton of M admits $n - i + 1$ linearly independent sections.

Consider a principal $\mathsf{O}(n)$-bundle $\pi : P \to M$ together with an open cover $\{U_i\}_{i \in I}$ of M. This bundle is locally defined by a Čech 2-cocycle $g_{ij} : U_i \cap U_j \to \mathsf{O}(n)$. Using the

two-fold covering map $\rho : \mathsf{Pin}(n) \to \mathsf{O}(n)$, this cocycle can be used to define a function

$$p_{ijk} : U_i \cap U_j \cap U_k \to \mathsf{Spin}(n) : x \mapsto \gamma_{jk}(x)\gamma_{ik}^{-1}(x)\gamma_{ij}(x)\,, \tag{36.42}$$

where γ_{ij} is a lift of the cocycle g_{ij} to $\mathsf{Pin}(n)$. It can be shown that this defines a 2-cocycle with values in $\mathbb{Z}_2$, i.e. $[p_{ijk}] \in \check{H}^2(M;\mathbb{Z}_2)$. If this cocycle vanishes, the O-bundle admits a well-defined Pin-lift in the sense of Definition 35.5.1.

Property 36.3.9 (Stiefel–Whitney classes). For a $\mathsf{Pin}^+(n)$-extension of an $\mathsf{O}(n)$-bundle, the cocycle p satisfies $[p] = w_2$. However, for a $\mathsf{Pin}^-(n)$-extension of an $\mathsf{O}(n)$-bundle, the cocycle p satisfies $[p] = w_2 + w_1 \smile w_1$. So, in general $[p]$ is not a Stiefel–Whitney class. However, note that if w_1 vanishes, i.e. if the bundle is orientable, the cocycle does reduce to the second Stiefel–Whitney class.

The classification of spin manifolds can be stated in terms of characteristic classes. However, instead of the usual $\mathbb{R}$- or $\mathbb{Z}$-valued cohomology classes, one needs classes in $\mathbb{Z}_2$-cohomology.

Property 36.3.10. Consider the orientation line bundle (Definition 34.8.4) of an $\mathsf{O}(n)$-bundle. The cocycle of this bundle define a cohomology class $w_1 \in H^1(M;\mathbb{Z}_2)$. It can be shown that this class is the first Stiefel–Whitney class.

Corollary 36.3.11 (Orientability). A smooth manifold is orientable if and only if its first Stiefel–Whitney class vanishes.

Moreover, this class has a geometric interpretation. Consider any triangulation of the base manifold. A choice of orientation consists of a consistent assignment of a sign to every edge in the 1-skeleton (or, equivalently, a trivialization of the bundle over the 1-skeleton). The total change

@@ COMPLETE THIS RESULT @@

Property 36.3.12 (Spin manifold). A smooth, orientable manifold M is spin if and only if its second Stiefel–Whitney class vanishes. Furthermore, the distinct spin structures form an affine space over $H^1(M;\mathbb{Z}_2)$.

Definition 36.3.13 (Integral classes). Consider the Bockstein morphism $\beta : H^i(M;\mathbb{Z}_2) \to H^{i+1}(M;\mathbb{Z})$ associated to the mod-2 reduction $\mathbb{Z} \to \mathbb{Z}_2$. The integral Stiefel–Whitney classes are defined as the images $\beta(w_i) \in H^{i+1}(M;\mathbb{Z})$.

Property 36.3.14 ($\mathsf{Spin}^{\mathbb{C}}$-structures). An orientable vector bundle admits a $\mathsf{Spin}^{\mathbb{C}}$-structure if and only if its third integral Stiefel–Whitney class vanishes.

36.3.3 Dirac operators ♣

In this section, the partial derivatives ∂_i and gradient operator $\sum_{i=1}^n e_i\partial_i$ are generalized to Clifford algebras and Clifford modules.

Definition 36.3.15 (Clifford bundle). Consider a (pseudo-)Riemannian manifold (M, g) of signature (p, q). For every point $p \in M$, one can construct a Clifford algebra associated to the tangent space (T_pM, g_p). Using these Clifford algebras, one can construct an associated bundle to TM that has $C\ell_{p,q}(\mathbb{R})$ as its typical fibre. A vector bundle with a Clifford algebra as typical fibre, for which the local trivializations respect the algebra structure, is called a Clifford bundle.[5]

The following criterion gives an algebraic characterization of Spin-manifolds.

Property 36.3.16 (Plymen's criterion). A Riemannian manifold admits a $\mathrm{Spin}^{\mathbb{C}}$-structure if and only if the C^*-algebra of continuous section of its Clifford bundle is *Morita-equivalent* to the C^*-algebra of continuous functions.[6]

Spinor bundles are a particular instance of the following notion.

Definition 36.3.17 (Clifford module bundle). Consider a (pseudo-)Riemannian manifold (M, g) with its associated Clifford bundle $C\ell(TM)$. Any vector bundle that admits a left $C\ell(TM)$-action is called a Clifford module (bundle) over M.

To be able to define a Dirac operator on spinor bundles, one first needs to define the Dirac operator on $\mathbb{R}^n$. This Dirac operator is obtained by composing the ordinary gradient

$$\partial := \sum_{i=1}^{n} e_i \partial_i \tag{36.43}$$

with the linear injection $\iota_{C\ell} : e_i \mapsto \gamma_i$ that sends a basis of $\mathbb{R}^n$ to the corresponding generators of $C\ell_n(\mathbb{R})$:

$$\underline{\partial} := \sum_{i=1}^{n} \gamma_i \partial_i \,. \tag{36.44}$$

To extend this definition to Clifford modules, one simply needs to replace partial derivatives by covariant derivatives as usual.

Property 36.3.18 (Induced Clifford connection). Let (M, g) be a (pseudo-)Riemannian manifold and let ∇ be the associated Levi-Civita connection. For every Clifford module E over M, there exists a unique connection $\nabla^E : \Gamma(E) \to \Gamma(T^*M \otimes E)$ that respects the Clifford action:

$$\nabla^E(\iota_{C\ell}(X) \cdot \sigma) = \iota_{C\ell}(\nabla X) \cdot \sigma + \iota_{C\ell}(X) \cdot \nabla^E \sigma \,, \tag{36.45}$$

[5] Note that one can use this construction to turn any vector bundle that admits a fibre metric into a Clifford bundle. Moreover, this is a particular instance of the more general notion of *C^*-algebra bundles* (see also Definition 59.4.1).

[6] Note that Morita-equivalence of C^*-algebras is more involved than the similar notion (Definition 2.4.56) for ordinary rings.

where $\iota_{C\ell} : TM \to C\ell(TM)$ is the canonical map that embeds a vector field in $C\ell(TM)$. This connection is induced by the following canonical connection form on the spin frame bundle F_{spin}:

$$\omega_{\text{spin}} := \rho_*^{-1}(\xi^*\omega)\,, \tag{36.46}$$

where ω is the Levi-Civita connection form.

Definition 36.3.19 (Dirac operator). Consider a (pseudo-)Riemannian manifold (M, g) together with a Clifford module E. If ∇^E is the compatible connection from the previous property, the Dirac operator on E is defined as follows:

$$\not{D} := \sum_{i=1}^{n} \gamma_i \cdot \nabla^E_{e_i}\sigma\,, \tag{36.47}$$

where $\{e_i\}_{i\leq n}$ is a local (orthonormal) frame field.

Property 36.3.20 (Ellipticity). The Dirac operator is a self-adjoint elliptic differential operator (Definition 34.11.1).

Definition 36.3.21 (Index). In many cases, the Clifford module E will be obtained by taking the tensor product of some vector bundle with a spinor bundle S. Such a spinor bundle decomposes as

$$S \cong S^+ \oplus S^- \tag{36.48}$$

under Clifford multiplication, which in turn induces a decomposition of E. The Dirac operator $\not{D}$ interchanges these spaces:

$$\not{D}S^+ = S^- \qquad\qquad \not{D}S^- = S^+\,. \tag{36.49}$$

By the previous property, $\not{D}$ is elliptic and, if M is compact, one can show that this implies that it is also Fredholm (Definition 26.1.37). The index of $\not{D}$ is defined as the Fredholm index of $\not{D}|_{E^+}$:

$$\mathsf{ind}(\not{D}) := \dim\ker(\not{D}|_{E^+}) - \dim\mathsf{coker}(\not{D}|_{E^+})\,. \tag{36.50}$$

The symbol $\sigma(\mathrm{D}) : \pi^*E \to \pi^*F$, where $\pi : T^*M \to M$ is the cotangent bundle projection, of every elliptic differential operator restricts to an isomorphism on all sections except the zero section, so it induces a class in $K(T^*M, T^*M_0)$. One can then use the (relative) Chern character (Construction 41.2.45) to obtain a (compactly supported[7]) cohomology class on T^*M and apply the Thom isomorphism 34.9.26 to obtain a (compactly supported) cohomology class on M.

[7]The Chern character gives a relative cohomology class in $H^\bullet(T^*M, T^*M_0) \cong H^\bullet(D(T^*M), S(T^*M))$. The (unit) sphere bundle over a compact manifold is compact, so this class injects into the compactly supported cohomology (Definition 34.9.15) of T^*M.

However, because K-theory is a (multiplicative generalized Eilenberg–Steenrod) cohomology theory, it also admits a Thom isomorphism. So, one could first map relative K-theory (or the reduced K-theory of the Thom space) to the K-theory of T^*M and then apply the Chern character.

These two approaches are not equivalent. The Thom class, which induces the Thom isomorphism through the cup product, is sent to the image of the Todd class under the Thom isomorphism:

$$\mathsf{ch}(\mathcal{T}(1)) = \mathsf{ch}(\mathsf{th}(E)) = \mathcal{T}(\mathsf{td}(E)), \tag{36.51}$$

where 1 denotes the multiplicative unit in K-theory and td is the Todd class (Example 35.6.15). This also implies that

$$\mathsf{ch}(\mathcal{T}_K(x)) = \mathcal{T}\big(\mathsf{ch}(x) \cup \mathsf{td}(E)\big), \tag{36.52}$$

for every class $x \in K(M)$. These considerations lead to the following definition of the index of D (here, it is assumed that a Riemannian metric is provided such that TM and T^*M can be identified).

Definition 36.3.22 (Topological index). Consider a compact n-dimensional manifold M together with an elliptic differential operator D.

$$\begin{aligned}\mathsf{ind}(D) &:= (-1)^{n(n+1)/2} \int_M \mathcal{T}^{-1}\big(\mathsf{ch}(\sigma(\mathsf{D}))\big) \cup \mathsf{td}(T_{\mathbb{C}}M) \\ &= (-1)^n \int_{TM} \mathsf{ch}(\sigma(\mathsf{D})) \cup \mathsf{td}(T_{\mathbb{C}}M), \end{aligned} \tag{36.53}$$

where $\mathcal{T}$ is the Thom isomorphism on TM.

Corollary 36.3.23. The index of an elliptic differential operator on an odd-dimensional manifold vanishes.

Property 36.3.24. The $\widehat{A}$-genus (Example 35.6.16) of a spin manifold is an integer.

Formula 36.3.25 (Lichnerowicz formula). Let $\not{D}$ be the Dirac operator on a spinor bundle S. Unlike in the Euclidean case, the Dirac operator does not square to the (Bochner) Laplacian:

$$\not{D}^2 = \nabla^*\nabla + \frac{1}{4}R, \tag{36.54}$$

where R is the scalar (Ricci) curvature.

36.4 Conformal structures

Definition 36.4.1 (Conformal transformation). Consider two (pseudo-)Riemannian manifolds (M, g) and (M', g'). A smooth function $f : M \to M'$ is said to be conformal

if it leaves the metric invariant up to a scale transformation (compare this to Definition 30.2.12):

$$f^* g' = \Omega g \tag{36.55}$$

for some smooth positive function $\Omega : M \to \mathbb{R}^+$. If f is a diffeomorphism, it is called a **conformal transformation**.

Infinitesimally, these maps are characterized by a special type of vector field.

Definition 36.4.2 (Conformal Killing vector). Consider a pseudo-Riemannian manifold (M, g). A vector field X is called a conformal Killing vector field, with **conformal factor** $\kappa : M \to \mathbb{R}$, if it satisfies

$$\mathcal{L}_X g = \kappa g \,. \tag{36.56}$$

In local coordinates, this amounts to

$$\nabla_\mu X_\nu + \nabla_\nu X_\mu = \kappa g_{\mu\nu} \,, \tag{36.57}$$

where ∇ is the Levi-Civita connection associated to (M, g). Equivalently, a vector field is a conformal Killing vector field if its flow determines a conformal transformation.

By parametrizing an infinitesimal transformation as $x^\mu \to x^\mu + \varepsilon^\mu$, one obtains the following infinitesimal generators:

- **Translations**: $a^\mu \partial_\mu$,
- **Rotations** (orthogonal transformations): $\omega^\mu{}_\nu x^\nu \partial_\mu$,
- **Dilations**: $\lambda x^\mu \partial_\mu$, and
- **Special conformal transformations**: $x^2 b^\mu \partial_\mu - 2(b \cdot x) x^\mu \partial_\mu$.

As usual, exponentiating these generators gives the finite transformations. One immediately notices that the Poincaré group is a subgroup of the conformal group. However, the conformal group of a (pseudo-)Riemannian manifold M is not just the group of conformal transformations of M.

Definition 36.4.3 (Conformal compactification). Consider a (pseudo-)Riemannian manifold (M, g). A conformal compactification of M is a compact manifold N such that there exists a conformal transformation that embeds M as a dense, open subspace of N.

Definition 36.4.4 (Conformal group). The conformal group $\mathsf{Conf}(M)$ is the connected component of the identity of the conformal diffeomorphism group of the conformal compactification of M.

Property 36.4.5. The conformal group of a (pseudo-)Euclidean space of signature (p, q) is isomorphic to $\mathsf{SO}(p+1, q+1)$.

36.5 Distributions ♣

For more information on the theory of distributions on Euclidean space, see Chapter 19. There are two ways to introduce distributions on general manifolds. Either one uses the locally Euclidean character, defines distributions on charts and glues them together using some compatibility data (see for example Choquet-Bruhat et al. (1991)), or one defines them as the dual of the space of smooth functions (with compact support) as in the Euclidean case. In this section, the second approach is followed.

The base manifold M will be required to be paracompact and second countable. Moreover, it is assumed that a Riemannian metric g is given. This data allows to turn the space of smooth sections of any tensor bundle over M into a Fréchet space (Definition 25.3.10) using a generalization of the seminorms (19.2), where the (partial) derivatives ∂_i are replaced by covariant derivatives ∇_i. The norm will now also be the one induced (fibrewise) by g. In a similar way, one can, for every compact subset $K \subseteq M$, define the space $\mathcal{D}(K, \otimes^p)$ of smooth p-tensor fields with support in K. By taking the direct limit (with its associated topology), one obtains the space of smooth, compactly supported p-tensor fields $\mathcal{D}(M, \otimes^p)$.

Definition 36.5.1 (Tensor distribution). The space of tensor distributions of order $p \in \mathbb{N}$ is the continuous dual of $\mathcal{D}(M, \otimes^p)$.

Much of the theory of distributions on Euclidean space can be generalized to smooth manifolds without too much trouble (for example, one again obtains a dense inclusion $\mathcal{D} \hookrightarrow \mathcal{D}'$). An interesting generalization is the definition of the covariant derivative.

Definition 36.5.2 (Covariant derivative). Let (M, g) be a Riemannian manifold with associated Levi-Civita connection ∇. The covariant derivative of a tensor distribution T is defined using duality as follows (as in the case of Euclidean space, this can be interpreted as an extension of the integration-by-parts formula):

$$\langle \nabla T, \sigma \rangle := -\langle T, g \cdot \nabla \sigma \rangle \,, \tag{36.58}$$

where $g \cdot \nabla \sigma$ denotes the **internal contraction** (generalizing the divergence of a vector field) which, in local coordinates, is given by

$$(g \cdot \nabla \sigma)^{i_1 \dots i_p} = \nabla_j \sigma^{j i_1 \dots i_p} \,. \tag{36.59}$$

Definition 19.3.21 can easily be generalized to smooth manifolds.

Definition 36.5.3 (Wave front set). The wave front set of a distribution $\phi \in D'(M)$ is defined as follows:

$$\mathrm{WF}(\phi) := \left\{ (x, v) \in T^* M_0 \mid v \in \Sigma_x(\phi) \right\} , \tag{36.60}$$

where the singular fibre $\Sigma_x(\phi)$ is defined as in the Euclidean case (since, by localization, one can restrict to a chart containing x and work in local coordinates).

Chapter 37

Symplectic Geometry

References for this chapter are Cardin (2015); McDuff and Salamon (2017).

Chapter Contents

37.1 Symplectic manifolds **706**

37.2 Poisson manifolds **709**

37.3 Lagrangians **711**

37.3.1 Lagrangian submanifolds 711

37.3.2 Maslov classes 713

37.3.3 Polarizations 716

37.4 Hamiltonian dynamics **717**

37.4.1 Dynamical systems 717

37.4.2 Hamilton–Jacobi equation 719

37.4.3 Integrability 720

37.5 Symplectic reduction **723**

37.5.1 Hamiltonian actions 723

37.5.2 Symplectic reduction 725

37.5.3 Poisson reduction 725

37.5.4 Lie–Poisson reduction 726

37.5.5 Infinite-dimensional manifolds 727

37.6 Metaplectic structures **727**

37.1 Symplectic manifolds

Definition 37.1.1 (Symplectic form). Le M be a smooth manifold. A two-form $\omega \in \Omega^2(M)$ is said to be symplectic if it satisfies the following properties:

1. **Closedness**: $\mathrm{d}\omega = 0$, and
2. **Nondegeneracy**: $\iota_X \omega = 0 \implies X = 0$.

If the closedness condition is dropped, one obtains an **almost symplectic form**.

Definition 37.1.2 (Symplectic manifold). A manifold equipped with a symplectic form.

Definition 37.1.3 (Hamiltonian vector field). Let (M, ω) be a symplectic manifold. For every function $f \in C^\infty(M)$, the associated Hamiltonian vector field X_f is defined by the following equation[1]:

$$\omega(X_f, \cdot) = -\mathrm{d}f\,. \tag{37.1}$$

This can be rewritten using $\omega^\sharp$ as

$$X_f = \omega^\sharp(-\mathrm{d}f, \cdot)\,. \tag{37.2}$$

Hamiltonian vector fields form a Lie subalgebra of the Lie algebra of symplectic vector fields. The flow associated to a Hamiltonian vector field is sometimes called a **Hamiltonian symplectomorphism**.[2]

Property 37.1.4 (Dimension). From the antisymmetry and the nondegeneracy of the symplectic form, it follows that the manifold is necessarily even-dimensional.

Theorem 37.1.5 (Darboux). *Let (M, ω) be a symplectic manifold. For every neighbourhood Ω in T^*M, there exists an adapted coordinate system (q^i, p^i) such that*

$$\omega|_\Omega = \sum_i \mathrm{d}p^i \wedge \mathrm{d}q^i\,. \tag{37.3}$$

The adapted charts in this theorem are called **Darboux charts**. Any symplectic manifold admits a covering by Darboux charts.

Remark 37.1.6. The Darboux theorem shows that all symplectic manifolds of the same dimension are locally isomorphic and, therefore, admit no local invariants. This is in stark contrast to, for example, Riemannian manifolds.

[1] A lot of different conventions exist in the literature. Here, the one that is compatible with the *Hamiltonian equations* 52.3.7 (which are universally accepted) is used.

[2] The fact that the Hamiltonian flow indeed preserves the symplectic form follows from the closedness of ω.

Formula 37.1.7. In Darboux coordinates, the components of the symplectic form ω are given by

$$\omega = \begin{pmatrix} 0 & -\mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix}. \tag{37.4}$$

Using the nondegeneracy condition, one can define the 'dual' or inverse $\omega^\sharp$ as

$$\omega^\sharp = \begin{pmatrix} 0 & \mathbb{1} \\ -\mathbb{1} & 0 \end{pmatrix}. \tag{37.5}$$

Note that the literature is very divided on what convention to use. Some authors use the opposite of the above convention.

Property 37.1.8. In the language of G-structures (35.5), one can restate the definition of symplectic manifolds. A smooth $2n$-dimensional manifold is almost symplectic exactly if it admits an $\mathsf{Sp}(2n, \mathbb{R})$-structure. It is symplectic exactly if the $\mathsf{Sp}(2n, \mathbb{R})$-structure is integrable, which by Darboux's theorem is equivalent to first-order integrability, i.e. $\mathrm{d}\omega = 0$.

Definition 37.1.9 (Symplectic potential). By the Poincaré lemma 34.9.8, the symplectic form ω locally defines a one-form:

$$\omega = \mathrm{d}\theta\,. \tag{37.6}$$

This one-form is sometimes called the symplectic potential of (M, ω).

Construction 37.1.10 (Liouville one-form). Let M be a smooth manifold. The cotangent bundle $\pi : T^*M \to M$ comes equipped with a canonical symplectic form that can be derived from a **tautological one-form**. In local coordinates on T^*M, it is given by

$$\alpha := p_i \mathrm{d}q^i\,. \tag{37.7}$$

In coordinate-free notation, this can be written as follows:

$$\alpha|_y(X) = y(\pi_* X)\,, \tag{37.8}$$

where $X \in T_y T^*M$. This one-form serves as a symplectic potential for the cotangent bundle:

$$\omega = \mathrm{d}\alpha\,.$$

Definition 37.1.11 (Liouville vector field). Consider a symplectic manifold (M, ω). A vector field X is called a Liouville vector field if it preserves the symplectic form:

$$\mathcal{L}_X \omega = \omega\,. \tag{37.9}$$

By Cartan's magic formula, a global symplectic potential is then given by

$$\theta := X \lrcorner\, \omega\,. \tag{37.10}$$

It follows that the existence of a Liouville vector field implies the exactness of M.

Definition 37.1.12 (Multisymplectic structure). The definition of a symplectic structure can be generalized to multisymplectic or n**-plectic** structures as $(n+1)$-forms ω that satisfy the conditions:

1. **Closedness**: $\mathrm{d}\omega = 0$, and
2. **Nondegeneracy**: $\iota_X \omega = 0 \implies X = 0$.

Construction 37.1.13 (Tautological form). The construction of the tautological one-form on cotangent bundles T^*M can be generalized to exterior powers of the cotangent bundle through Eq. (37.8):

$$\alpha|_y(X_1, \dots, X_n) := y\big(\pi_* X_1, \dots, \pi_* X_n\big)\,, \tag{37.11}$$

where $X_1, \dots, X_n \in T_y \Lambda^n T^* M$.

Example 37.1.14 (Killing form). The transgression 3-cocycle (Example 32.6.11) induced by the Killing form of a compact simple Lie group turns this group into a 2-plectic manifold.

Definition 37.1.15 (Symplectomorphism). An isomorphism of symplectic manifolds, i.e. a diffeomorphism $f : (M, \omega_M) \to (N, \omega_N)$ satisfying

$$f^* \omega_N = \omega_M\,. \tag{37.12}$$

These functions form an infinite-dimensional Lie group called the **symplectomorphism group**. This should not be confused with the symplectic group $\mathsf{Sp}(n)$.

Definition 37.1.16 (Symplectic vector field). A vector field whose flow preserves the symplectic form ω:

$$\mathcal{L}_X \omega = 0\,. \tag{37.13}$$

Equivalently, a vector field is symplectic if its flow is a symplectomorphism. These vector fields form a Lie subalgebra of $\mathfrak{X}(M)$.

Theorem 37.1.17 (Gromov's nonsqueezing theorem[3]). *Consider the Euclidean ball and cylinder $B^{2n}(r)$ and $Z^{2n}(R)$ in $\mathbb{R}^{2n}$ with their standard (induced) symplectic forms. If there exists a symplectic embedding $B^{2n}(r) \hookrightarrow Z^{2n}(R)$, then $r \leq R$.*

This theorem implies that one cannot arbitrarily squeeze a symplectic manifold while preserving its (symplectic) volume. This hints towards the existence of certain (global) invariants.

Definition 37.1.18 (Symplectic capacity). Consider the category of symplectic manifolds with symplectic embeddings as morphisms. A symplectic capacity is a functor C from this category to the poset of (nonnegative) real numbers (this just means that it is a monotonic function on symplectic manifolds) satisfying the following conditions:

[3] Also called the theorem of the **symplectic camel** (after a paper by *Stewart*).

1. **Homogeneity**: If $\alpha \neq 0$, $C(M, \alpha\omega) = |\alpha| C(M, \omega)$, and
2. **Normalization**: $C(B^{2n}(r)) = C(Z^{2n}(r)) = \pi r^2$.

Example 37.1.19. Inspired by *Gromov*'s theorem, one can obtain capacities by taking the largest ball that embeds into the manifold and the smallest cylinder such that the manifolds embeds in it. These define the minimal and maximal capacities, respectively.

37.2 Poisson manifolds

Definition 37.2.1 (Poisson bracket). Let (M, ω) be a symplectic manifold. The Poisson bracket of two functions $f, g \in C^\infty(M)$ is defined as[4]

$$\{f, g\} := X_f(g) \tag{37.14}$$

or, equivalently, as

$$X_{\{f,g\}} := [X_f, X_g], \tag{37.15}$$

where X_f, X_g are the Hamiltonian vector fields associated to f and g. In Darboux coordinates, the Poisson bracket of the coordinates is represented by the dual matrix $\omega^\sharp$. The bracket operation represented by the symplectic form itself is often called the **Lagrange bracket**.

Property 37.2.2. The Poisson bracket induced by the symplectic form turns the structure $(C^\infty(M), \{\cdot, \cdot\})$ into a Lie algebra and the second equation above gives a (surjective[5]) Lie algebra morphism

$$(C^\infty(M), \{\cdot, \cdot\}) \to (\{X \mid X \text{ is a HVF on M}\}, [\cdot, \cdot]).$$

Furthermore, together with the pointwise multiplication, this structure becomes a Poisson algebra (Definition 32.2.5).

Definition 37.2.3 (Poisson manifold). A smooth manifold M for which the Poisson bracket on $C^\infty(M)$ induces a Poisson algebra structure (Definition 32.2.5). This is equivalently encoded in a bivector field $\Pi \in \Gamma(\Lambda^2 TM)$, called the **Poisson tensor**, such that

$$\{f, g\} = \mathrm{d}f \wedge \mathrm{d}g(\Pi) \tag{37.16}$$

and

$$[\Pi, \Pi] = 0 \tag{37.17}$$

[4]Note that, as for Hamiltonian vector fields, different conventions are used throughout the literature.
[5]The kernel is given by the constant functions and, hence, it is not a bijection.

with respect to the Schouten–Nijenhuis bracket (Definition 34.3.24). The first equation can also be rephrased in terms of the Schouten–Nijenhuis bracket:

$$\{f, g\} = \big[[\Pi, f], g\big] \,. \tag{37.18}$$

Every bivector field that satisfies Eq. (37.17) arises as the Poisson tensor of some Poisson structure.

Property 37.2.4. Every symplectic manifold is a Poisson manifold. The converse, however, is not true.

Property 37.2.5 (Symplectic foliation). The Poisson tensor $\Pi \in \Gamma(\Lambda^2 TM)$ on a Poisson manifold M induces a linear map

$$\pi^\# : \Omega^1(M) \to \mathfrak{X}(M) \tag{37.19}$$

through the interior product. It can be shown that, on **regular Poisson manifolds**, i.e. where the fibrewise images of the above map all have the same dimension, $\pi^\#(T^*M)$ is an integrable subbundle. Moreover, the associated foliation is symplectic, i.e. the restriction of Π to each leaf is nondegenerate.

Definition 37.2.6 (Casimir function). Let $(M, \{\cdot,\cdot\})$ be a Poisson manifold. A function $C \in C^\infty(M)$ is called a Casimir (function) if it satisfies

$$\{C, f\} = 0 \tag{37.20}$$

for all $f \in C^\infty(M)$.

The following property implies that the level sets of a Casimir function retrieve the symplectic foliation of a Poisson manifold.[6]

Property 37.2.7. On symplectic manifolds, Casimir functions are constant on connected components. Moreover, a function on a Poisson manifold M is Casimir if and only if it is constant on every leaf of the symplectic foliation of M.

By generalizing Definition 37.2.3 even more, one obtains Jacobi manifolds.

Definition 37.2.8 (Jacobi manifold). A smooth manifold on which the algebra of smooth functions can be equipped with a Lie algebra structure $\{\cdot,\cdot\}$ such that

$$\mathsf{supp}(\{f, g\}) \subseteq \mathsf{supp}(f) \cap \mathsf{supp}(g) \,. \tag{37.21}$$

This is equivalent to the existence of a vector field $v \in \Gamma(TM)$ and a bivector field $\Pi \in \Gamma(\Lambda^2 TM)$ such that

$$\{f, g\} = (f\mathsf{d}g - g\mathsf{d}f)(v) + \mathsf{d}f \wedge \mathsf{d}g(\Pi) \,. \tag{37.22}$$

[6] To actually obtain the individual leaves, a set of Casimir function is required (similar to the conditions for integrability in Section 37.4.3).

Remark 37.2.9 (Multisymplectic geometry). Most of this section can be generalized to multisymplectic manifolds. For example, the definitions of Hamiltonian vector fields and the induced Lie algebra structure remain virtually the same. However, given an $(n-1)$-form ζ there might not exist a vector field X_ζ that satisfies Eq. (37.1). If one restricts to the subspace $\mathsf{Ham}(M)$ of the **Hamiltonian forms** that do induce a Hamiltonian vector field, one can define a generalized Poisson structure as follows:

$$\{\zeta, \xi\} := \mathcal{L}_{X_\xi}\zeta\,. \tag{37.23}$$

Since the Lie derivative of a function along a vector field is equal to the action of the vector field on the function, it can be seen that this definition reduces to the ordinary one in the case of $n = 1$. For $n > 1$, the structure is not exactly Poisson because the bracket is only antisymmetric up to an exact form. Following Baez, Hoffnung, and Rogers (2010), this will be called the **hemibracket**.

For $n = 1$, one can equivalently define the Poisson (semi)bracket as

$$\{f, g\}_s := \iota_{X_g}\iota_{X_f}\omega\,. \tag{37.24}$$

For $n > 1$, however, this structure differs from the hemibracket by an exact form:

$$\{\zeta, \xi\} = \{\zeta, \xi\}_s + \mathrm{d}\iota_{X_\zeta}\xi\,. \tag{37.25}$$

The **semibracket** $\{\cdot,\cdot\}_s$ now satisfies antisymmetry, but it fails to satisfy the Jacobi identity. Although these brackets do not define a Lie algebra, they do define a Lie 2-algebra (Definition 4.9.6), where $L_1 := C^\infty(M)$ and $L_0 := \mathsf{Ham}(M)$. They are isomorphic as Lie 2-algebras.

Remark 37.2.10 (Presymplectic manifolds). A manifold equipped with a closed two-form (sometimes required to satisfy certain conditions on its rank). In contrast to symplectic manifolds, which also satisfy a nondegeneracy condition, there is no surjective correspondence between smooth functions and Hamiltonian vector fields as given in Property 37.2.2. In this setting, a Hamiltonian vector field should really be viewed as a pair (X, h) of a smooth function h and a vector field X such that $X = X_h$.

37.3 Lagrangians

37.3.1 Lagrangian submanifolds

Definition 37.3.1 (Symplectic complement). Let (M, ω) be a symplectic manifold and let $S \subset M$ be an embedded submanifold $\iota : S \hookrightarrow M$. The symplectic orthogonal complement $T_p^\perp S$ (sometimes denoted by $T_p^\omega S$ to distinguish it from the Riemannian orthogonal complement) at the point $p \in S$ is defined as the subspace

$$T_p^\perp S := \{v \in T_pM \mid \omega(v, \iota_* w) = 0, \forall w \in T_pS\}\,. \tag{37.26}$$

Definition 37.3.2 (Isotropic submanifold). Let (M, ω) be a symplectic manifold. An embedded submanifold $\iota : S \hookrightarrow M$ is said to be isotropic if $\iota_* T_p S \subset T_p^\perp S$ or, equivalently, if $\omega|_S = 0$. It is said to be coisotropic if $T_p^\perp S \subset \iota_* T_p S$.

Property 37.3.3 (Characteristic distribution). Consider a symplectic manifold M and let $\iota : S \hookrightarrow M$ be a coisotropic submanifold. The orthogonal complement $T^\perp S$ defines an integrable distribution where S is foliated by isotropic leaves. More generally, if $\iota : S \hookrightarrow M$ is an embedded submanifold, the intersection $TS \cap T^\perp S$ defines an isotropic foliation of S (if the intersection is of constant rank).

Definition 37.3.4 (Lagrangian submanifold). Let (M, ω) be a symplectic manifold. An embedded submanifold $\iota : S \hookrightarrow M$ is said to be Lagrangian if $\iota_* T_p S = T_p^\perp S$. This is equivalent to S being isotropic and satisfying $\dim(S) = \frac{1}{2}\dim(M)$. Therefore, they are sometimes called maximal isotropic submanifolds.

Example 37.3.5 (Closed sections). Consider a closed one-form $\sigma \in \Omega^1(M)$. The graph of σ is a Lagrangian submanifold of T^*M. In particular, this implies that the graph of the differential $\mathrm{d}f$ of any smooth function $f \in C^\infty(M)$ is a Lagrangian submanifold of T^*M.

Property 37.3.6 (Projectable Lagrangians). Consider a Lagrangian submanifold $L \subseteq T^*M$. The following statements are equivalent:

- L is **projectable**, i.e. $\pi|_L$ is a diffeomorphism.
- $L = \mathrm{im}(\sigma)$ for some closed $\sigma \in \Omega^1(M)$.

Example 37.3.7 (Conormal bundle). Consider a submanifold $\iota : S \hookrightarrow M$ of a symplectic manifold. The conormal bundle

$$N^*S := \{(p, \alpha) \in T^*M \mid \forall v \in T_p S : \alpha(v) = 0\} \tag{37.27}$$

is a Lagrangian submanifold of T^*M.

Property 37.3.8 (Symplectomorphisms). A diffeomorphism $f : (M, \omega) \to (M', \omega')$ is a symplectomorphism if and only if the graph of f is a Lagrangian submanifold of $\overline{M} \times M'$.

Property 37.3.9 (Lagrangian subbundles). Let $E \to M$ be a symplectic vector bundle and let $L \subset E$ be a Lagrangian subbundle, i.e. a subbundle whose fibres are Lagrangian subspaces. E is symplectomorphic to $L \oplus L^*$. Moreover, E admits a $\mathrm{GL}(n)$-structure, where $2n = \dim(E)$.

Theorem 37.3.10 (Maslov–Hörmander). *Let M be a smooth manifold and consider a smooth function $W : M \times \mathbb{R}^k \to \mathbb{R}$ with $k \geq 0$. If 0 is a regular value of the map*

$$\frac{\partial W}{\partial q} : M \times \mathbb{R}^k \to \mathbb{R}^k\,, \tag{37.28}$$

the subset $\Lambda \subset T^*M$, locally defined by the equations

$$\frac{\partial W}{\partial u^i} = 0 \qquad\qquad p_i = \frac{\partial W}{\partial q^i}\,, \tag{37.29}$$

is a Lagrangian submanifold. Conversely, if $\Lambda \overset{\iota}{\hookrightarrow} T^*M$ is a Lagrangian submanifold, then at every $\lambda_0 \in \Lambda$ there exists an integer

$$k_0 \geq \dim(M) - \mathrm{rk}\big(D(\pi \circ \iota)|_{\lambda_0}\big) \tag{37.30}$$

such that, locally around λ_0, the submanifold Λ is described by some function $W : M \times \mathbb{R}^{k_0}$ satisfying the above equations.

Any function W generating a Lagrangian submanifold through the above equations will be called a **generating function**. Functions satisfying both the equations and the regularity condition are called **Morse families**.

Definition 37.3.11 (Liouville class). Consider an exact symplectic manifold (M, ω) and a Lagrangian submanifold L. Because $\omega|_L = 0$, every (global) symplectic potential θ defines a cohomology class $[\theta] \in H^1_{\mathrm{dR}}(L)$.[7] For M a cotangent bundle and θ the Liouville one-form, this is called the Liouville class of L.

In Property 37.3.6, projectable Lagrangians were characterized as the graphs of closed sections.

Definition 37.3.12 (Exact Lagrangian). Consider a Lagrangian submanifold $\iota : L \hookrightarrow T^*M$. The following statements are equivalent:

- L is **exact**, i.e. L is the graph of an exact one-form.
- The Liouville class $[\iota^*\alpha]$ vanishes, i.e. the Liouville one-form pulls back to an exact form.

37.3.2 Maslov classes

Definition 37.3.13 (Lagrangian Grassmannian). The set of isomorphism classes of Lagrangian subspaces of $\mathbb{R}^{2n}$. It is often denoted by $\mathsf{LGr}(n)$. It can be shown to be isomorphic to the homogeneous space $\mathsf{U}(n)/\mathsf{O}(n)$.

Definition 37.3.14 (Maslov class of Lagrangian subspaces). The square of the determinant function gives an isomorphism $\pi_1(\mathsf{LGr}(n)) \cong \pi_1(S^1)$, where the generator of $\pi_1(S^1)$ is identified with the path of unitary matrices $\mathrm{diag}(e^{i\pi t}, 1, 1, \ldots)$. By the Hurewicz theorem and the definition of (singular) cohomology, one obtains $H^1(\mathsf{LGr}(n)) \cong \mathbb{Z}$, which is just the degree of the induced element in $\pi_1(S^1)$. This class is called the Maslov class or **Maslov index**. (The image of the canonical generator of $\pi_1(S^1)$ is called the **universal Maslov class** μ_n.)

[7] This class depends on the choice of potential.

Example 37.3.15 ($\mathbb{R}^2$). It is not hard to show that the Lagrangian subspaces of $\mathbb{R}^2$ are exactly the lines through the origin, i.e. $\mathsf{LGr}(1) \cong \mathbb{RP}^2$. Consequently, the Lagrangians L_θ are characterized by an angle $\theta \in [0, \pi[$. The Maslov index of L_θ is given by:

$$\tau(L_\theta) = \begin{cases} 1 - \frac{2\theta}{\pi} & \text{if } \theta \neq 0\,, \\ 0 & \text{otherwise}\,. \end{cases} \tag{37.31}$$

For $\mathbb{R}^{2n}$, diagonalization of unitary matrices leads to a reduction to the case of $\mathbb{R}^2$. The Maslov index is then simply the sum of n 1-dimensional Maslov indices.

Definition 37.3.16 (Maslov class of Lagrangian submanifolds). The Maslov index (for loops) of a Lagrangian submanifold $L \subset M$ is defined analogously. Every loop in L defines a loop in $\mathsf{LGr}(n)$, since the tangent spaces are Lagrangian subspaces. The Maslov index $\mu(L) : \pi_1(L) \to \mathbb{Z}$ is defined as the Maslov index of this induced loop.

This definition can be extended to arbitrary paths of Lagrangians. For any Lagrangian $L_0 \in \mathsf{LGr}(n)$, one can stratify the Grassmannian by the sets

$$\Lambda_k := \{L \in \mathsf{LGr}(n) \mid \dim(L \cap L_0) = k\}\,. \tag{37.32}$$

The union $\Sigma(L_0) := \bigcup_{i=1}^n \Lambda_i \cong \overline{\Lambda_1}$ is called the **Maslov cycle** (with respect to L_0). It consists of all subspaces that intersect nontrivially with L_0. The Maslov index μ is the Poincaré dual of this codimension-1 cycle. It counts the number of intersections of the given path with the Maslov cycle.

Another approach passes through relative homology. Because the stratum Λ_0 is contractible, the long exact sequence in homology implies the existence of an isomorphism

$$H_1(\mathsf{LGr}(n), \Lambda_0) \cong H_1(\mathsf{LGr}(n)) \cong \mathbb{Z}\,. \tag{37.33}$$

Combined with the Hurewicz isomorphism and degree map, one again gets a Maslov index for paths with endpoints in Λ_0. It immediately follows that a path entirely in Λ_0 has vanishing Maslov index.

Remark 37.3.17. There exists a generalization (Robbin & Salamon, 1993), based on intersection theory, for paths with endpoints in nonzero strata, i.e. where the endpoints satisfy $\gamma(0) \in \Lambda_{k_0}$ and $\gamma(1) \in \Lambda_{k_1}$ for $k_0, k_1 \geq 1$. The important difference is that for these paths the Maslov index can be a half-integer:

$$\mu + \frac{k_0 - k_1}{2} \in \mathbb{Z}\,. \tag{37.34}$$

Definition 37.3.18 (Maslov index of Lagrangian subbundles). Let L, L' be Lagrangian subbundles of a trivial symplectic vector bundle $E \to M$. The Maslov class $\mu(L, L') \in H^1(M; \mathbb{Z})$ is defined as follows:

$$\mu(L, L') := (f_L^* - f_{L'}^*)\mu_n\,, \tag{37.35}$$

where $f_L, f_{L'} : M \to \mathsf{LGr}(n)$ are induced by the the trivialization of E. (It can be shown that this definition if independent of the chosen trivialization.) For nontrivial E, the Maslov class is defined by requiring that

$$\mu(\gamma^* L, \gamma^* L') = \gamma^* \mu(L, L') \tag{37.36}$$

for all loops $\gamma : S^1 \to M$. (The pullbacks along such a loop are always trivial. @@ WHY??? @@)

@@ CHECK DEFINITIONS FOR POINTS vs PATHS vs SUBBUNDLES @@

Property 37.3.19 (Homotopy invariance). The Maslov index is invariant under homotopies rel endpoints.

Property 37.3.20 (Additivity). If γ_1 and γ_2 are two paths that can be concatenated, then

$$\mu(\gamma_2 * \gamma_1) = \mu(\gamma_1) + \mu(\gamma_2)\,. \tag{37.37}$$

Alternative Definition 37.3.21 (Maslov index). The above two properties give rise to an axiomatic characterization of the Maslov index:[8]

1. **Homotopy invariance**: $\gamma_1 \sim \gamma_2 \iff \mu(\gamma_1) = \mu(\gamma_2)$,
2. **Additivity**: $\mu(\gamma_1 * \gamma_2) = \mu(\gamma_1) + \mu(\gamma_2)$,
3. **Direct sum**: $\mu(\gamma_1 \oplus \gamma_2) = \mu(\gamma_1) + \mu(\gamma_2)$, and
4. **Normalization**: $\mu(t \mapsto e^{2\pi i t}) = 1$.

Formula 37.3.22. Consider a loop $\gamma : [0,1] \to \mathsf{LGr}(n)$. An explicit formula for its Maslov index is given by

$$\mu(L) = \frac{\alpha(1) - \alpha(0)}{\pi}\,, \tag{37.38}$$

where $\exp(i\alpha(t))$ is the determinant of any unitary lift of γ.

Definition 37.3.23 (Pairwise Maslov index). One can also extend the definition of the Maslov index to a pair of paths. To this end, one should note that the diagonal map $\Delta : M \to M \times M$ gives a Lagrangian embedding. The pairwise Maslov index is then defined as follows:

$$\mu(\gamma_1, \gamma_2) := \mu(\gamma_1 \times \gamma_2; \Delta)\,, \tag{37.39}$$

where the right-hand side denotes the Maslov index with respect to the Δ-stratification of $\mathsf{LGr}(2n)$. However, this index is only well defined if the pair (γ_1, γ_2) is **admissible**, i.e. if $\gamma_1(0) \cap \gamma_2(0) = \emptyset = \gamma_1(1) \cap \gamma_2(1)$. For pairs that do not satisfy this condition,

[8] Some authors introduce additional factors of 2 in these conditions to account for nonorientablity.

one should choose a *compatible almost-complex structure* J (see Chapter 39). Then, the Maslov index is defined as follows

$$\mu(\gamma_1, \gamma_2) := \mu(\gamma_1, e^{-\theta J}\gamma_2), \tag{37.40}$$

where θ is chosen such that $(\gamma_1, e^{-\theta' J}\gamma_2)$ is admissible for all $0 < |\theta'| \leq \theta$. Homotopy invariance then implies that the definition does not depend on the precise choice of θ.

37.3.3 Polarizations

Definition 37.3.24 (Real polarization). A (real) polarization of a symplectic manifold (M, ω) is a foliation by Lagrangian submanifolds, i.e. a subbundle $P \subset TM$ such that the following conditions are satisfied:

1. **Maximality**: $\dim(TM) = 2\dim(P)$,
2. **Isotropy**: $\iota_X \omega = 0$ for all $X \in P$, and
3. **Involutivity**: $[X, Y] = 0$ for all $X, Y \in P$.

The last condition characterizes P as an integrable subbundle by Frobenius' theorem. In fact, this implies that TM is locally spanned by Hamiltonian vector fields.

More generally, one can define a (complex) polarization.

Definition 37.3.25 (Polarization). An integrable Lagrangian subbundle $\mathcal{P}$ of the complexified tangent bundle $TM^{\mathbb{C}}$ with the additional property that $\dim(\mathcal{P} \cap \overline{\mathcal{P}} \cap TM)$ is constant throughout the entire manifold.

A real polarization is (after complexifying it) the same as a complex polarization for which $\mathcal{P} = \overline{\mathcal{P}}$.

Remark 37.3.26. The constant rank condition implies that there exists a real subbundle $D \subset TM$ such that $D \otimes \mathbb{C} \cong \mathcal{P} \cap \overline{\mathcal{P}}$.

Example 37.3.27 (Vertical polarization). Consider the cotangent bundle T^*M of a smooth manifold M. Define the bundle $\mathcal{P}$ at every point $\alpha \in T^*M^{\mathbb{C}}$ as

$$\ker(\pi_*) = T_\alpha T^*_{\pi(\alpha)}M \otimes \mathbb{C} = \operatorname{span}_{\mathbb{C}}\left\{ \frac{\partial}{\partial p_i} \,\middle|\, p_i \text{ is a Darboux coordinate on } T^*M \right\},$$

where $\pi : T^*M^{\mathbb{C}} \to M$ is the (complexified) cotangent bundle projection. It can be shown that this polarization is real.

Definition 37.3.28 (Admissible polarization). Let $\mathcal{P}$ be a polarization of a manifold M. This defines two new subbundles $D := \mathcal{P} \cap \overline{\mathcal{P}} \cap TM$ and $E := (\mathcal{P} + \overline{\mathcal{P}}) \cap TM$ that are each other's symplectic complement. These are sometimes called the **isotropic** and **coisotropic** distributions, respectively. Since $\mathcal{P}$ is integrable, D is too. A polarization

is said to be (strongly) admissible[9] if E is also integrable and the leaf sets M/D and M/E are smooth manifolds. (Sometimes, the projection $M/D \to M/E$ is required to be a submersion.)

Definition 37.3.29 (Kähler polarization). A polarization $\mathcal{P}$ such that

1. $\mathcal{P} \cap \overline{\mathcal{P}} = \emptyset$, and
2. the Kähler form $\omega(\cdot, J\cdot)$ is positive definite (cf Section 39.4).

In this case, $\mathcal{P}$ is always admissible and $E = TM$. Every *Kähler manifold* (see Chapter 39) admits such a polarization, namely its *(anti)holomorphic tangent bundle*. Conversely, the existence of a Kähler polarization implies that the manifold is *(pseudo)Kähler*, where the *(anti)holomorphic tangent bundle* is given by the polarization. More generally, the *(anti)holomorphic tangent bundle* of a *complex bundle* is called the **(anti)holomorphic polarization**.

37.4 Hamiltonian dynamics

37.4.1 Dynamical systems

Definition 37.4.1 (Dynamical system). Let (M, ω) be a symplectic manifold and let $H \in C^\infty(M)$ be a distinguished 'observable'. The triple (M, ω, H) is called a dynamical system with **Hamiltonian** H. The time derivative of any function $F \in C^\infty(M)$ is defined by[10]

$$\dot{F} := \{H, F\}\,, \tag{37.41}$$

where $\{\cdot,\cdot\}$ is the Poisson bracket on M. The time evolution is completely governed by the Hamiltonian flow $\exp(tX_H)$.

Definition 37.4.2 (Nonautonomous system). A 'dynamical system' induced by a function $H \in C^\infty(M \times \mathbb{R})$, where (M, ω) is a symplectic manifold. The additional factor $\mathbb{R}$ indicates that the Hamiltonian is explicitly time dependent. In this case, the time derivative from the previous definition has to be modified:

$$\dot{f} := \{H, f\} + \partial_t f\,. \tag{37.42}$$

Definition 37.4.3 (Conserved quantity). Let (M, ω, H) be a dynamical system. A function $F \in C^\infty(M)$ is said to be conserved if it satisfies $\dot{F} = 0$. Such a function is sometimes called a **first integral**.

[9] Or strongly integrable.

[10] Note that this construction can, in fact, be generalized to Poisson manifolds.

Theorem 37.4.4 (Noether). *Every function whose Poisson bracket leaves the Hamiltonian invariant is a conserved quantity:*

$$\{H, Q\} = 0 \iff \{Q, H\} = 0\,. \tag{37.43}$$

From here on, a specific type of Hamiltonian function, called a **mechanical Hamiltonian**, is considered. Let (Q, g) be a Riemannian manifold and equip the cotangent bundle $T^*Q \xrightarrow{\pi} Q$ with its canonical symplectic structure. The Hamiltonians that will be considered are of the form (in local Darboux coordinates)

$$H(q, p) = \frac{1}{2}g(p, p) + V(q)\,, \tag{37.44}$$

where $V : Q \to \mathbb{R}$ is smooth. These Hamiltonians admit two types of symmetries (and, accordingly, conserved quantities).

Definition 37.4.5 (Kinematical symmetry). Consider a conserved quantity C. The symmetry is said to be kinematical if $\pi_*(X_C) \in \Gamma(TQ)$ exists and satisfies

$$\mathcal{L}_{\pi_*(X_C)}g = 0\,. \tag{37.45}$$

This condition says that $\pi_*(X_C)$ is a Killing vector (Definition 36.1.24).

Definition 37.4.6 (Dynamical symmetry). Any symmetry that is not a kinematical symmetry.

The following algorithm gives a way to find conditions to determine whether a given observable is conserved.

Method 37.4.7 (Van Holten algorithm). Let the conserved quantity be analytic, i.e.

$$C(q, p) = \sum_{k=0}^{N} \frac{1}{k!} a^{(n_1 \dots n_k)}(q) p_{n_1} \dots p_{n_k} \tag{37.46}$$

for some $N \in \mathbb{N}$, where the brackets around indices denote symmetrization. For a manifold where the metric g does not depend on q, one can rewrite $\{C, T + V\} = 0$ as

$$\sum_{k=1}^{N} \left[\frac{1}{(k-1)!} a^{n_1 \dots n_{k-1} i} p_{n_1} \dots p_{n_{k-1}} \frac{\partial V}{\partial q^i} - \frac{2}{k!} \frac{\partial}{\partial q^i} a^{n_1 \dots n_k} p_{n_1} \dots p_{n_k} g^{im} p_m \right] = 0\,. \tag{37.47}$$

Because two polynomials are equal if and only if their corresponding coefficients are equal, one obtains the following conditions:

- 0th order:

$$a^k \frac{\partial V}{\partial q^k} = 0\,, \tag{37.48}$$

- 1st order:

$$a^{(n_1 i)} \frac{\partial V}{\partial q^i} - 2 \frac{\partial a}{\partial q^i} g^{i n_1} = 0\,, \tag{37.49}$$

- $\cdots$, and
- N^{th} order:

$$\frac{1}{N!}a^{(n_1\ldots n_N i)}\frac{\partial V}{\partial q^i} - \frac{2}{(N-1)!}\frac{\partial}{\partial q^i}a^{(n_1\ldots n_{N-1}}g^{i)n_N} = 0\,, \tag{37.50}$$

where one should pay attention to the symmetrization brackets in the second term of the last equation. Pulling down the indices by multiplying with the metric gives

$$a_{(m_1\ldots m_N)}{}^i\partial_i V - 2N\partial_{(m_N}a_{m_1\ldots m_{N-1})} = 0\,. \tag{37.51}$$

The upper bound N in the series expansion is determined by the generalized Killing condition (36.25):

$$\partial_{(m_{N+1}}a_{m_1\ldots m_N)} = 0 \implies a_{(m_1\ldots m_{N+1})} = 0\,. \tag{37.52}$$

Remark 37.4.8. The above algorithm still holds for curved manifolds upon replacing all partial derivatives ∂_i by (Levi-Civita) covariant derivatives ∇_i.

Definition 37.4.9 (Lagrangian). Consider a smooth function $H \in C^\infty(M)$ on a symplectic manifold (M,ω). The associated Lagrangian (function) is given by the following Legendre transformation:

$$L := \iota_{X_H}\omega - f\,. \tag{37.53}$$

By Cartan's magic formula, one also has that

$$\mathcal{L}_{X_H}\theta = \mathrm{d}L\,, \tag{37.54}$$

where θ is a symplectic potential.

37.4.2 Hamilton–Jacobi equation

Definition 37.4.10 (Hamilton–Jacobi equation). Consider a smooth manifold M whose cotangent bundle comes equipped with a Hamiltonian function $H : T^*M \to \mathbb{R}$. The Hamilton–Jacobi equation (HJE) for H is the differential equation for a (smooth) function $S : M \to \mathbb{R}$ of the following form:

$$H \circ \mathrm{d}S = H\left(q, \frac{\partial S}{\partial q}\right) = 0\,. \tag{37.55}$$

Property 37.4.11. If S is a solution of the HJE, then the following properties hold:

- The differential $\mathrm{d}S$ determines a submanifold of $H^{-1}(0)$, which, by Example 37.3.5, is Lagrangian. (More generally, it determines a Lagrangian submanifold tangent to X_H.)

- The graph of $\mathrm{d}S$ is transversal to the fibres of the projection $\pi : T^*M \to M$.
- $\pi_{T^*M}|_L$ is a diffeomorphism.

Conversely, every submanifold satisfying these conditions corresponds to a solution of the HJE.

Remark 37.4.12. If the third condition is relaxed to $\pi_{T^*M}|_L$ being a local diffeomorphism, the submanifolds are only locally described by a solution to the HJE or, equivalently, they are described by a closed section $\sigma : M \to T^*M$ such that $H \circ \sigma = 0$.

Even more generally, one obtains the following notion.

Definition 37.4.13 (Geometric solution). A Lagrangian submanifold of the level set $H^{-1}(0)$ of a smooth function $H : T^*M \to \mathbb{R}$.

@@ ADD LINK WITH Maslow–Hormander AND generating functions @@

Definition 37.4.14 (Caustic). Let $L \subseteq T^*M$ be a geometric solution of the HJE. Projections of points where L is not transversal to π_{T^*M} are called caustics.

37.4.3 Integrability

Definition 37.4.15 (Integrable system). Consider a smooth vector-valued function

$$F \equiv (F_1, \dots, F_n) : M \to \mathbb{R}^n$$

on a symplectic manifold (M, ω). This map defines a **completely integrable system** (CIS) if it satisfies the following conditions:

1. The dimension is maximal, i.e. $\dim(M) = 2n$.
2. The Hamiltonian vector fields $\{X_{F_i}\}_{i \leq n}$ are almost everywhere linearly independent or, equivalently, the Jacobian DF has full rank almost everywhere.
3. For all $i, j \leq n : \{F_i, F_j\} = 0$.

The function F is sometimes called the **moment map** of the system. A Hamiltonian $H : M \to \mathbb{R}$ is said to be **completely integrable** or **Liouville integrable** if there exists a CIS consisting of first integrals of H.

Property 37.4.16. Because the Poisson brackets (Definition 37.2.1) are related to the commutator of Hamiltonian vector fields, a CIS gives rise to a maximal set of mutually commuting vector fields. Frobenius' theorem 34.6.6 then says that a CIS also gives rise to a maximally integrable distribution and, hence, an n-dimensional regular foliation.

Definition 37.4.17 (Liouville foliation). Consider a CIS $F : M \to \mathbb{R}^n$ on a symplectic manifold (M, ω). The decomposition $\bigsqcup_{v \in F(M)} F^{-1}(v)$ is called a Liouville foliation of

M. The regular leaves, i.e. those containing only regular points, densely fill M and are Lagrangian (Definition 37.3.4). Moreover, the singular leaves, which form a set of measure zero (with respect to the Liouville measure), are still isotropic.

Theorem 37.4.18 (Liouville–Arnol'd). *Consider a CIS $F : M \to \mathbb{R}^n$ on a symplectic manifold (M, ω) and let $v \in \mathbb{R}^n$ be a regular value. If the fibre over v is compact and connected, it is an embedded n-torus. Moreover, in this case, there exist*

- *open sets $V, V' \subseteq \mathbb{R}^n$ with $v \in V$ and $0 \in V'$,*
- *a symplectomorphism $\varphi : \mathbb{T}^n \times V' \to F^{-1}(V)$, and*
- *a diffeomorphism $\psi : V \to V'$ with $\psi(v) = 0$*

satisfying the following relations:

- *φ maps ω to the standard Darboux form:*

$$\varphi^* \omega = \sum_{i=1}^{n} \mathrm{d}p_i \wedge \mathrm{d}q^i \,, \tag{37.56}$$

  and

- *the composite $\psi^{-1} \circ F \circ \varphi : \mathbb{T}^n \times V' \to V'$ satisfies*

$$\psi^{-1} \circ F \circ \varphi = \pi_2 \,. \tag{37.57}$$

Remark 37.4.19. Dropping the compactness and connectedness conditions leads to a more general result:

- No compactness: fibres can be diffeomorphic to $\mathbb{T}^k \times \mathbb{R}^{n-k}$.
- No connectedness: the fibres can consist of multiple (up to countable infinity) connected components.

Definition 37.4.20 (Action-angle coordinates). Consider the composite Hamiltonian $h := F \circ \varphi : \mathbb{T}^n \times V' \to \mathbb{R}^n$ from the Liouville–Arnol'd theorem. Because of Eq. (37.57), the evolution equations become trivial:

$$\begin{cases} \dot{q}^i = \omega^i(p) \,, \\ \dot{p}_i = 0 \,. \end{cases} \tag{37.58}$$

for some '**frequency**' functions $\omega^i : \mathbb{R}^n \to \mathbb{R}^n$. The p_i's are called the **action** coordinates and the q^i's are called the **angle coordinates** (since they parameterize the tori $\mathbb{T}^n$). The only use of the further map ψ in the Liouville–Arnol'd theorem is to normalize/center the coordinates. Note that, since $\dot{p}_i = 0$, the tori are left invariant by the Hamiltonian flow.

Definition 37.4.21 (Bifurcation diagram). Consider a CIS $F : M \to \mathbb{R}^n$ and denote its critical points by $\mathrm{Crit}(F)$. The image $F\big(\mathrm{Crit}(F)\big)$ is called the bifurcation diagram of F.

Property 37.4.22 (Critical orbits). The Hamiltonian flow preserves the rank of critical points. Consequently, aside from fixed points, critical points come in families with dimension given by their rank.

The following property extends Definition 34.10.1.

Property 37.4.23 (Nondegenerate critical points). Let (M, ω) be a four-dimensional symplectic manifold equipped with a CIS $F : M \to \mathbb{R}^2$. A fixed point $p \in \mathsf{Crit}(F)$ is said to be nondegenerate if

1. **Independence**: The Hessians of F_1 and F_2 are linearly independent.
2. **Full rank**: There exist coefficients $\alpha, \beta \in \mathbb{R}$ such that

$$\alpha\Omega^{-1}\mathsf{Hess}(F_1) + \beta\Omega^{-1}\mathsf{Hess}(F_2) \tag{37.59}$$

   has four distinct eigenvalues, where Ω is a matrix representation of ω.

For rank-1 critical points $p \in M$, the situation is different. Here, there always exists a choice of coefficients $r_1, r_2 \in \mathbb{R}$ such that $r_1 \mathrm{d}F_1 + r_2 \mathrm{d}F_2 = 0$ at p. p is said to be nondegenerate if the linear combination $r_1\mathsf{Hess}(F_1) + r_2\mathsf{Hess}(F_2)$ is invertible on the quotient L^ω/L, where $L := \mathsf{span}(\{X_{F_1}, X_{F_2}\})$.

Property 37.4.24. Let (M, ω) be a symplectic manifold equipped with a CIS $F : M \to \mathbb{R}^{2n}$. Around a nondegenerate critical point $p \in M$, there exists a (local, Darboux) coordinate patch $U \equiv \{x_1, \dots, x_n, y_1, \dots, y_n\}$ and functions $f_1, \dots, f_n \in C^\infty(U)$ such that

$$\{F_i, f_j\} = 0 \tag{37.60}$$

for all $i, j \leq n$. Moreover, these functions satisfy one of the following expressions:

- **Elliptic**: $f_i(x, y) = \frac{1}{2}(x_i^2 + y_i^2)$.
- **Hyperbolic**: $f_i(x, y) = x_i y_i$.
- **Focus-focus**: $\begin{cases} f_i(x, y) & = x_i y_{i+1} - x_{i+1} y_i\,, \\ f_{i+1}(x, y) & = x_i y_i + x_{i+1} y_{i+1}\,. \end{cases}$
- **Regular**: $f_i(x, y) = y_i$.

For nondegenerate fixed points, the classification can also be performed based on the eigenvalues from the previous property:

- Elliptic: pair of imaginary eigenvalues $\pm i\beta_i$.
- Hyperbolic: pairs of real eigenvalues $\pm\alpha_i$.
- Focus-Focus: Complex eigenvalues $\pm\alpha \pm i\beta$.

Remark 37.4.25 (Focus-focus points). Assume that $n = 2$, i.e. consider a 4-manifold. Although the Arnol'd–Liouville theorem can be extended to regular-elliptic points and elliptic-eliptic points quite easily, the focus-focus case is something else. The fibres over these points are pinched tori (the pinches exactly corresponding to focus-focus points). Restricting to the case of single pinches, the Hamiltonian flow has two orbits: the pinched point and its complement.

37.5 Symplectic reduction

37.5.1 Hamiltonian actions

Alternative Definition 37.5.1 (Hamiltonian torus action). Let (M, ω) be a symplectic manifold. First, consider the case of an action of $G = \mathbb{R}$ or $G = S^1$ on M. If G acts by Hamiltonian symplectomorphisms, the action is said to be **Hamiltonian**. For $G = \mathbb{R}^n$ or $G = \mathbb{T}^n$, the action is said to be Hamiltonian if the restriction to every factor $\mathbb{R}$ or S^1 is Hamiltonian.

For general Lie groups, one needs an additional concept.

Definition 37.5.2 (Moment map). Consider a Lie group G with associated Lie algebra $\mathfrak{g}$ acting on a symplectic manifold M. The G-action on M is said to be **Hamiltonian** with moment map $\mu : M \to \mathfrak{g}^*$ if the following conditions are satisfied:

1. For every $\xi \in \mathfrak{g}$, the map

$$\mu^\xi : M \to \mathbb{R} : p \mapsto \langle \mu(p), \xi \rangle \tag{37.61}$$

   is the Hamiltonian function for the fundamental vector field 35.3.4 associated to ξ, i.e. the vector field X_ξ generated by the one-parameter group $e^{t\xi}$.

2. μ is an intertwiner between the G-action on M and the coadjoint representation (Definition 32.3.9) on $\mathfrak{g}^*$. Equivalently, the assignment of Hamiltonian vector fields $\xi \mapsto X_\xi$ is G-equivariant, i.e. $X_{g^{-1}\xi g} = g^* X_\xi$.

If only the first condition holds, the action is said to be **weakly Hamiltonian**.

Property 37.5.3. A symplectic S^1-action on a symplectic manifold for which $[\omega] \in H^2(M)$ is a positive multiple of the first Chern class (Definition 35.6.7) is Hamiltonian.

Property 37.5.4 (Obstruction). If G is compact and connected, G-equivariance of a weakly Hamiltonian action is equivalent to the assignment $\xi \mapsto H_\xi$ being a Lie algebra morphism with respect to the Poisson algebra structure on $C^\infty(M)$. This also explains the terminology 'moment map'. An action is Hamiltonian if the moment map is constant along the Hamiltonian flow, just like ordinary linear and angular momenta.

The obstruction to a weakly Hamiltonian action being Hamiltonian is given by a class in Lie algebra cohomology (Section 32.6). Let $H : \mathfrak{g} \to C^\infty(M) : \xi \mapsto H_\xi$ be a map assigning Hamiltonian functions to infinitesimal generators. The obstruction

$$\tau(\xi, \zeta) := \{H_\xi, H_\zeta\} - H_{[\xi,\zeta]} \tag{37.62}$$

satisfies the 2-cocycle condition

$$\tau\big([\xi,\zeta],\theta\big) + \tau\big([\zeta,\theta],\xi\big) + \tau\big([\theta,\xi],\zeta\big) = 0\,, \tag{37.63}$$

i.e. this obstruction determines a class $[\tau] \in H^2(\mathfrak{g})$. Furthermore, this class vanishes if and only if the action is Hamiltonian.

Definition 37.5.5 (Toric manifold). A closed, connected, symplectic $2n$-manifold that comes equipped with a faithful, Hamiltonian torus action of $\mathbb{T}^n$.

Theorem 37.5.6 (Atiyah–Guillemin–Sternberg). *Let (M, ω) be a toric manifold with moment map $\mu : M \to \mathbb{R}^n$. The image $\mu(M)$ is a convex polytope. More precisely, the fixed point set of the action is a finite disjoint union of connected, symplectic submanifolds C_i, the moment map is constant on every component and the image $\mu(M)$ is the convex hull of the points $\mu(C_i)$.*

Definition 37.5.7 (Delzant polytope). A polytope $\Delta \subset \mathbb{R}^n$ satisfying the following conditions:

1. **Simplicity**: There are n edges meeting at every vertex.
2. **Rationality**: The edges meeting at a given vertex p are of the form $p + \lambda v_i$ with $v_i \in \mathbb{Z}^n$.
3. **Smoothness**: For every vertex p, the slopes $\{v_i\}_{i\leq n}$ form a basis of $\mathbb{Z}^n$.

Theorem 37.5.8 (Delzant). *The symplectomorphism classes of toric manifolds are in bijection with the isomorphism classes of Delzant polytopes.*

Theorem 37.5.9 (Duistermaat–Heckman). *Consider a symplectic manifold (M, ω) and a Hamiltonian G-action with moment map $\mu : M \to \mathfrak{g}^*$. The pushforward of the Liouville measure along μ is a piecewise polynomial measure:*

$$\int_M (f \circ \mu)\frac{\omega^n}{n!} = \int_{\mathfrak{g}^*} fP\, d\lambda \tag{37.64}$$

for all $f \in L^1(\mathfrak{g}^*)$, where P is piecewise polynomial and λ is the Lebesgue measure on $\mathfrak{g}^*$.

Corollary 37.5.10 (Localization). Consider a circle action on a closed, symplectic n-manifold (M, ω), generated by a Morse function (Definition 34.10.2). The Fourier transform of the Liouville measure is a piecewise polynomial function:

$$\int_M e^{-\lambda f}\frac{\omega^n}{n!} = \sum_{p\in\mathsf{Crit}(f)} \frac{e^{-\lambda f(p)}}{\lambda^n w(p)}\,, \tag{37.65}$$

where $w : M \to \mathbb{Z}$ is the product of the weights obtained by considering the tangent space T_pM as an S^1-representation.

37.5.2 Symplectic reduction

Property 37.5.11 (Coisotropic reduction). If the leaf space of the characteristic distribution (Property 37.3.3) of a coisotropic submanifold $\iota : S \hookrightarrow (M, \omega)$ is itself a smooth manifold, it admits a symplectic form that pulls back to $\iota_* \omega$. This is, for example, the case when the coisotropic submanifold is **regular**, i.e. when there exists a submanifold $W \subset S$ such that $TS = TW \oplus T^\perp S$ and W intersects every leaf of S only once.

Definition 37.5.12 (Reduced vector field). Let M be a smooth manifold and G a Lie group that acts freely and properly on M. This implies that the quotient space M/G is a smooth manifold by Property 35.1.4. Now, suppose that G acts as a symmetry group on some vector field $X \in \mathfrak{X}(M)$, i.e. $g^* X = X$ for all $g \in G$. The reduced vector field $\overline{X} \in \mathfrak{X}(M/G)$ is defined through the following equation:

$$\overline{X}(\pi(m)) := \pi_* X(m) \,, \tag{37.66}$$

where $\pi : M \to M/G$ is the quotient projection.

@@ CHECK (cursus Antwerpen) @@

Theorem 37.5.13 (Marsden–Weinstein & Meyer). *Consider a Hamiltonian action of a connected Lie group G on a symplectic manifold (M, ω) with moment map $\mu : M \to \mathfrak{g}^*$. Let $M_0 := \mu^{-1}(0)$ and consider the quotient space $\overline{M} := M_0/G$. If G acts freely and properly on M_0, $\overline{M}$ is a smooth manifold and, moreover, it is symplectic with the symplectic form $\overline{\omega}$ defined by the following equation:*

$$\iota^* \omega = \pi^* \overline{\omega} \,, \tag{37.67}$$

where $\iota : M_0 \to M$ is the canonical inclusion and $\pi : M_0 \to \overline{M_0}$ the quotient projection.

Remark 37.5.14. The quotient $M /\!\!/ G := \mu^{-1}(0)/G$ is called the **Marsden–Weinstein reduction** of M by G. There exists a more general construction, where instead of the level set of 0, the inverse image of a coadjoint orbit is considered. Furthermore, if one only requires 0 to be a regular value, the reduction process still applies, but the result is only a *symplectic orbifold*.

37.5.3 Poisson reduction

Definition 37.5.15 (Poisson map). Let $(M, \{\cdot,\cdot\})$ and $(N, [\cdot,\cdot])$ be two Poisson manifolds. A Poisson map $\Phi : M \to N$ is a smooth function satisfying the following equality for all $f, g \in C^\infty(N)$:

$$\Phi^*[f, g] = \{\Phi^* f, \Phi^* g\} \,. \tag{37.68}$$

Definition 37.5.16 (Poisson action). Let G be a Lie group and let $(M, \{\cdot,\cdot\})$ be a Poisson manifold. A G-action on M is called a Poisson action or **canonical action** if every $g \in G$ acts by a Poisson map.

Theorem 37.5.17 (Poisson reduction). *Let G be a Lie group that acts freely and properly on a Poisson manifold $(M, \{\cdot,\cdot\})$. If the action is canonical, the Poisson bracket on M descends (uniquely) to a Poisson bracket on the quotient manifold M/G. Furthermore, the projection $\pi : M \to M/G$ is a Poisson map with respect to this structure.*

Property 37.5.18. Let $H : M \to \mathbb{R}$ be a G-invariant Hamiltonian function. Its Hamiltonian vector field X_H is also G-invariant and the Hamiltonian vector field of the reduced Hamiltonian $h : M/G \to \mathbb{R}$, defined by $H = h \circ \pi$, is given by the reduced vector field of X_H.

37.5.4 Lie–Poisson reduction

In the case of Lie–Poisson reductions, one considers the cotangent bundle T^*G of a Lie group G as the configuration manifold Q. It is not too hard to show that $T^*G/G \cong \mathfrak{g}^*$.

First, one can define a Poisson manifold structure on $\mathfrak{g}^*$.

Definition 37.5.19 (Lie–Poisson bracket). The Poisson bracket on $\mathfrak{g}^*$ induced by the following Poisson tensor:

$$\Pi = \frac{1}{2} c^{ij}{}_k x^k \partial_i \wedge \partial_j \,, \tag{37.69}$$

where $c^{ij}{}_k$ are the structure constants of $\mathfrak{g}$ and $x^c \in \mathfrak{g}$ is interpreted as a (linear) function on $\mathfrak{g}^*$. The induced bracket is given by

$$\{F, G\}(\mu) = \left\langle \mu, \left[\frac{\delta F}{\delta \mu}, \frac{\delta F}{\delta \mu}\right]\right\rangle . \tag{37.70}$$

Formula 37.5.20 (Lie–Poisson equations). First, assign to any vector field $X \in \Gamma(TQ)$ a linear function $\mu_X : T^*Q \to \mathbb{R}$ by the following formula:

$$\mu_X\big(\alpha|_q\big) := \alpha(X)|_q \,. \tag{37.71}$$

For these functions, one has $\{\mu_X, \mu_Y\} = -\mu_{[X,Y]}$.

Now, choose a basis $\{E^i\}_{i \leq \dim(\mathfrak{g})}$ for $\mathfrak{g}^*$. This induces a basis $\{(E^i)_L\}$ of left-invariant one-forms on G. The projection of a one-form $\alpha \in T^*G$ onto its component associated to the basis element $(E^i)_L$ defines a linear map $\mu_i : \mathfrak{g}^* \to \mathbb{R}$ by the following formula:

$$\mu_i \circ \pi : \alpha_k (E^k)_L \mapsto \alpha_i \,, \tag{37.72}$$

where $\pi : T^*G \to T^*G/G \cong \mathfrak{g}$ is the quotient map $(E^i)_L \mapsto E^i$. It can be shown that $\mu_i \circ \pi$ is exactly the linear map associated to the corresponding left-invariant vector field $(E_i)_L$.

The Lie–Poisson equations for G are the following set of equations:

$$\dot{\mu}_i = \{\mu_i, h\}_{\mathfrak{g}^*} = -C^k_{ij} \mu_k \frac{\partial h}{\partial \mu_j} \,, \tag{37.73}$$

where the Poisson bracket on $\mathfrak{g}^*$ is defined by applying the Poisson reduction theorem to T^*G.

37.5.5 Infinite-dimensional manifolds

The general definition of a symplectic manifold (M, ω) remains the same, i.e. it is a smooth manifold M equipped with a closed, nondegenerate 2-form ω. Even though the 2-plectic nature is preserved, the content of Remark 37.2.9 also applies to infinite-dimensional manifolds, i.e. Hamiltonian functions do not necessarily exist. On the space of Hamiltonian functions (and vector fields), one can define a Poisson structure as follows:

$$\{F, G\} := \omega(X^F, X^G) . \tag{37.74}$$

Recall the Lie–Poisson structure on Lie algebras from Definition 37.5.19.

Example 37.5.21 (Poisson–Vlasov bracket). Consider a Poisson manifold $(M, \{\cdot,\cdot\})$ and regard $C^\infty(M)$ as a Lie algebra under the Poisson bracket. The induced Lie–Poisson structure, where the dual of $C^\infty(M)$ is identified with the space of densities on M (Remark 34.8.11), is given as follows:

$$\{F, G\}(f) := \int_M f \left\{ \frac{\delta F}{\delta f}, \frac{\delta G}{\delta f} \right\} \mathsf{Vol}_M . \tag{37.75}$$

@@ CHECK IF CORRECT (is this globally defined? does this make sense in general?) @@

@@ COMPLETE (e.g. Palais, cursus Antwerpen) + VERY INCOHERENT RIGHT NOW @@

37.6 Metaplectic structures

Definition 37.6.1 (Metaplectic group). Consider the symplectic group $\mathsf{Sp}_{2n}(\mathbb{R})$ as defined in Definition 22.4.60. This group admits a double covering called the metaplectic group $\mathsf{Mp}_{2n}(\mathbb{R})$.

Remark 37.6.2. In contrast to the Spin-groups that are the double covers of $\mathsf{SO}(n)$ by Property 27.4.13, the metaplectic groups are not matrix groups, i.e. they do not admit a faithful finite-dimensional representation.

Definition 37.6.3 (Metaplectic structure). Consider a symplectic $2n$-manifold (M, ω). By Property 37.1.8, the frame bundle FM can be reduced to a $\mathsf{Sp}_{2n}(\mathbb{R})$-bundle $\pi_{\mathrm{Sp}} : F_{\mathrm{Sp}}M \to M$.

The smooth manifold M is said to have a metaplectic structure if there exists a principal $\mathsf{Mp}_{2n}(\mathbb{R})$-bundle $\pi_{\mathrm{Meta}} : P_{\mathrm{Meta}} \to M$ and an equivariant 2-fold lifting of F_{Sp} to P_{Meta}, i.e. a morphism $\xi : P_{\mathrm{Meta}} \to F_{\mathrm{Sp}}M$ together with a 2-fold covering map $\rho : \mathsf{Mp}_{2n}(\mathbb{R}) \to \mathsf{Sp}_{2n}(\mathbb{R})$ that satisfies:

- $\pi_{\mathrm{Sp}} \circ \xi = \pi_{\mathrm{Meta}}$, and
- $\xi(p \lhd g) = \xi(p) \cdot \rho(g)$

for all $g \in \mathsf{Mp}_{2n}(\mathbb{R})$, where $\lhd$ and $\cdot$ denote the right actions of the respective structure groups.

The following result is the symplectic analogue of Property 36.3.12.

Property 37.6.4 (Obstruction). A smooth, symplectic manifold M admits a metaplectic structure if and only if its second Stiefel–Whitney class vanishes.[11] Furthermore, the distinct metaplectic structures form an affine space over $H^1(M; \mathbb{Z}_2)$.

[11] The first Stiefel–Whitney class vanishes trivially since all symplectic manifolds are orientable.

Chapter 38

Contact Geometry

Chapter Contents

38.1 Contact structures **729**

38.1.1 Contact form 729

38.1.2 Reeb vector fields 730

38.1 Contact structures

38.1.1 Contact form

Definition 38.1.1 (Contact element). Let M be a smooth n-dimensional manifold. A contact element at the point $p \in M$, called the **contact point**, is a $(n-1)$-dimensional subspace of the tangent space T_pM.

Property 38.1.2. Because every $(n-1)$-dimensional subspace of the tangent space can be constructed as the kernel of a linear functional (an element of T_p^*M), one can construct the space of contact elements as a quotient of the cotangent bundle:

$$PT^*M = (T^*M \backslash \{0_M\})/\sim, \tag{38.1}$$

where the equivalence relation $\sim$ is defined by $\omega \sim \rho \iff \exists \lambda \in \mathbb{R}_0 : \omega = \lambda\rho$.

Definition 38.1.3 (Contact structure). Let M be a $(2n+1)$-dimensional smooth manifold. A distribution ξ of contact elements on M is called a contact structure on M if the (locally) defining one-form α satisfies the following nonintegrability condition[1]:

$$\alpha \wedge (\mathrm{d}\alpha)^n \neq 0\,. \tag{38.2}$$

[1] In fact it is maximally nonintegrable. (Compare with Frobenius' theorem 34.6.7.)

If the one-form α is defined globally on M, it is called a **contact form** and the pair (M, α) is accordingly called a **contact manifold**. The similarity to Eq. (34.75) explains why contact manifolds are sometimes said to be 'maximally nonintegrable'.

Property 38.1.4 (Coorientable distribution). A contact form α such that $\xi = \ker(\alpha)$ can be defined globally if and only if the distribution ξ is coorientable, i.e. if the line bundle TM/ξ is trivial (or orientable).

38.1.2 Reeb vector fields

Definition 38.1.5 (Reeb vector field). Let (M, α) be a contact manifold. A Reeb vector field on M is a vector field X such that $\alpha(X) = 1$ and $\iota_X \mathrm{d}\alpha = 0$.

Property 38.1.6. Given a contact manifold, there exists a unique Reeb vector field associated to it.

Chapter 39

Complex Geometry

Chapter Contents

39.1 Complex structures **731**
39.2 Complex differential forms **735**
39.3 Holomorphic vector bundles **736**
39.4 Kähler manifolds **737**
39.4.1 Killing vectors 740
39.4.2 Dirac operators 741
39.5 Complex curves **741**
39.5.1 Riemann surfaces and orbifolds 741
39.5.2 Pseudoholomorphic maps 746
39.6 Cohomology **748**
39.6.1 Dolbeault cohomology 748
39.6.2 Lagrangian Floer homology ♣ 748

39.1 Complex structures

Definition 39.1.1 (Almost complex structure). Let M be a smooth manifold. An almost complex structure on M is a (complexified) smooth $(1,1)$-tensor field $J : TM \to TM$ such that $J|_p : T_pM \to T_pM$ satisfies $J|_p^2 = -1$ for all $p \in M$. Such a structure allows to treat the tangent spaces as complex vector spaces:

$$(a + ib)X := aX + bJX\,. \tag{39.1}$$

A vector bundle equipped with such a tensor field is called a **complex vector bundle**. The underlying (real) vector bundle of a complex vector bundle E is often denoted by $E_{\mathbb{R}}$.

Property 39.1.2. An almost complex manifold is even-dimensional and orientable.

An almost complex structure induces a decomposition of the complexified tangent bundle in so-called holomorphic and antiholomorphic components:

$$TM^{\mathbb{C}} = TM^{+} \oplus TM^{-},$$

where both bundles have the same dimension (and are isomorphic as real vector bundles to TM). When the coordinates on M are denoted by $\{x^k\}_{k\leq 2n}$, bases for these two subbundles are given by

$$\left\{ \frac{\partial}{\partial z^k} := \frac{1}{2}\left(\frac{\partial}{\partial x^{2k-1}} - i\frac{\partial}{\partial x^{2k}} \right) \right\}_{k\leq n} \tag{39.2}$$

and

$$\left\{ \frac{\partial}{\partial \bar{z}^k} := \frac{1}{2}\left(\frac{\partial}{\partial x^{2k-1}} + i\frac{\partial}{\partial x^{2k}} \right) \right\}_{k\leq n}, \tag{39.3}$$

respectively.

Remark 39.1.3. The reason that the complexified tangent bundle is required has to do with the fact that J is only diagonalizable on a complex vector space (because it squares to a negative value).

Example 39.1.4 (Complex vector spaces). Consider a complex vector space V. By looking at Property 23.3.24 and using the canonical isomorphism $V \cong T_v V$ for vector spaces, one can see that the automorphism $v \mapsto iv$ induced by the imaginary unit gives rise to an almost complex structure on V.

Property 39.1.5 (Reduction of structure group). A $2n$-dimensional manifold M admits an almost complex structure if and only if the structure group of the tangent bundle TM can be reduced from $\mathsf{GL}_{2n}(\mathbb{R})$ to $\mathsf{GL}_n(\mathbb{C})$.

Moreover, the set of of almost complex structures on $\mathbb{C}^n$ is given by the homogeneous space $\mathsf{GL}_{2n}(\mathbb{R})/\mathsf{GL}_n(\mathbb{C})$. By globalizing this, one obtains that the set of almost complex structures on a complex vector bundle (E, J) is given by $\mathsf{Aut}(E_{\mathbb{R}})/\mathsf{Aut}(E, J)$.

Definition 39.1.6 (Complex dimension). The integer $n \in \mathbb{N}$ in this property is called the complex dimension of M. It is denoted by $\mathsf{dim}_{\mathbb{C}}(M)$.

Definition 39.1.7 (Complex manifold). A topological space M for which there exists an open cover $\{U_i\}_i$ such that for every U_i there exists a homeomorphism $\varphi_i : U_i \to \mathbb{C}^n$ onto some open subset of $\mathbb{C}^n$. The transition functions $\varphi_{ji} : \varphi_i(U_i \cap U_j) \to \varphi_j(U_i \cap U_j)$ are also required to be holomorphic.

Property 39.1.8. An almost complex manifold is complex if and only if the $\mathsf{GL}_n(\mathbb{C})$-structure is integrable.

The integrability condition can be rephrased algebraically as follows.

Theorem 39.1.9 (Newlander–Nirenberg). *An almost complex manifold is complex if and only if the* ***Nijenhuis tensor*** N_J *vanishes:*

$$N_J(X,Y) := [JX,JY] - J[JX,Y] - J[X,JY] - [X,Y] = 0 \tag{39.4}$$

for all vector fields $X, Y \in \mathfrak{X}(M)$*. Locally, this can be written as*

$$J^\nu_\rho \partial_\nu J^\mu_\sigma - J^\nu_\sigma \partial_\nu J^\mu_\rho - J^\mu_\nu \partial_\rho J^\nu_\sigma + J^\mu_\nu \partial_\sigma J^\mu_\rho = 0\,. \tag{39.5}$$

Property 39.1.10. The Nijenhuis tensor is, equivalently, the Frölicher–Nijenhuis bracket (Definition 34.4.32) of J with itself:

$$N_J = [J,J]_{\mathrm{FN}}\,. \tag{39.6}$$

Theorem 39.1.11 (Oka coherence theorem). *The structure sheaf of a complex manifold, i.e. its the sheaf of holomorphic functions, is coherent (Definition 8.6.3).*

Property 39.1.12 (Characteristic classes). Let E be a complex vector bundle.

$$c_i(E) \bmod 2 = w_{2i}(E_{\mathbb{R}}) \tag{39.7}$$

for all $i \leq \dim_{\mathbb{C}}(E)$.

Definition 39.1.13 (Stein manifold). A complex manifold M satisfying the following two conditions, where $\mathcal{O}(M)$ denotes the ring of holomorphic functions on M:

1. **Holomorphic convexity**: For every compact subset $K \subset M$, the **holomorphic convex hull**

$$\overline{K} := \left\{ z \in M \mid \forall f \in \mathcal{O}(M) : |f(z)| \leq \sup_{z' \in K} |f(z')| \right\} \tag{39.8}$$

   is compact.

2. **Holomorphic separability**: For every two distinct points $z, z' \in M$, there exist a function $f \in \mathcal{O}(M)$ such that $f(z) \neq f(z')$.

Alternative Definition 39.1.14. A Stein manifold of (complex) dimension $n \in \mathbb{N}$ is a complex manifold admitting a proper holomorphic immersion into $\mathbb{C}^n$.

Property 39.1.15. Stein manifold are noncompact.

The following type of cover resembles that of Definition 31.1.16 and Property 31.1.17.

Property 39.1.16 (Stein cover). Every complex manifold admits a cover by Stein manifolds such that any finite intersection of covering sets is again Stein.

Property 39.1.17 (Oka–Grauert principle). The classification of holomorphic and topological vector bundles over a Stein manifold coincide or, more generally, all cohomological problems on Stein manifolds only have topological obstructions.

Remark 39.1.18 (Oka manifolds). This principle has a more general homotopical/model-theoretic (Chapter 13) formulation. There exists a class of complex manifolds, the Oka manifolds, such that the inclusion

$$\mathsf{Map}_{\mathrm{hol}}(A, S) \hookrightarrow \mathsf{Map}(A, S) \tag{39.9}$$

of holomorphic functions into continuous functions is a weak homotopy equivalence for all Oka manifolds A and Stein manifolds S.

Theorem 39.1.19 (Cartan's theorem A). *Every coherent sheaf (Definition 8.6.3) on a Stein manifold is spanned by its global sections.*

Theorem 39.1.20 (Cartan's theorem B). *The sheaf cohomology of any coherent sheaf over a Stein manifold vanishes in positive degree:*

$$H^{\geq 1}(M; \mathcal{F}) = 0 \tag{39.10}$$

for all $\mathcal{F} \in \mathsf{Coh}(M)$.

Definition 39.1.21 (Metalinear structure). Consider the determinant morphism

$$\det : \mathsf{GL}_n(\mathbb{C}) \to \mathbb{C}^\times .$$

The metalinear group can be considered as the domain of the holomorphic square root of det:

$$\mathsf{ML}_n(\mathbb{C}) := \left\{(A, z) \in \mathsf{GL}_n(\mathbb{C}) \times \mathbb{C}^\times \mid \det(A) = z^2\right\} . \tag{39.11}$$

An equivalent definition, which will be used in the remainder of the text, makes use of the special linear group:

$$\mathsf{ML}_n(\mathbb{C}) = \frac{\mathsf{SL}_n(\mathbb{C}) \times \mathbb{C}}{2\mathbb{Z}}, \tag{39.12}$$

where $\mathbb{Z}$ acts on the product group as $k : (A, z) \mapsto (e^{-2\pi i k/n} A, z + 2\pi i k/n)$. This group is the double cover of $\mathsf{GL}_n(\mathbb{C})$.

Similar to the definition of spinor and metaplectic structures (Definition 36.3.3 and Definition 37.6.3), one can also define metalinear structures on a manifold. The metalinear frame bundle is a lift of the (complex) frame bundle along the canonical morphism $\mathsf{ML}_n(\mathbb{C}) \to \mathsf{GL}_n(\mathbb{C})$ such that it 'commutes' with the bundle map $F_{\mathrm{ML}}M \to FM$.

Property 39.1.22 (Existence). A smooth manifold M admits a metalinear structure if and only if its first Stiefel–Whitney class $w_1 \in H^1(M; \mathbb{Z}_2)$ squares to 0. In particular, every orientable manifold admits a metalinear structure. The set of nonequivalent metalinear structures is parametrized by $H^1(M; \mathbb{Z}_2)$.

Remark 39.1.23. The above definitions can be restricted to real manifolds and real metalinear structures.

Corollary 39.1.24. A smooth manifold M admits a metalinear structure if and only if its canonical line bundle admits a square root.

This result can be generalized as follows.

Definition 39.1.25 (Half-form). Consider a smooth manifold M equipped with a metalinear frame bundle $F_{\mathrm{ML}}M$. The bundle of half-forms $\Omega^{1/2}(M)$ is defined as the associated $\mathbb{C}$-line bundle constructed from the action $(g,\lambda) \mapsto z\lambda$, where $g \equiv (A,z) \in \mathsf{ML}_n(\mathbb{C})$.

This can be seen as a (holomorphic) square root of the determinant line bundle. Consider the bundle of 1-densities $|\Omega^1|(M)$ from Remark 34.8.11. There exists a map $\Omega^{1/2}(M)\otimes \Omega^{1/2}(M) \to |\Omega|^1(M)$ defined by sending the pair (μ,ν) to the (tensor) product $\mu\otimes\overline{\nu}$ along the covering map $F_{\mathrm{ML}}M \to FM$. If one does not use the conjugation, a section of the determinant line bundle $\Omega^n(M)$ is obtained.

Property 39.1.26 (Metaplectic structure). Let (M,ω) be a symplectic manifold and consider a Lagrangian subbundle $L \subset TM$. The tangent bundle TM admits a metaplectic structure if and only if L admits a metalinear structure.

39.2 Complex differential forms

Property 39.2.1. On a complex manifold, there exist local coordinates $\{z^\mu\}_{\mu\leq n}$ such that the almost complex structure J can be written as (when extended to $TM^{\mathbb{C}}$)

$$J = i\partial_\mu \otimes \mathrm{d}z^\mu - i\partial_{\overline{\mu}} \otimes \mathrm{d}\overline{z}^\mu \,. \tag{39.13}$$

This coordinate expression can be used to find a coordinate transformation from the real coordinates $\{x^\mu\}_{\mu\leq 2n}$ to the complex coordinates $\{z^\mu,\overline{z}^\mu\}_{\mu\leq n}$.

Remark 39.2.2. Note that, on the complexified tangent bundle, there exist two kinds of imaginary units: i and J. The differential forms $\mathrm{d}z$ and $\mathrm{d}\overline{z}$ are both linear with respect to the scalar i, but only $\mathrm{d}z$ is linear with respect to J, i.e. $\mathrm{d}z \circ J = \mathrm{d}z$ and $\mathrm{d}\overline{z} \circ J = -\mathrm{d}\overline{z}$.

Using the basis forms $\mathrm{d}z^\mu, \mathrm{d}\overline{z}^\mu$ one can also define complex Grassmann algebras $\Omega^{p,q}(M)$, analoguous to $\Omega^k(X)$ for smooth manifolds:

$$\begin{aligned}
\Omega^{1,0}(M) &:= \mathsf{span}_{C^\infty(M,\mathbb{C})}\{\mathrm{d}z^\mu\} \\
\Omega^{0,1}(M) &:= \mathsf{span}_{C^\infty(M,\mathbb{C})}\{\mathrm{d}\overline{z}^\mu\} \\
\Omega^{p,q}(M) &:= \left(\bigwedge_{i=1}^{p} \Omega^{1,0}(M)\right) \wedge \left(\bigwedge_{j=1}^{q} \Omega^{0,1}(M)\right).
\end{aligned} \tag{39.14}$$

Property 39.2.3. The spaces $\Omega^{1,0}(M)$ and $\Omega^{0,1}(M)$ are stable, i.e. they transform tensorially, under holomorphic coordinate transformations. On the space

$$\Omega^k(M) = \bigoplus_{p+q=k} \Omega^{p,q}(M) \tag{39.15}$$

of forms of total degree $k \in \mathbb{N}$, one can then define the canonical projection maps $\pi^{p,q} : \Omega^k \to \Omega^{p,q}$.

Definition 39.2.4 (Dolbeault operator). Consider a general $(p+q)$-form $\omega \in \Omega^{p,q}(M)$. The de Rham differential maps this form to a $(p+q+1)$-form. This form is in general an element of $\sum_{r+s=p+q+1} \Omega^{r,s}(M)$. Using the projection maps $\pi^{p,q}$ one can define two additional differential operators:

$$\begin{aligned} \partial &:= \pi^{p+1,q} \circ \mathsf{d}\,, \\ \overline{\partial} &:= \pi^{p,q+1} \circ \mathsf{d}\,. \end{aligned} \tag{39.16}$$

The latter is called the Dolbeault operator.

Definition 39.2.5 (Holomorphic form). A differential form $\omega \in \Omega^k(M)$ satisfying

$$\overline{\partial}\omega = 0\,. \tag{39.17}$$

This is in analogy with the classical Cauchy–Riemann conditions (17.9).

Property 39.2.6 (Holomorphic decomposition). By explicitly writing out the action of the de Rham differential d on a general (p,q)-form, one obtains the following decomposition:

$$\mathsf{d} = \partial + \overline{\partial}\,. \tag{39.18}$$

Note that, for an almost complex manifold, this relation does not hold in general. An almost complex manifold is integrable if and only if this expression holds. By using the coboundary property of d, one also obtains

$$\begin{aligned} \partial^2 = \overline{\partial}^2 &= 0\,, \\ \partial\overline{\partial} + \overline{\partial}\partial &= 0\,. \end{aligned} \tag{39.19}$$

Remark 39.2.7 (Integrability). It can be shown that J is integrable, i.e. the almost complex structure is complex, if and only if the induced Dolbeault operator $\overline{\partial}$ squares to zero.

39.3 Holomorphic vector bundles

More generally, a complex vector bundle E is said to be **holomorphic** if it admits a trivialization by holomorphic transition functions or, equivalently[1], if it admits a Dolbeault operator $\overline{\partial} : \Omega^{\bullet,\bullet}(M;E) \to \Omega^{\bullet,\bullet+1}(M;E)$ that squares to zero. Note that, in contrast to the case of TM, a holomorphic vector bundle only admits the natural definition of a $\overline{\partial}$-operator. To have a ∂-operator, one should consider antiholomorphic vector bundles.

[1] Another equivalent definition is a complex vector bundle for which the projection is holomorphic.

Theorem 39.3.1 (Koszul–Malgrange). *Holomorphic vector bundles $E \to M$ are equivalent to complex vector equipped with a (unique) connection ∇ such that $\nabla^{0,1} = \overline{\partial}$.*

Property 39.3.2 (Holomorphic sections). Let $E \to M$ be a holomorphic vector bundle over a compact manifold M. The subspace $\Gamma_{\text{hol}}(E) \subset \Gamma(E)$ of holomorphic sections is finite dimensional.

Definition 39.3.3 (Positive line bundle). A holomorphic line bundle whose curvature form is, after multiplication by i, a positive-definite form.

Formula 39.3.4. Analogous to the definition of the codifferential in Definition 36.1.4, one can define the adjoints of the Dolbeault operators:

$$\begin{aligned} \partial^\dagger &:= - * \overline{\partial} *, \\ \overline{\partial}^\dagger &:= - * \partial *, \end{aligned} \tag{39.20}$$

where the fact that the real dimension of a complex manifold is even is used: $(-1)^{n(k+1)+1} = -1$.

Formula 39.3.5 (Kähler identities). Consider the **Lefschetz operator**

$$L : \Omega^{p,q}(M) \to \Omega^{p+1,q+1}(M) : \alpha \mapsto \omega \wedge \alpha \tag{39.21}$$

and its adjoint, sometimes called the **contraction operator**,

$$L^\dagger : \Omega^{p,q}(M) \to \Omega^{p-1,q-1}(M) : \alpha \mapsto *L * \alpha\,. \tag{39.22}$$

- $[\partial, L] = [\overline{\partial}, L] = 0$
- $[\partial^\dagger, L] = -i\overline{\partial}$,
- $[\overline{\partial}^\dagger, L] = i\partial$,
- $[\partial^\dagger, L^\dagger] = [\overline{\partial}^\dagger, L^\dagger] = 0$,
- $[\partial, L^\dagger] = -i\overline{\partial}^\dagger$, and
- $[\overline{\partial}, L^\dagger] = i\partial^\dagger$.

Using the adjoint Dolbeault operators, one can also write the Hodge Laplacian (Definition 36.1.5) as

$$\Delta = 2(\partial\partial^\dagger + \partial^\dagger\partial) = 2(\overline{\partial}\overline{\partial}^\dagger + \overline{\partial}^\dagger\overline{\partial})\,. \tag{39.23}$$

The Hodge Laplacian commutes with all of the operators above.

39.4 Kähler manifolds

In analogy with the definition of Riemannian manifolds (Definition 36.1.2), one can also define metrics for complex vector bundles.

Definition 39.4.1 (Hermitian manifold). A complex vector bundle equipped with a Hermitian bundle metric. A connection that is compatible with this metric is called a Hermitian connection.

Definition 39.4.2 (Kähler manifold). Consider a smooth manifold M equipped with a Riemannian structure g, a symplectic structure ω and an almost complex structure J. This manifold is called a Kähler manifold if the structures satisfy any of the following equivalent sets of compatibility conditions:

1. The almost complex structure J is integrable[2], and
2. The symplectic form is **compatible** with the almost complex structure:

$$\omega(v, w) = \omega(Jv, Jw) \tag{39.24}$$

and

$$\omega(v, Jv) > 0\,; \tag{39.25}$$

or

1. M is Hermitian with metric $h(v, w) := g(v, w) + ig(v, Jw)$, and
2. The two-form $\omega(v, w) := g(v, Jw)$ is closed and, hence, symplectic[3];

or

1. M is Hermitian with metric $h(v, w) := g(v, w) + ig(v, Jw)$, and
2. J is parallel with respect to the Levi-Civita connection on (M, g):

$$\nabla_X J = 0\,. \tag{39.26}$$

Remark 39.4.3. The property that says that J acts isometrically can be interchanged for the statement that J acts as a symplectomorphism. These two statements are equivalent for a Kähler manifold.

Property 39.4.4 (Tame structures). The compatibility conditions between a symplectic form ω and an almost complex structure J can be weakened to only be $\omega(v, Jv) > 0$. In this case, J is said to be **ω-tame**.

The set of all almost complex structures that are tamed by a given symplectic form (on a finite-rank vector bundle) is nonempty and contractible. This also holds for the stronger compatibility condition. The converse also holds, i.e. given an almost complex structure, the set of all symplectic forms that tame it (or that are compatible with it) is nonempty and convex (hence, also contractible).

[2] If not, the manifold is said to be almost Kähler.

[3] The nondegeneracy condition is automatically satisfied because of the nondegeneracy of the metric.

Remark 39.4.5. Note that every ω-tame almost complex structure J also induces a Riemannian metric after symmetrization. When the tame structure is also compatible, this symmetrized metric coincides with $\omega(\cdot, J\cdot)$.

Definition 39.4.6 (Kähler form). When any of the equivalent sets of conditions in Definition 39.4.2 is satisified, the central object is the Kähler form or **fundamental form**:

$$\omega(v, w) := g(v, Jw)\,. \tag{39.27}$$

Because it is closed, it determines a cohomology class $[\omega] \in H^2_{\mathrm{dR}}(M;\mathbb{R})$. This class is called the **Kähler class** of M.

Formula 39.4.7. The metric $g \equiv g_{\mu\nu}\mathrm{d}x^\mu \otimes \mathrm{d}x^\nu$ can be rewritten as

$$g = g_{\mu\bar{\nu}}\left(\mathrm{d}z^\mu \otimes \mathrm{d}\bar{z}^\nu + \mathrm{d}\bar{z}^\nu \otimes \mathrm{d}z^\mu\right)\,. \tag{39.28}$$

The Kähler form can then be written as

$$\omega = i g_{\mu\bar{\nu}}\mathrm{d}z^\mu \wedge \mathrm{d}\bar{z}^\nu\,. \tag{39.29}$$

Definition 39.4.8 (Kähler potential). Using the $\partial\bar{\partial}$-*lemma* (see 39.6.4), one can locally write the Kähler form as

$$\omega = i\partial\bar{\partial}K(z, \bar{z})\,, \tag{39.30}$$

where the function $K : M \to \mathbb{R}$ is called the **Kähler potential**. Equation (39.29) implies that one can locally rewrite the metric as

$$g_{\mu\bar{\nu}} = \partial_\mu \partial_{\bar{\nu}} K(z, \bar{z})\,. \tag{39.31}$$

Property 39.4.9. The Christoffel symbols associated to the Levi-Civita connection on (M, g) admit a simple expression when M is Kähler: only the ${\Gamma_{\mu\nu}}^\lambda$ and ${\Gamma_{\bar{\mu}\bar{\nu}}}^{\bar{\lambda}}$ components do not vanish. They are given by

$$\begin{aligned} {\Gamma_{\mu\nu}}^\lambda &= g^{\lambda\bar{\rho}}\partial_\mu g_{\nu\bar{\rho}}\,, \\ {\Gamma_{\bar{\mu}\bar{\nu}}}^{\bar{\lambda}} &= g^{\bar{\lambda}\rho}\partial_{\bar{\mu}} g_{\bar{\nu}\rho}\,. \end{aligned} \tag{39.32}$$

Accordingly, the only nonvanishing component of the Riemann curvature tensor is

$$R_{\bar{\mu}\nu\bar{\lambda}\rho} = g_{\bar{\lambda}\kappa}\partial_{\bar{\mu}}{\Gamma_{\nu\rho}}^\kappa\,. \tag{39.33}$$

Definition 39.4.10 (Kähler transformation). From Definition 39.4.8, one can conclude that the Kähler potential is not unambiguously defined. The following transformation leaves the Kähler form invariant:

$$K'(z, \bar{z}) = K(z, \bar{z}) + f(z) + \overline{f(\bar{z})}\,. \tag{39.34}$$

On overlapping coordinate charts, the transformation between Kähler potentials is exactly of this form.

Definition 39.4.11 (Calabi–Yau manifold). A Kähler manifold with trivial canonical bundle $\Omega^{n,0}(M)$. Equivalently, a $2n$-dimensional Riemannian manifold with special holonomy group in $\mathsf{SU}(n)$. This implies, in particular, that the manifold is Ricci flat.[4]

Conjecture 39.4.12 (Calabi–Yau). For a compact Calabi–Yau manifold, the first Chern class vanishes.

Definition 39.4.13 (Hypercomplex manifold). A smooth manifold equipped with three distinct complex structures $I, J, K : TM^{\mathbb{C}} \to TM^{\mathbb{C}}$ that satisfy the quaternion algebra relation:

$$I \circ J = K\,. \tag{39.35}$$

Definition 39.4.14 (Hyperkähler manifold). A hypercomplex manifold that admits a (Riemannian) metric that is Kähler with respect to all complex structures.

39.4.1 Killing vectors

Definition 39.4.15 (Holomorphic Killing vector). Consider the set of Killing vector fields associated to the metric g. Within this set of vector fields, one can consider those k_A that satisfy

$$\mathcal{L}_{k_A} J = 0\,. \tag{39.36}$$

or, equivalently by the Kähler condition,

$$\mathcal{L}_{k_A} \omega = 0\,. \tag{39.37}$$

These are called holomorphic Killing vector fields because their components are locally holomorphic. This can easily be shown by writing the Killing condition in terms of covariant derivatives and by using Eq. (39.13).

Definition 39.4.16 (Moment map). Let $k \in \mathfrak{X}(M)$ be a holomorphic Killing vector field. From $\mathrm{d}\omega = 0$, one can, using Cartan's magic formula 34.4.18 and the above condition, derive that $\iota_k \omega$ is closed. Poincaré's lemma further implies that there exists a real function $\mathcal{P}(z, \bar{z})$ such that

$$\iota_k \omega = \mathrm{d}\mathcal{P}\,. \tag{39.38}$$

Using Eq. (39.29), one can then find the following expression for the Killing vector fields:

$$k^\mu = -i g^{\mu\bar{\nu}} \partial_{\bar{\nu}} \mathcal{P}\,. \tag{39.39}$$

[4]The original definition by *Yau* was that of a compact, Ricci-flat Kähler manifold with vanishing first Chern class.

39.4.2 Dirac operators

Property 39.4.17. Let M be a compact Kähler (or Hermitian) manifold. M admits a Spin-structure if and only if there exists a square root of its (holomorphic) canonical line bundle by Property 39.1.12 or, equivalently, if M admits a metaplectic structure.

Property 39.4.18 (Spin$^{\mathbb{C}}$-structures). By Property 36.3.14 and Property 39.1.12, every almost complex manifold admits a Spin$^{\mathbb{C}}$-structure. If there exists a line bundle L, such that $c_1(L) = w_2(M) \bmod 2$, then M admits a Spin$^{\mathbb{C}}$-structure with L as its determinant line bundle.

Definition 39.4.19 (Kähler–Dirac operator). Consider a Riemannian manifold (M, g). The Kähler–Dirac operator is a square root of the Hodge Laplacian (Definition 36.1.5):

$$\mathsf{D} := \mathsf{d} + \delta\,. \tag{39.40}$$

Definition 39.4.20 (Dolbeault–Dirac operator). Let M be a Kähler manifold equipped with a Spin-structure. By the first property above, this implies that the canonical line bundle admits a square root $\sqrt{\Omega^{n,0}}$. It can be shown that the spin bundle on M satisfies

$$S \cong \Omega^{0,\bullet}(M) \otimes \sqrt{\Omega^{n,0}} \tag{39.41}$$

and that the Dirac operator can be identified with

$$\overline{\partial} + \overline{\partial}^{\dagger}\,, \tag{39.42}$$

where $\overline{\partial}$ is the Dolbeault operator (Definition 39.2.4). For the action on the θ-characteristic, choose a connection ∇ on the canonical line bundle:

$$\begin{aligned} \overline{\partial}(\omega \otimes \psi) &:= (\overline{\partial}\omega) \otimes \psi + \sum_{i=1}^{2n} \pi^{0,k+1}(\mathrm{d}x^i \wedge \omega) \otimes \nabla_i \psi \\ \overline{\partial}^{\dagger}(\omega \otimes \psi) &:= (\overline{\partial}^{\dagger}\omega) \otimes \psi - \sum_{i=1}^{2n} (\partial_i \lrcorner\, \omega) \otimes \nabla_i \psi\,, \end{aligned} \tag{39.43}$$

where $\omega \in \Omega^{0,k}(M)$.

Remark 39.4.21. More generally, for Spin$^{\mathbb{C}}$-structures on complex manifolds, the associated spinor bundle satisfies

$$S \cong \Omega^{0,\bullet}(M)\,. \tag{39.44}$$

The associated Dirac operator is given by the Dirac–Dolbeault operator on M: $\not{D} = \overline{\partial} + \overline{\partial}^{\dagger}$.

39.5 Complex curves

39.5.1 Riemann surfaces and orbifolds

Definition 39.5.1 (Riemann surface). A complex manifold of (complex) dimension one. It is often assumed to be connected.

Property 39.5.2 (Genus classification). Compact Riemann surfaces are, topologically, characterized by their genus.

Example 39.5.3 (Riemann sphere). The sphere S^2 admits the structure of a Riemann surface: $\mathbb{CP}^1$. The automorphism group of $\mathbb{CP}^1$ is the modular group $\mathsf{PSL}_2(\mathbb{C})$ from Definition 11.2.1. It can be proven that, given any two triples of points, there exists a unique Möbius transformation mapping them onto each other.

Theorem 39.5.4 (Uniformization theorem). *Every simply connected Riemann surface is biholomorphically equivalent to either the Riemann sphere, the complex plane or the upper half plane.*

Corollary 39.5.5. Because the universal cover of a Riemann surface is again a Riemann surface, the uniformization theorem implies that every Riemann surface can be obtained as the quotient of $\mathbb{CP}^1, \mathbb{C}$ or $\mathcal{H}$ by a freely-acting discrete group.

Moreover, $\mathbb{CP}^1$ only covers itself and $\mathbb{C}$ only admits $\mathbb{Z}$ or a discrete lattice as freely-acting discrete automorphism groups (leading to $\mathbb{C}$, $\mathbb{R} \times S^1 \cong \mathbb{C}\backslash\{0\}$ and $\mathbb{T}^2$ as quotients). All other Riemann surfaces are obtained from the halfplane $\mathcal{H}$.

Definition 39.5.6 (Stable Riemann surface). A Riemann surface Σ of genus $g \in \mathbb{N}$ with $n \in \mathbb{N}$ marked points is said to be stable if its punctured Euler characteristic

$$\chi(\Sigma\backslash\{z_1, \dots, z_n\}) = 2 - 2g - n \tag{39.45}$$

is negative.

Property 39.5.7 (Automorphism group). For a stable Riemann surface (Σ, J), every element of $\mathsf{Aut}_n(\Sigma, J)$ that is not the identity is also not homotopic to the identity. In particular, (Σ, J) has a finite automorphism group. For nonstable Riemann surfaces, the automorphism group is always a smooth Lie group.

Construction 39.5.8 (Moduli space). The **Teichmüller space** $\mathcal{T}_{g,n}$ is the set of equivalence classes of Riemann surfaces of genus $g \in \mathbb{N}$ with $n \in \mathbb{N}$ marked points, where two surfaces X, Y are said to be equivalent if there exists a holomorphic diffeomorphism, i.e. a biholomorphism, $X \cong Y$ that is isotopic to the identity and that preserves the marked points:

$$\mathcal{T}_n(\Sigma) := \mathcal{J}(\Sigma)/\mathsf{Diff}^{\mathrm{hol}}_{n,0}(\Sigma)\,, \tag{39.46}$$

where Σ is an oriented, closed surface. The full moduli space $\mathcal{M}_{g,n}$ can then be obtained by further quotienting out the action of the mapping class group (Definition 7.1.10).

Remark 39.5.9 (Markings). A confusing point in Teichmüller theory (and the general study of Riemann surfaces) is that a point in the Teichmüller space, i.e. an equivalence class of complex structures with biholomorphisms, is also called a marking or **marked Riemann surface**.

Property 39.5.10 (Alternative characterization). Markings can also be described differently. Consider a surface Σ of genus $g \in \mathbb{N}$. A point in $\mathcal{T}_n(\Sigma)$ is an equivalence class of pairs (X, φ), where $\varphi : \Sigma \to X$ is a diffeomorphism, where two pairs $(X, \varphi), (X', \varphi')$ are identified if $\varphi' \circ \varphi^{-1}$ is isotopic to a biholomorphism.

Even when restricting to stable surfaces, the moduli space does not have the structure of a smooth manifold due to the existence of the marked points. To accurately describe the geometrical structure one needs to generalize the notion of a manifold (cf. Chapter 31).

Definition 39.5.11 (Orbifold). Let X be a topological space. An orbifold chart on X is a tuple $(U, G, \varphi : U \to V/G)$ such that $U \subset M$ is open, $V \subset \mathbb{R}^n$ is open and connected, G is a finite group and φ is a homeomorphism.

A **subchart** $(U', G', \varphi' : U' \to V'/G')$ of $(U, G, \varphi : U \to V/G)$ is a triple such that there exist inclusions $U' \subset U$, $V' \subset V$ and a morphism $G' \to G$ that all commute in the obvious way with the additional property that the stabilizer of every point in U' is preserved. This property implies that the stabilizer of any point in X can be uniquely defined as the stabilizer of any of its preimages.

Two orbifold charts are said to be **compatible** if their intersection is contained in a subchart of both charts. An **orbifold atlas** is defined as a cover of X by compatible orbifold charts.

Definition 39.5.12 (Morphism of orbifolds). A general definition of orbifold morphisms is quite technical. Here, a specific situation is considered, that where the fibres over the orbifold are manifolds themselves.

A morphism of orbifolds 'with manifold fibres' is a continuous function $f : X \to Y$ with, for every point $y \in Y$, a choice of orbifold chart $\varphi_y : U_y \to V_y/G$ (containing y), a smooth intertwiner $F : V_x \to V_y$ and an isomorphism of V_x/G with a suborbifold of X such that the following equation is satisfied:

$$\varphi_y^{-1} \circ F = f \circ \varphi_x^{-1} . \tag{39.47}$$

Most notions of differential geometry carry over to the orbifold setting quite naturally. For example, a differential form on a chart $(U, G, \varphi : U \to V/G)$ is defined as a G-invariant differential form on V and integration is defined by averaging over the preimage of a chain:

$$\int_C \omega := \frac{1}{|G|} \int_{\varphi^{-1}(C)} \omega_U , \tag{39.48}$$

where ω_U is the orbifold representative of ω for the chart $(U, G, \varphi : U \to V/G)$ containing C. A vector bundle on an orbifold consists of an ordinary vector bundle with a fibrewise lift of the G-action.

Definition 39.5.13 (Euler characteristic). Let X be a G-orbifold. The Euler characteristic of X is defined as the average of the Euler characteristics of its fixed-point spaces:

$$\chi(X) := \frac{1}{|G|} \sum_{g \in G} \chi(X^g) \,. \tag{39.49}$$

Now that orbifolds have been introduced, the structure of $\mathcal{M}_{g,n}$ can be analyzed. When $2 - 2g - n < 0$, the moduli space has the structure of a complex orbifold of (complex) dimension $3g - 3 + n$ and the stabilizer at a point is equal to the automorphism group of a representative of that equivalence class. To endow the moduli space with such a structure, one can pull back a 'universal curve'.

Definition 39.5.14 (Family of Riemann surfaces). A family of Riemann surfaces of genus $g \in \mathbb{N}$ with $n \in \mathbb{N}$ marked points is a function $p : C \to U$ that admits n disjoint sections $s_i : U \to C$ and for which the fibre over every point is a Riemann surface. The intersections of the fibres with the sections are exactly the marked points.

@@ CHECK THIS DEFINITION @@

Given two such families p, p' and a subset $V \subset U'$, the restriction $p'|_V$ is called a **pull-back** of p if there exists a function $\varphi : V \to U$ such that $C'|_V \cong \varphi^* C$.

Property 39.5.15 (Universal curve over $\mathcal{M}_{g,n}$). Let C be Riemann surface of genus $g \in \mathbb{N}$ with $n \in \mathbb{N}$ marked points (with negative Euler characteristic). Denote its (finite) automorphism group by G. There exist

- an open, bounded, connected set $U \subset \mathbb{C}^{3g-3+n}$,
- a family $p : \mathcal{C} \to U$ of Riemann surfaces of genus g with n marked points, and
- a G-action on $\mathcal{C}$ descending to a G-action on U

such that the following conditions are satisfied:

1. The fiber $C_0 := p^{-1}(0)$ is (isomorphic to) $\mathcal{C}$.
2. The G-action preserves C_0 and acts as its automorphism group.
3. For any family $p' : \mathcal{C}' \to U'$ of Riemann surfaces of genus g with n marked points such that $p'^{-1}(w) \cong C$ for some $w \in U'$, there exists an open subset $b \in V \subset U'$ and a map $\varphi : V \to U$, unique up to the G-action, such that $p'|_V = \varphi^* p$.

When one takes all sets U/G corresponding to the Riemann surfaces of genus g with n marked points, one gets an orbifold covering. The corresponding orbifold is the moduli space $\mathcal{M}_{g,n}$. The sets $\mathcal{C}$ give an orbifold covering of an orbifold $\mathcal{C}_{g,n}$. Moreover, there exists an orbifold morphism $\pi : \mathcal{C}_{g,n} \to \mathcal{M}_{g,n}$, called the universal curve over $\mathcal{M}_{g,n}$. The fibres of the universal curve are exactly the Riemann surfaces, each only appearing once.

Although a proper geometric structure has been placed on the moduli space, a further problem remains. In general, this space is not compact, which might make it difficult to handle for some purposes. Consider for example the case $g = 0, n = 4$ as shown in Fig. 39.1. By Example 39.5.3, one can uniquely map three marked points to, for example, the points 0, 1 and ∞. The fourth point can, however, still be mapped freely to any point $c \in \mathbb{CP}^1 \setminus \{0, 1, \infty\}$. (This number is the **modulus** that gives rise to the term 'moduli space'.) For $c \to 0$, the marked points x_1 and x_4 would coincide. However, after a coordinate transformation $x \longrightarrow x/c$, the marked points x_2 and x_3 would coincide. To remove this ambiguity, one should replace $\mathbb{CP}^1$ by two copies of $\mathbb{CP}^1$ that intersect (transversally) at a single point. From the point of view of x_2 and x_3 (the original

Figure 39.1: Riemann surface with 4 marked points.

coordinate chart), the marked points x_1 and x_4 have collapsed, while from the point of view of the latter (after the transformation $x \longrightarrow x/c$) the former have collapsed. This 'bubbling' is what will generally happen at singular points. Curves that have this kind of singularity are called 'stable curves'.

Definition 39.5.16 (Stable curve). A compact, complex algebraic curve with $n \in \mathbb{N}$ marked points that satisfies the following conditions:

1. It has only simple **nodal singularities** (i.e. singularities of the form $xy = 0$).
2. The marked points are distinct and do not coincide with the nodes.
3. The curve has a finite automorphism group.

One can **smoothen** a stable curve by replacing the neighbourhood of every node that consists of two disks by a cylinder. A stable curve can also be **normalized** by replacing these intersecting disks by disjoint disks.

Remark 39.5.17. The last condition above is equivalent to requiring that every connected component of the normalization has negative Euler characteristic, i.e. has itself a finite automorphism group.

Property 39.5.18 (Deligne–Mumford compactification). There exists a morphism of compact, complex orbifolds $\overline{\pi} : \overline{\mathcal{C}}_{g,n} \to \overline{\mathcal{M}}_{g,n}$ such that:

- $\mathcal{M}_{g,n} \subset \overline{\mathcal{M}}_{g,n}$ is a suborbifold with preimage $\mathcal{C}_{g,n}$ under $\overline{\pi}$,
- the fibres of $\overline{\pi}$ are stable curves of genus $g \in \mathbb{N}$ with $n \in \mathbb{N}$ marked points,
- each stable curve is isomorphic to a unique fibre, and

- the stabilizer of a point in $\overline{\mathcal{M}}_{g,n}$ is isomorphic to the automorphism group of the corresponding stable curve.

39.5.2 Pseudoholomorphic maps

Definition 39.5.19 (Cauchy–Riemann operator). Consider a complex vector bundle (E, J) over an almost complex manifold (M, j). A Cauchy–Riemann operator on E is a complex-linear map

$$\mathsf{D} : \Gamma(E) \to \Gamma\big(\overline{\mathsf{Hom}}(TM, E)\big) \tag{39.50}$$

satisfying the Leibniz rule

$$\mathsf{D}(f\sigma) = (\overline{\partial} f)\sigma + f(\mathsf{D}\sigma)\,, \tag{39.51}$$

where $\overline{\mathsf{Hom}}$ denotes the bundle of antilinear morphisms, i.e. those vector bundle morphisms that anticommute with the almost complex structures.

This strongly resembles the definition of a Koszul connection (Definition 34.7.1). The following property shows that this is no coincidence.

Property 39.5.20. For every Cauchy–Riemann operator D on a Hermitian vector bundle (E, J) over an almost complex manifold (M, j), there exists a Hermitian connection ∇ such that

$$\mathsf{D}\sigma = \nabla\sigma + J \circ \nabla\sigma \circ j \tag{39.52}$$

for all sections $\sigma \in \Gamma(E)$.

Definition 39.5.21 (Pseudoholomorphic function). Consider two almost complex manifolds (M, J) and (N, j). A function $f : (M, J) \to (N, j)$ is said to be pseudoholomorphic if it satisfies the **Cauchy–Riemann equations**

$$f_* + j \circ f_* \circ J = 0 \tag{39.53}$$

or, equivalently,

$$j \circ f_* = f_* \circ J\,, \tag{39.54}$$

i.e. the differential is complex linear. This can rephrased in terms of a Cauchy–Riemann-like operator

$$\overline{\partial}_J : C^\infty(M, N) \to \Gamma\big(\overline{\mathsf{Hom}}(TM, TM)\big) : f \mapsto f_* + j \circ f_* \circ J\,, \tag{39.55}$$

The kernel of this operator consists of exactly the pseudoholomorphic functions. Note that the image of f under the above nonlinear Cauchy–Riemann operator is actually

an element of $\Gamma\big(\overline{\mathsf{Hom}}(TM, f^*TN)\big)$, since the tangent space at a point $f \in C^\infty(M,N)$ is modelled on $\Gamma(f^*TN)$.[5] By Property 34.4.25, this operator can also be written in terms of the differential $\mathsf{d}f$.

Remark 39.5.22 (Pseudoholomorphic curves). In practice, one often restricts to pseudoholomorphic curves, i.e. pseudoholomorphic functions where the domain is a Riemann surface. On one hand, in two dimensions, one does not lose generality by only considering complex manifolds, since the integrability condition $\overline{\partial}^2 = 0$ is always satisfied. On the other hand, it is better to restrict to two-dimensional manifolds in the domain, because, in general, there are no nonconstant pseudoholomorphic functions (even locally) when the domain has a higher dimension.

Property 39.5.23 (Regularity). If a function of class C^1 satisfies the Cauchy–Riemann equations, it is automatically smooth.

Definition 39.5.24 (Pseudoholomorphic polygon). For every polygon in $\mathbb{C}$, i.e. a complex disk with a finite number of marked points, the *Riemann mapping theorem* gives a biholomorphism to the interior of the disk D. The *Schwarz–Christoffel formula* gives an expression for this map, which can even be extended to the n-punctured disk D_n. A pseudoholomorphic n-gon in an almost complex manifold M is a pseudoholomorphic map $f : D_n \to M$ such that $\lim_{w\to w_i} f(w) = q_i$, where q_i is a vertex of the polygon and w_i is a puncture of D_n.

Property 39.5.25 (Symplectic submanifolds). Let (M,ω) be a symplectic manifold and consider an ω-tame almost complex structure J. Every (complex) line in a tangent space is also a symplectic subspace. Globally, this means that every J-holomorphic curve corresponds to a symplectic submanifold.

Definition 39.5.26 (Energy). Consider a symplectic manifold (M,ω) equipped with an ω-tame almost complex structure J. By symmetrizing the form $\omega(\cdot, J\cdot)$, one obtains a (Riemannian) metric g. The energy of a J-holomorphic curve $f : (\Sigma, J_0) \to (M,J)$ is defined as follows:

$$\mathsf{E}(f) := \int_\Sigma f^*\omega = \frac{1}{2}\int_\Sigma \|\mathsf{d}f\|^2 \,\mathsf{Vol}_\Sigma \,. \tag{39.56}$$

This quantity is always nonnegative and is zero if and only if f is locally constant. (Note that for the second expression, one needs the induced metric on both Σ and M.)

Construction 39.5.27 (Moduli space). Choose two integers $g, n \geq 1$ and a homology class $A \in H_2(M)$. The moduli space $\mathcal{M}^A_{g,n}$ is defined as the set of equivalence classes of tuples $(\Sigma, J_0, f, z_1, \dots, z_n)$, where (Σ, J_0) is a Riemann surface of genus g, $f : (\Sigma, J_0) \to (M,J)$ is a pseudoholomorphic curve such that $f_*[\Sigma] = A$ and $\{z_i\}_{i\leq n}$ are marked points of Σ. Two such tuples are deemed equivalent if there exists an automorphism, i.e. a

[5] To make this statement precise, one needs to endow $C^\infty(M,N)$ with the right topology, since this manifold is infinite dimensional.

biholomorphic or conformal diffeomorphism, that preserves the order of the marked points.

Example 39.5.28. When $M = \{*\}$, the moduli space reduces to $\mathcal{M}_{g,n}$, the moduli space of Riemann surfaces of genus $g \in \mathbb{N}$ with $n \in \mathbb{N}$ marked points.

39.6 Cohomology

39.6.1 Dolbeault cohomology

Theorem 39.6.1 (Hodge decomposition). *Let M be a compact Kähler manifold.*

$$H^k_{\mathrm{dR}}(M) \cong \bigoplus_{p+q=k} H^{p,q}(M) \tag{39.57}$$

for all $k \in \mathbb{N}$

By analogy with the Poincaré lemma for smooth manifolds, one can prove the following theorems.

Theorem 39.6.2 (∂-lemma). *Let $\alpha \in \Omega^{p,q}(M)$. If $\partial\alpha = 0$, then, locally, there exists a complex form $\beta \in \Omega^{p-1,q}$ such that $\alpha = \partial\beta$.*

Theorem 39.6.3 ($\overline{\partial}$-lemma). *Let $\alpha \in \Omega^{p,q}(M)$. If $\overline{\partial}\alpha = 0$, then, locally, there exists a complex form $\beta \in \Omega^{p,q-1}$ such that $\alpha = \overline{\partial}\beta$.*

Theorem 39.6.4 ($\partial\overline{\partial}$-lemma). *Let $\alpha \in \Omega^{p,q}(M)$. If $\mathrm{d}\alpha = 0$, then, locally, there exists a complex form $\beta \in \Omega^{p-1,q-1}$ such that $\alpha = \partial\overline{\partial}\beta$.*

Theorem 39.6.5 (Kodaira vanishing theorem). *Let L be a positive holomorphic line bundle on a compact complex manifold M.*

$$H^p(M, L \otimes K_M) = 0 \tag{39.58}$$

for all $p > 0$, where K_M is the canonical class of M.

39.6.2 Lagrangian Floer homology ♣

In this section, a (co)homology theory is constructed from the intersection theory of Lagrangian submanifolds (see Chapter 37 for an introduction).

Recall the pseudoholomorphic polygons from Section 39.5.2. In the study of symplectic manifolds, one often has a set of boundary conditions $\{L_i, q_i\}_{i \leq n}$, where the L_i are Lagrangian submanifolds and the $q_i \in L_i \cap L_{i+1 \bmod n}$ are intersection points. A pseudoholomorphic n-gon satisfies these boundary conditions if the edges of the domain are mapped to the submanifolds L_i and the marked points to the intersection points q_i.

Property 39.6.6 (Moduli space). To fully appreciate Floer theory, one needs to study the moduli space of pseudoholomorphic polygons. One of the most important parts being compactness and the *Gromov compactification*. For certain one-parameter families of polygons, the limit is a degenerate configuration that is not part of the moduli space itself. Consider, for example, the situation in Fig. 39.2. The Lagrangian submanifolds are indicated by black lines (infinite straight lines are Lagrangian submanifolds of the complex plane). The interior of the polygon is sketched by a red line and the marked points are indicated by red dots.

Figure 39.2: Pseudoholomorphic 4-gon.

Now, one can make a one-parameter family of 4-gons that satisfies the same boundary conditions by, instead of going directly from vertex 3 to vertex 4, first going a bit down and then going up again, i.e. making a 'slit'. This still satisfies all the properties to apply the *Riemann mapping theorem*, but gives a different polygon map. At a certain point, a new vertex is created on the lower Lagrangian, a so-called **nodal** vertex, and the 4-gon is broken up into two 3-gons. From the domain point of view, what happens is that two boundary punctures have collided and a new disk (with two punctures) has been attached. Adding such configurations to the moduli space to account for these limit operations gives rise to a Deligne–Mumford compactification (Property 39.5.18). Because finding all possible ways to arrange $n+1$ points on different attached disks (without altering their order) is equivalent to finding all possible parenthesations of $n+1$ symbols, one finds that the compactified moduli space of boundary-punctured disks $\overline{\mathcal{M}}_{n+1}$ is isomorphic to the nth Stasheff polytope K_n.

In the case of bigons (for simplicity), three types of degeneracies can occur:

- **Strip breaking**: Here, energy concentrates at one of the marked points. There exists a sequence of bigons $(f_n)_{n\in\mathbb{N}}$, viewed as homotopies, and a diverging sequence of numbers $(a_n)_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} f_n(s-a_n,t)$ is not the constant strip. This corresponds to the situation in Fig. 39.2.

- **Disk bubbling**: Here, energy concentrates at a point on the boundary. There exists a sequence of bigons that can be rescaled such that the limit is a pseudoholomorphic disk entirely contained in one of the boundary Lagrangians.

- **Sphere bubbling**: Here, energy concentrates at a point in the interior of a bigon. There exists a sequence of bigons that can be rescaled such that the limit is a pseudoholomorphic sphere entirely contained in the interior of M.

The latter two issues can be avoided by imposing topological restrictions such as by restricting to *monotone Lagrangian submanifolds* or by requiring that

$$[\omega] \cdot \pi_2(M, L) = 0 = [\omega] \cdot \pi_2(M, L') = 0\,. \tag{39.59}$$

Definition 39.6.7 (Floer complex). Let M be a symplectic manifold equipped. For every two transversally intersecting Lagrangian submanifolds, the chain group is defined as the free vector space generated by the intersection points: $CF(L, L') := \Lambda^{L \cap L'}$, where, to avoid technical difficulties, one should take the field Λ to be a *Novikov field*.

The differential is defined by counting pseudoholomorphic bigons:

$$\partial\langle p\rangle := \sum_{q \in L \cap L' \setminus \{p\}} n(p, q)\langle q\rangle\,, \tag{39.60}$$

where $n(p, q)$ denotes the number of (isomorphism classes of[6]) pseudoholomorphic bigons with boundary conditions $\{L, L', p, q\}$ of finite symplectic energy.

To be more precise, denote by $\mathcal{M}(p, q; [f])$ the moduli space of pseudoholomorphic bigons of finite symplectic energy in a fixed homotopy class $[f] \in \pi_2(M, L \cap L')$. The Maslov index of such a class can be defined using the spectral flow approach.

$$\mu([f]) := \mathsf{ind}(\mathsf{D}_{\bar{\partial}_J})\,, \tag{39.61}$$

where $\mathsf{D}_{\bar{\partial}_J}$ denotes the linearization of the Cauchy–Riemann operator associated to J (it can be shown that this operator is Fredholm and, hence, admits a well-defined index). When all bigons are regular, i.e. when the linearized Cauchy–Riemann operator is surjective everywhere, the moduli space has dimension $\mu([f]) - 1$. Moreover, the *Gromov compactness theorem* (see the property above) states that this manifold is compact.

Putting a well-defined grading on $CF(L, L')$ is a bit more subtle. Let $\mathsf{LGr}(TM)$ denote the Lagrangian Grassmann bundle over M. The Maslov index of a bigon should only depend on the difference in degrees of its marked points and not on its homotopy class. This is implemented as follows (only a $\mathbb{Z}$-grading is considered here).

Definition 39.6.8 (Maslov covering). A **Maslov covering** of M is a $\mathbb{Z}$-covering $\mathcal{L} \to \mathsf{LGr}(TM)$ such that the fibre over every Lagrangian is given by the $\mathbb{Z}$-covering of $\mathsf{LGr}(\mathbb{R}^{\dim(M)})$. This corresponds to patching together the universal covers of the Lagrangian Grassmannians at every point of M.

[6]The domain, i.e. the disk with two boundary punctures, has as automorphism group $\mathbb{R}$. It scales the distance between the punctures.

A Maslov covering exists if and only if the first Chern class is 2-torsion:

$$2c_1(M) = 0\,. \tag{39.62}$$

This can be restated in cohomological conditions involving the Chern class (Moshayedi, 2021). $c_1(M)$ should be 2-torsion and $\mu \in H^1$ should vanish for both L and L'. This allows to lift a Lagrangian submanifold to a graded Lagrangian submanifold, a section of the universal cover of the Lagrangian Grassmann bundle.

$$\partial\langle p\rangle = \sum_{\substack{q\in L\cap L'\\ \mathrm{ind}([f])=1}} |\mathcal{M}(p,q;[f])|\, T^{\varepsilon([f])}\langle q\rangle\,, \tag{39.63}$$

where T is the generator of the *Novikov field* Λ and $\varepsilon([f])$ denotes the symplectic energy of $[f]$.

Property 39.6.9 (Hamiltonian isotopy). Consider two Lagrangian submanifolds L, L'. If there exists a Hamiltonian isotopy $L' \to L''$, then

$$HF(L,L') \cong HF(L,L'')\,. \tag{39.64}$$

If there exists a Hamiltonian isotopy $L \to L'$, then

$$HF(L,L') \cong H^\bullet(L)\,, \tag{39.65}$$

where the right-hand side denotes singular cohomology.

Definition 39.6.10 (Symplectic action functional). For every two Lagrangian submanifolds $L, L' \subset M$, one can consider the space of smooth paths connecting L and L':

$$\mathcal{P}(L,L') := \{\gamma \in C^1([0,1],M) \mid \gamma(0)\in L, \gamma(1)\in L'\}\,. \tag{39.66}$$

The symplectic action functional is defined as follows:

$$A : \widetilde{\mathcal{P}}(L,L') \to \mathbb{R} : [\gamma,h] \mapsto \int_{[0,1]\times[0,1]} h^*\omega\,, \tag{39.67}$$

where ω is the symplectic form on M and $\widetilde{\mathcal{P}}$ denotes the universal cover (of a connected component) of $\mathcal{P}$, i.e. the set of equivalence classes of pairs (γ, h) where h is a homotopy between γ and a fixed point in the connected component.[7] If the symplectic form is trivializable on h, Stokes' theorem implies that this integral is, up to a constant depending on the choice of base point, equal to

$$\int_{[0,1]} \gamma^*\theta\,, \tag{39.68}$$

where θ is the symplectic potential.

[7] To be correct, one should use the so-called *Novikov covering*.

One can also calculate the differential of the action functional or, by introducing a (family of) compatible almost complex structure(s), the gradient of the action functional:

$$\mathrm{grad}A([\gamma, h]) = J_t \frac{\partial \gamma}{\partial t} . \tag{39.69}$$

It follows, that the critical points of the functional correspond to constant paths, i.e. intersection points of the Lagrangian submanifolds. Moreover, the flow equation of this gradient is exactly the Cauchy–Riemann equation of J-holomorphic curves with the given Lagrangian boundary conditions. In (finite-dimensional) Morse homology (Section 34.10), one counts flow lines of the Hamiltonian flow, so it follows that one can interpret Lagrangian Floer homology as the infinite-dimensional analogue of Morse homology for the symplectic action functional.

Definition 39.6.11 (Fukaya category). Consider a symplectic manifold (M, ω). The associated Fukaya category consists of the following data:

1. **Objects**: Lagrangian submanifolds of M.
2. **Morphisms**: When $L \pitchfork L'$, $\mathrm{Hom}(L, L') := CF(L, L')$, the (Lagrangian) Floer chain group.

This definition can be generalized to obtain an A_∞*-category*. Whenever the Lagrangian submanifolds intersect transversally, a multiplication map

$$\mu : \mathrm{Hom}(L', L'') \otimes \mathrm{Hom}(L, L') \to \mathrm{Hom}(L, L'') \tag{39.70}$$

can be defined. Consider intersection points $q_1 \in L \cap L', q_2 \in L' \cap L''$ and $q \in L \cap L''$. The coefficient of q in $\mu(q_2, q_1)$ is obtained by counting pseudoholomorphic 3-gons with boundary on the Lagrangians and marked points q_1, q_2 and q.

Remark 39.6.12 (Bubbles). Aside from breaking and splitting of holomorphic polygons, another situation where degenerate polygons arise is the so-called bubbling phenomenon, where a holomorphic sphere (0-gon) pops up. Unless the symplectic manifold and the boundary Lagrangian are exact, these give rise to a 'vacuum constant' in the sense that the A_∞-structure is modified to a curved A_∞-structure with $m_0 \neq 0$.

Chapter 40

Calculus of Variations

The standard references for (global) variational calculus are Anderson (1992); Takens (1979).

Chapter Contents

40.1 Introduction **754**

40.2 Jet bundles **756**

40.2.1 Topology 756

40.2.2 Contact structure 758

40.2.3 Differential operators 758

40.3 Variational bicomplex ♣ **759**

40.3.1 Differential structure 759

40.3.2 Euler operators 763

40.3.3 Functional complex 765

40.3.4 Structure of the bicomplex 766

40.3.5 Variational problems 770

40.3.6 Inverse problem 772

40.3.7 Finite jet bundles 773

40.4 Partial differential equations **774**

40.4.1 Algebraic formulation 774

40.4.2 Symmetries 776

40.4.3 Pseudogroups ♣ 777

40.1 Introduction

Definition 40.1.1 (Variational symmetry). Consider an integral quantity I defined by a Lagrangian function:

$$I_M = \int_M L(q, u, u_I)\, dq\,, \tag{40.1}$$

where u are (analytic) functions of the variables q. A transformation $(q, u) \longrightarrow (q', u')$ of the variables[1] is called a variational (or Noether) symmetry if it satisfies

$$\int_{M'} L(q', u', u'_I)\, dq' = \int_M L(q, u, u_I)\, dq \tag{40.2}$$

for arbitrary M.

Following *Lie*, the notion of a group of transformations is introduced.

Definition 40.1.2 (Finite continuous group). A collection of analytic functions, closed under inverses and composition, such that every function depends analytically on a finite number of parameters. In this chapter, these groups will be denoted by $\mathfrak{G}_k$, where $k \in \mathbb{N}$ is the number of independent parameters.

Remark 40.1.3. It should be clear that this is the same as a finite-dimensional Lie group (Definition 32.1.1).

Instead of structure parameters, one can also generalize to structure functions.

Definition 40.1.4 (Infinite continuous group). A collection of analytic functions, closed under inverses and composition, such that every function depends analytically on a finite number of arbitrary (analytic) functions. In this chapter, these groups will be denoted by $\mathfrak{G}_{\infty,k}$, where $k\mathbb{N}$ is the number of independent functions.

Remark 40.1.5. In physics terminology, the infinite groups would be obtained by 'gauging' a global symmetry group $\mathfrak{G}_k$.

Theorem 40.1.6 (Noether). *Consider an integral quantity I that is invariant under a continuous group $\mathfrak{G}$.*

I) If $\mathfrak{G}$ is finite continuous of the form $\mathfrak{G}_k$, there exist $k \in \mathbb{N}$ independent (linear) combinations among the Lagrangian expressions that are equal to divergences. Conversely, if there exist k independent combinations among the Lagrangian expressions that are divergences, then I is invariant under a group of the form $\mathfrak{G}_k$.

II) If $\mathfrak{G}$ is infinite continuous of the form $\mathfrak{G}_{\infty,k}$, there exist $k \in \mathbb{N}$ independent relations among the Lagrangian expressions and their derivatives[2]. Conversely, if k such relations exist,

[1] The transformations of the derivatives ∂u are induced by the ones for u.

[2] The order up to which the derivatives occur is equal to the order of derivatives up to which the transformations depend on the k arbitrary functions.

then I is invariant under a group of the form $\mathfrak{G}_{\infty,k}$.

Remark 40.1.7. In fact, the first theorem is also valid in the limit of an infinite number of parameters.

Property 40.1.8 (Improper relations). Divergence relations $\sum_i \psi_i \bar{\delta} u_i = \nabla \cdot B$ obtained in a variational problem with symmetry group $\mathfrak{G}_k$ can be classified into two groups:

- If the quantities B are linear combinations of Lagrangian expressions (and their derivatives) and divergence-free quantities, the divergence relations are said to be improper.
- In all other cases, the divergence relations are said to be proper.

Theorem 40.1.9 (Noether's third theorem). *A finite continuous symmetry group $\mathfrak{G}_k$ of an integral quantity is a subgroup of an infinite continuous symmetry group $\mathfrak{G}_{\infty,k}$ if and only if the divergence relations are improper.*

For Lagrangians describing 'point particles', where $M \subseteq \mathbb{R}$, the following result is obtained.

Example 40.1.10 (One dimension). Consider the following infinitesimal transformations

$$\begin{aligned} q^i &\longrightarrow q^i + \varepsilon \xi^i(q^k, t)\,, \\ t &\longrightarrow t + \varepsilon \tau(q^k, t)\,, \\ \dot{q}^i &\longrightarrow \dot{q}^i + \varepsilon(\dot{\xi}^i - \dot{q}^i \dot{\tau})\,, \end{aligned} \tag{40.3}$$

where the transformation of the 'velocities' on the last line, induced by the coordinate transformations, is called a **prolongation** (see also Definition 40.3.3 further below). These transformations generate Noether symmetries if they leave the Lagrangian invariant up to a total derivative (at first order) for every subinterval $[t_0, t_1] \subseteq [a, b]$ and for some function $f(q, t)$:

$$\int_{\tilde{t}_0}^{\tilde{t}_1} L(\tilde{q}, \dot{\tilde{q}}, \tilde{t})\, d\tilde{t} = \int_{t_0}^{t_1} L(q, \dot{q}, t)\, dt + \varepsilon \int_{t_0}^{t_1} \frac{\mathrm{d} f}{\mathrm{d} t}\, dt + O(\varepsilon^2)\,. \tag{40.4}$$

This is equivalent to requiring that the transformation is a solution of the following differential equation (sometimes called the **Rund–Trautman identity**):

$$\frac{\partial L}{\partial t}\tau + \frac{\partial L}{\partial q^i}\xi^i + \frac{\partial L}{\partial \dot{q}^i}(\dot{\xi}^i - \dot{q}^i \dot{\tau}) + L\dot{\tau} = \dot{f}\,. \tag{40.5}$$

By Noether's (first) theorem one obtains, for every such symmetry, a conserved quantity of the following form

$$F := f - \left[L\tau + \frac{\partial L}{\partial \dot{q}^i}(\xi^i - \dot{q}^i \tau) \right]. \tag{40.6}$$

It is important to note that the left-hand side of Eq. (40.5) is simply the Lie derivative of the Lagrangian form Ldt with respect to the vector field that generates the transformations (40.3). (This will turn out to be an important concept in the calculus of variations. See Definition 40.3.40 further on.)

40.2 Jet bundles

40.2.1 Topology

Although the following constructions can be defined in the general context of fibred manifolds, they will only be considered in the case of smooth fibre bundles. Only the notion of a jet will be defined in general for functions between smooth manifolds.

Definition 40.2.1 (Jet). Consider two smooth manifolds M, N and let $r \in \mathbb{N}$. Smooth functions $f, g \in C^\infty(M, N)$ with local coordinates (f^i) and (g^i) are said to define the same r-jet at a point $p \in M$ if and only if

$$\left.\frac{\partial^\alpha f^i}{\partial x^\alpha}\right|_p = \left.\frac{\partial^\alpha g^i}{\partial x^\alpha}\right|_p \tag{40.7}$$

for all $0 \leq i \leq \dim(M)$ and every multi-index α with $0 \leq |\alpha| \leq r$. It is clear that this defines an equivalence relation. The r-jet at $p \in M$ of a representative f is denoted by $j^r_p f$. The integer r is called the **order** of the jet. The set of all r-jets of functions between manifolds M and N is denoted by $J^r(M, N)$.

Definition 40.2.2 (Jet projections). Let M, N be smooth manifolds and consider the jet space $J^r(M, N)$. The **source** and **target projections** are defined as follows:

$$\begin{aligned} \pi_r &: J^r(M, N) \to M : j^r_p f \mapsto p\,, \\ \pi_{r,0} &: J^r(M, N) \to N : j^r_p f \mapsto f(p)\,. \end{aligned} \tag{40.8}$$

One can also define a **k-jet projection** $\pi_{r,k}$ as the map

$$\pi_{r,k} : J^r(M, N) \to J^k(M, N) : j^r_p f \mapsto j^k_p f\,, \tag{40.9}$$

where $k \leq r$. The k-jet projections satisfy the transitivity property $j_{k,m} = j_{r,m} \circ j_{k,r}$.

Definition 40.2.3 (Prolongation). Let $f : M \to N$ be a smooth function. The r-jet prolongation $j^r f$ is defined as the following map:

$$j^r f : M \to J^r(M, N) : p \mapsto j^r_p f\,. \tag{40.10}$$

Property 40.2.4 (Topology). Every diffeomorphism $f : M \to M'$ induces a morphism of jet spaces $J^r(g, N) : J^r(M', N) \to J^r(M, N)$ by pullback. Similarly, every smooth function $g : N \to N'$ induces a morphism of jet spaces $J^r(M, g) : J^r(M, N) \to J^r(M, N')$ by pushforward.

Now, consider two charts (U, φ) and (V, ψ) of M and N, respectively. Construct the morphism

$$J^r(\varphi^{-1}, \psi) := J^r(\varphi^{-1}, V) \circ J^r(\varphi(U), \psi) : J^r(U, V) \to J^r\big(\varphi(U), \psi(V)\big) . \tag{40.11}$$

This morphism is invertible and, hence, can be used as a local chart. The topology on $J^r(M, N)$ is induced by the manifold topology corresponding to the maximal atlas of all such charts.

Definition 40.2.5 (Whitney C^k-topology). Let M, N be two smooth manifolds and consider the manifold of k-jets $J^k(M, N)$. A basis for the Whitney C^k-topology on $C^\infty(M, N)$ is given by the sets

$$S^k(U) := \{ f \in C^\infty(M, N) \mid j^k f \in U \} , \tag{40.12}$$

where U is open in $J^k(M, N)$.

Property 40.2.6. When the manifold M is compact, the Whitney and compact-open topologies on $C^\infty(M, N)$ coincide. In general, the Whitney topology is the topology of global uniform convergence.

Remark 40.2.7 (Fibre bundles). The above definitions can also be used to define jet manifolds of (local) sections of bundles. One just sets $M = B$ and $N = E$ and restricts to the subset satisfying

$$j^r_{\sigma(p)}\pi \circ j^r_p\sigma = j^r_p \mathbb{1}_M . \tag{40.13}$$

This is just the prolongation of the definition of sections $\pi \circ \sigma = \mathbb{1}_M$.

The r-jet bundle corresponding to the projection π is defined as the triple $(J^r(E), B, \pi_r)$. The bundle charts (U_i, φ_i) on E define induced bundle charts on $J^r(E)$ in the following way:

$$U^r_i := \{ j^r_p\sigma \mid \sigma(p) \in U_i \} \tag{40.14}$$

$$\varphi^r_i := \left(x^k, u^\alpha, \left.\frac{\partial^I u^\alpha}{\partial x^I}\right|_p \right), \tag{40.15}$$

where I is a multi-index such that $0 \leq |I| \leq r$. The partial derivatives

$$\left.\frac{\partial^I u^\alpha}{\partial x^I}\right|_p$$

are called the **derivative coordinates** on $J^r(E)$.

Definition 40.2.8 (Holonomic section). Consider a fibre bundle $\pi : E \to M$. A (local) section σ of π gives rise to a (local) section of π_r given by the r-jet prolongation of σ. Sections of $J^r(E)$ that lie in the image of j^r are said to be holonomic.

Definition 40.2.9 (Infinite jet bundle). The inverse limit (Definition 1.8.4) of the projections $\pi_{k,k-1} : J^k(E) \to J^{k-1}(E)$. It can be shown (in an algebro-geometric fashion) that the smooth functions on the infinite jet bundle $J^\infty(E)$ are (at least locally) given by smooth functions on some finite jet bundle. This just means that the infinite jet bundle is defined by taking its algebra of smooth functions to be the direct limit of those on the finite jet bundles. By extension, it can be shown that any smooth morphism $J^\infty(E) \to E'$ into a finite-dimensional manifold factorizes through a finite jet bundle. Furthermore, a map $Q \to J^\infty(E)$ or $J^\infty(E) \to J^\infty(E')$ is said to be smooth if the composition with any smooth map is again smooth.

Property 40.2.10 (Jet comonad). The assignment of jet bundles constitutes a comonad: $J^\infty : \mathbf{Bundle}(X) \to \mathbf{Bundle}(X)$. The counit sends the prolongation of a section to the section evaluated at a point. The comultiplication reshuffles derivatives.

40.2.2 Contact structure

Every jet manifold J^k carries a natural constact structure (Definition 38.1.3).

Definition 40.2.11 (Contact form). Given a fibre bundle $\pi : E \to M$ and one of its jet bundles $J^k(E)$, a differential form $\omega \in \Omega^\bullet(J^k(E))$ is called a contact form if it is annihilated by all jet prolongations, i.e. $(j^{k+1}\sigma)^*\omega = 0$ for all sections $\sigma \in \Gamma(E)$.

The space of such contact forms generates a differential ideal and, in turn, defines a distribution, called the **Cartan distribution**. For finite-order jet spaces, this distribution is completely nonintegrable. However, for infinite jet bundles (where Frobenius's theorem need not hold), the distribution becomes involutive and integrable.

Locally, the differential ideal is generated by the following contact forms:

$$\theta_I^\alpha := \mathbf{d}u_I^\alpha - \sum_\mu u_{I,\mu}^\alpha \mathbf{d}x^\mu \,, \tag{40.16}$$

where the x^μ and u_I^α are the independent and dependent variables, respectively.

Property 40.2.12 (Cartan connection). There exists a connection on the infinite jet bundle $J^\infty(E)$ where the horizontal subbundle is exactly given by the Cartan distribution. This connection can be shown to be flat.

40.2.3 Differential operators

Alternative Definition 40.2.13 (Differential operator). Let E_1, E_2 be smooth fibre bundles over the same base manifold M. A differential operator $\widetilde{D} : E_1 \to E_2$ is a bundle morphism $J^\infty(E_1) \to E_2$. This induces a map of sections $D : \Gamma(E_1) \to \Gamma(E_2)$ such that $D = \widetilde{D} \circ j^\infty$. It is said to be of **order** $k \in \mathbb{N}$ if it factors through the k-jet bundle, i.e. if j^∞ can be replaced by j^k. If the bundle morphism $\widetilde{D}$ is a vector bundle morphism, the differential operator is said to be **linear**.

This definition can easily be phrased in terms of the jet comonad. The category of differential operators is equivalent to the (co)Kleisli category (Definition 3.3.21) of the jet comonad J^∞.

Property 40.2.14 (Formal adjoints). Consider two differential operators

$$D, D^\dagger : \Gamma(E) \to \Gamma\big(E^* \otimes \Lambda^{\dim(M)}(M)\big)\,. \tag{40.17}$$

These operators are said to be formally adjoint if there exists a bilinear differential operator $K : \Gamma(E) \otimes \Gamma(E) \to \Omega^{\dim(M)}(M)$ such that the following condition is satisfied for all sections σ_1, σ_2 of E:

$$\langle D(\sigma_1), \sigma_2\rangle - \langle \sigma_1, D^\dagger(\sigma_2)\rangle = \mathrm{d}K(\sigma_1, \sigma_2)\,. \tag{40.18}$$

This formula can be interpreted as a generalization of Green's identities. In the case where M is compact, Stokes' theorem 34.8.25 shows that D and $D^\dagger$ are related through integration by parts.

Definition 40.2.15 (Generalized vector field). Consider a vector bundle $\pi : E \to M$. A generalized vector field on M is a 'vector field' whose components are, locally, given by smooth functions on the infinite jet bundle $J^\infty(E)$, i.e. it is a smooth map $X : J^\infty(E) \to TM$ and, hence, a smooth map $X : J^k(E) \to TM$ for some $k \in \mathbb{N}$ such that $X(\sigma) \in T_{\pi_k(\sigma)}M$.

40.3 Variational bicomplex ♣

In this section, the language of jet bundles, as introduced in the previous section, will be used to rephrase the classical theory of variational calculus in more general geometric terms. A smooth function on the infinite jet bundle will be denoted by $f[u]$, i.e. the arguments will be written inside square brackets.

40.3.1 Differential structure

A smooth fibre bundle $\pi : E \to M$ over a base manifold M will be considered.[3] First, the de Rham operator $\mathbf{d}$ on the infinite jet bundle $J^\infty(E)$ is decomposed in a horizontal and a vertical part:

$$\mathbf{d} = \mathrm{d} + \delta\,. \tag{40.19}$$

The **horizontal derivative** d lifts the de Rham differential on M to E (hence the name). On smooth functions in $C^\infty(J^\infty(E))$, it acts as follows:

$$\mathrm{d}f := (D_\mu f)\mathrm{d}x^\mu\,, \tag{40.20}$$

[3] In fact, the fibre bundle can be replaced with any fibred manifold.

where the **total derivative**[4]

$$D_\mu f := \frac{\partial f}{\partial x^\mu} + \frac{\partial f}{\partial u^\alpha} u^\alpha_{,\mu} + \frac{\partial f}{\partial u^\alpha_{,\nu}} u^\alpha_{,\nu\mu} + \cdots \tag{40.21}$$

is introduced. The horizontal de Rham operator can be extended to all of $\Omega^\bullet(J^\infty(E))$ through the Leibniz property and the condition

$$\mathsf{d} \circ \delta = -\delta \circ \mathsf{d}\,, \tag{40.22}$$

which follows from the nilpotency of the differentials. The differentials d, δ turn $\Omega^\bullet(J^\infty(E))$ into a bigraded complex, called the **variational bicomplex**.

Remark 40.3.1. Some authors use the term variational bicomplex for the bicomplex of **local forms** $\Omega^\bullet_{\mathsf{loc}}(M \times \Gamma(E))$, which is defined as the image of $\Omega^\bullet(J^\infty(E))$ under the prolongation map $M \times \Gamma(E) \to J^\infty(E)$. This way, they can work with forms over the (trivial) field bundle, while maintaining the property that all objects only depend on finite-order jets. Furthermore, when working in full generality the de Rham complex over M is twisted by the orientation bundle (Remark 34.8.11).

Definition 40.3.2 (Local Lagrangian). A top-degree horizontal form on $J^\infty(E)$. Because $\Omega^{\dim(M),0}(J^\infty(E))$ is one-dimensional, such forms are proportional to the volume form:

$$\mathbf{L} = L\,\mathsf{Vol}\,, \tag{40.23}$$

where L is a smooth function on the infinite jet bundle. The function L is called the **Lagrangian density**. By its very nature, this implies that L locally only depends on partial derivatives up to some finite order.

For any such horizontal form, one can define a functional on $\Gamma_c(M)$ as follows:

$$S : \phi \mapsto \int_M (j^\infty \phi)^* \mathbf{L}\,. \tag{40.24}$$

Functions of this form are called **local functionals**.

Now, consider a (generalized) vector field X on the total space E. This vector field can be lifted to $J^k(E)$ in a canonical way.

Definition 40.3.3 (Prolongation of vector fields). Given a generalized vector field X on a fibre bundle $\pi : E \to M$, there exists a unique vector field $j^k E$ on the jet bundle $J^k(E)$ defined by the following conditions:

1. X and $j^k X$ coincide on $C^\infty(E)$.
2. $j^k X$ preserves the contact ideal, i.e. if θ is a contact form, then $\mathcal{L}_{j^k X}\theta$ is also a contact form.

[4] In fact, this formula is virtually the same as the one for the true total derivative. However, partial derivatives are replaced by jet coordinates.

Locally, the prolongation of a vector field

$$X = X^\mu \partial_\mu + X^\alpha \partial_\alpha \tag{40.25}$$

is given by

$$(j^k X)^\alpha_I = D_I(X^\alpha - u^\alpha_\mu X^\mu) + u^\alpha_{\mu I} X^\mu \tag{40.26}$$

for all $|I| \leq k$.

A similar definitions exists for vector fields on the base manifold.

Definition 40.3.4 (Total vector field). Given a generalized vector field X on M, its total vector field $\mathrm{tot}X$ is defined by the following conditions:

1. X and $\mathrm{tot}X$ coincide on $C^\infty(M)$, and
2. $\mathrm{tot}X \lrcorner\, \omega = 0$ if ω is a contact form.

An explicit formula can be obtained by replacing partial derivatives by total derivatives:

$$X = X^\mu \partial_\mu \longrightarrow \mathrm{tot}X = X^\mu D_\mu \,. \tag{40.27}$$

In particular, the total vector fields associated to a coordinate-induced basis ∂_μ are exactly the total derivatives D_μ.

Definition 40.3.5 (Evolutionary vector field). A generalized vector field that projects to 0 on the base manifold, i.e. a π-vertical (generalized) vector field. The space of evolutionary vector fields on E is denoted by $\mathsf{Ev}(J^\infty(E))$.

The prolongation of an evolutionary vector field to $J^\infty(E)$ still projects to 0 on M. By extension, all vector fields on $J^\infty(E)$ that preserve the contact ideal and project to 0 on M are called evolutionary vector fields. Such vector fields are of the form

$$X = X^\alpha_I[u]\partial^I_\alpha \,. \tag{40.28}$$

The name stems from the fact that these vector fields define PDEs that (locally) describe the evolution of the fibres.

The prolongation of an evolutionary vector field can be written as follows:

$$j^\infty X = \sum_{|I|=0}^{+\infty} (D_I X^\alpha)\partial^I_\alpha \,. \tag{40.29}$$

Property 40.3.6. By writing out Cartan's magic formula 34.4.18 with respect to $\mathbf{d}$, one can prove that the prolongation of an evolutionary vector field also satisfies this formula with respect to δ and that $\iota_{j^\infty X}$ and d anticommute.

Property 40.3.7 (Evolutionary decomposition). Consider a generalized vector field

$$X = X^\mu \partial_\mu + X^\alpha \partial_\alpha \tag{40.30}$$

on E. By extending the tot-construction to generalized vector fields as

$$\mathsf{tot} X := \mathsf{tot}(\pi_* X)\,, \tag{40.31}$$

one can define the evolutionary part of X as follows:

$$X_{\mathsf{ev}} := X - (\pi_{\infty,0})_*(\mathsf{tot} X)\,. \tag{40.32}$$

Locally, this can be written as

$$X_{\mathsf{ev}} = \left(X^\alpha - u^\alpha_\mu X^\mu\right)\partial_\alpha\,. \tag{40.33}$$

Using this definition, the prolongation of X can be decomposed as follows:

$$j^\infty X = j^\infty X_{\mathsf{ev}} + \mathsf{tot} X\,. \tag{40.34}$$

The evolutionary part is sometimes also called the **characteristic** of X.

To close this section, the Cartan calculus of the variational bicomplex is summarized.

Formula 40.3.8 (Cartan calculus). Let X, Y denote a generalized vector field and an evolutionary vector field, respectively.

1. Exterior derivatives:

$$\mathsf{d}x^\mu = \mathbf{d}x^\mu\,, \tag{40.35}$$
$$\delta u^\alpha_I = \mathbf{d}u^\alpha_I - u^\alpha_{I\mu}\mathsf{d}x^\mu\,. \tag{40.36}$$

2. Interior product:

$$\{\mathsf{d}, \iota_{j^\infty Y}\}_+ = 0\,. \tag{40.37}$$

3. Lie derivatives:

$$\mathcal{L}_{j^\infty X} = \mathbf{d}\iota_{j^\infty X} + \iota_{j^\infty X}\mathbf{d}\,, \tag{40.38}$$
$$\mathcal{L}_{j^\infty Y} = \delta\iota_{j^\infty Y} + \iota_{j^\infty Y}\delta\,, \tag{40.39}$$
$$[\mathsf{d}, \mathcal{L}_{j^\infty Y}] = [\delta, \mathcal{L}_{j^\infty Y}] = 0\,. \tag{40.40}$$

40.3.2 Euler operators

Property 40.3.9 (Total differential operators). A differential operator $P : \mathsf{Ev}(J^\infty(E)) \to \Omega^\bullet(J^\infty(E))$ that can locally be written as

$$P(X) = \sum_{|I|=0}^{k} \big(D_I X^\alpha\big) P^I_\alpha \tag{40.41}$$

for some differential forms P^I_α. By (formally) integrating by parts, this can locally be written as

$$P(X) = \sum_{|I|=0}^{k} D_I\big(X^\alpha Q^I_\alpha\big)\,, \tag{40.42}$$

where the smooth forms Q^I_α can be expressed as follows:

$$Q^I_\alpha = \sum_{|J|=0}^{k-|I|} \binom{|I|+|J|}{|J|} (-1)^{|J|} D_J P^{IJ}_\alpha\,. \tag{40.43}$$

The zeroth-order part Q_α also defines a total differential operator:

$$E_P(X) := X^\alpha Q_\alpha\,. \tag{40.44}$$

This operator is called the **Euler operator** (associated to P).

Property 40.3.10 (Decomposition of total differential operators). Consider a total differential operator $P : \mathsf{Ev}(J^\infty(E)) \to \Omega^{\dim(M),k}(J^\infty(E))$. The Euler operator E_P is the unique globally defined (zeroth order) operator such that, on each chart U, one can find a total differential operator R_U that satisfies the following equation:

$$P(X) = E_P(X) + \mathsf{d}R_U(X)\,. \tag{40.45}$$

The operator R_U can locally be expressed as follows:

$$R_U(X) = \sum_{|I|=0}^{k-1} D_I\big(X^\alpha D_\mu \lrcorner\, Q^{\mu I}_\alpha\big)\,. \tag{40.46}$$

The decomposition above can, in fact, be shown to hold globally. However, in that case, the expression of R cannot be expressed easily in terms of the coefficients of P (only for P of order 2 does a canonical expression exist).

Example 40.3.11 (Euler–Lagrange operator). Consider a local Lagrangian $\mathbf{L}$. This form induces a total differential operator as follows:

$$P_{\mathbf{L}}(X) := \mathcal{L}_{j^\infty X}\mathbf{L} = \sum_{|I|=0}^{k} (D_I X^\alpha)(\partial^I_\alpha L)\,\mathsf{Vol}\,. \tag{40.47}$$

The coefficients (40.43) associated to this operator are given by the following formula (to turn these coefficients into true differential forms, one should multiple them by the volume form):

$$E^I_\alpha(L) := \sum_{|J|=0}^{k-|I|} \binom{|I|+|J|}{|J|} (-1)^{|J|} D_J\left(\partial^{IJ}_\alpha L\right). \tag{40.48}$$

The induced Euler operator is exactly the Euler–Lagrange operator associated to variational problems. For this reason, and the fact that they are induced by a Lie derivative, the coefficients E^I_α are called **Lie–Euler operators**.

Given a local Lagrangian $\mathbf{L}$, its **Euler–Lagrange form** is defined as

$$\delta_{\mathrm{EL}}\mathbf{L} := E_\alpha(L)\delta u^\alpha \wedge \mathrm{Vol}\,. \tag{40.49}$$

An explicit formula for the Euler–Lagrange derivative is given by the following formula (this is just Eq. (40.48) for $|I| = 0$):

$$\delta_{\mathrm{EL}}L := \left(\frac{\partial L}{\partial u^\alpha} - D_\mu \frac{\partial L}{\partial u^\alpha_\mu} + \cdots\right)\delta u^\alpha\,. \tag{40.50}$$

The set of functions that satisfy $\delta_{\mathrm{EL}}L = 0$ is called the **shell**. The functions for which also all higher-order derivatives vanish, i.e. the elements of $\{x \in J^\infty(E) \mid \forall I : D_I\delta_{\mathrm{EL}}L(x) = 0\}$, are said to be **on-shell**.

Property 40.3.12 (Naturality). The Euler–Lagrange operators are natural operators in the following sense:

- $\delta_{\mathrm{EL}}|_p$ only depends on the germ at $p \in J^\infty(E)$.
- If $\phi : E \to E'$ is fibre-preserving, then

$$\delta_{\mathrm{EL}}((j^\infty\phi)^*\mathbf{L}) = (j^\infty\phi)^*(\delta_{\mathrm{EL}}(\mathbf{L}))\,. \tag{40.51}$$

It can be shown that δ_{EL} is the unique (up to scaling) linear, natural differential operator from $\Omega^{\dim(M),0}(J^\infty(E))$ to $\Omega^{\dim(M),1}(J^\infty(E))$.

The following property generalizes the property that the Euler–Lagrange equations remain invariant under addition of a divergence to the Lagrangian.

Property 40.3.13 (Divergences). If a smooth function is locally an order-k divergence, i.e. $f = D_I A^I$ for smooth functions A^I and $|I| = k$, the Lie–Euler operators E^J_α vanish on f for all $|J| < k$.

Property 40.3.14 (Local variational formula). Decomposition 40.3.10 shows that the Euler–Lagrange operator is the unique operator such that locally, for all local Lagrangians $\mathbf{L}$ and all evolutionary vector fields X, the following equation holds

$$\mathcal{L}_{j^\infty X}\mathbf{L} = j^\infty X \lrcorner\, \delta_{\mathrm{EL}}(\mathbf{L}) + \mathrm{d}(j^\infty X \lrcorner\, \gamma) \tag{40.52}$$

for some $\gamma \in \Omega^{\dim(M)-1,1}(J^\infty(E))$.

40.3.3 Functional complex

Example 40.3.15 (Interior Euler operator). For every smooth form $\omega \in \Omega^{p,q}(J^\infty(E))$, one can define a total differential operator as follows:

$$P_\omega(X) := j^\infty X \lrcorner\, \omega\,. \tag{40.53}$$

As for the previous example, this operator induces (higher) Euler operators:

$$F^I_\alpha(\omega) := \sum_{|J|=0}^{k-|I|} \binom{|I|+|J|}{|J|} (-1)^{|J|} D_J\big(\partial^{IJ}_\alpha \lrcorner\, \omega\big)\,. \tag{40.54}$$

Since, in this case, they arise from interior multiplication, they are called interior Euler operators. For $p = \mathsf{dim}(M)$, one again obtains a globally defined Euler operator (also called the interior Euler operator):

$$I(\omega) := \frac{1}{q}\delta u^\alpha \wedge F_\alpha(\omega)\,. \tag{40.55}$$

The interior Euler operator defines a sequence of spaces, the so-called spaces of **functional forms**, as follows:

$$\mathcal{F}^q(J^\infty(E)) := \big\{\omega \in \Omega^{\mathsf{dim}(M),q}(J^\infty(E)) \,|\, I(\omega) = \omega\big\}\,. \tag{40.56}$$

Property 40.3.16. The interior Euler operator has the following important properties:

- I is a projection $I^2 = I$.
- I vanishes on (locally) d-exact forms.
- $\delta_V := I \circ \delta$ endows $\mathcal{F}^\bullet(J^\infty(E))$ with the structure of a cochain complex. It is sometimes called the **Helmholtz operator**.
- $\delta_{\mathrm{EL}}\mathbf{L} = \delta_V \mathbf{L}$ for all local Lagrangians $\mathbf{L}$.

Corollary 40.3.17. The Euler–Lagrange operator δ_{EL} vanishes on locally d-exact forms and commutes with the Lie derivative of projectable vector fields.

The degree-2 functional forms also admit a local characterization:

Property 40.3.18 (Local expression for $\mathcal{F}^2$). Consider a collection of smooth functions $A^I_{\alpha\beta} \in C^\infty(J^\infty(U))$ on some chart $U \subset E$ that satisfy

$$A^I_{\alpha\beta} = (-1)^{|I|+1} A^I_{\beta\alpha}\,. \tag{40.57}$$

The local $(\mathsf{dim}(M), 2)$-form

$$w^{|I|} := \delta u^\alpha \wedge \big(A^I_{\alpha\beta}\delta u^\beta_I + D_I(A^I_{\beta\alpha}\delta u^\beta)\big) \wedge \mathsf{Vol} \tag{40.58}$$

is an element of $\mathcal{F}^2(J^\infty(U))$. Furthermore, every degree-2 functional form can locally be expressed as a sum of the form

$$\omega = w^0 + w^1 + w^2 + \cdots\,. \tag{40.59}$$

Although local characterizations for functional forms $\omega \in \mathcal{F}^q(J^\infty(E))$ with $q \geq 2$ are still not well understood, there exists a more high-level characterization.

Property 40.3.19 (General characterization). A form $\omega \in \Omega^{\dim(M),q}(J^\infty(E))$ is functional if and only if there exists a linear, **formally skew-adjoint**[5] differential operator

$$P : \mathsf{Ev}(J^\infty(E)) \to \Omega^{\dim(M),q-1}(J^\infty(E))$$

such that

$$\omega = \delta u^\alpha \wedge P_\alpha \,. \tag{40.60}$$

This representation is unique if it exists.

The forms in $\mathcal{F}^q$ are said to be functional due to the following property.

Property 40.3.20 (Functionals). To every smooth k-form $\omega \in \Omega^\bullet(J^\infty(E))$, compact subset $K \subset E$ and $q := k - \dim(M)$ generalized vector fields on E, one can assign a functional on $\Gamma(U)$, where U is a chart containing K, by the following formula:

$$W_\omega(X_1, \dots, X_q)[\sigma] := \int_K (j^\infty\sigma)^*\omega(j^\infty X_1, \dots, j^\infty X_q) \,. \tag{40.62}$$

The integer q is called the **degree** of W_ω. In general, these functionals are invariant under the addition of a d-exact form. However, the assignment $\omega \mapsto W_\omega$ is a bijection for fixed K and X_i.

Construction 40.3.21. The differential δ_V on $\mathcal{F}^\bullet$ induces a differential on the space of functionals of the above form:

$$\delta W_\omega := W_{\delta_V\omega} \,. \tag{40.63}$$

An equivalent definition can be given by a formula similar to Formula 34.4.11, where an evolutionary vector field X acts on a degree-0 functional by Lie derivation:

$$X(W_\omega[\sigma]) := \int_V (j^\infty\sigma)^*(\mathcal{L}_{j^\infty X}\omega) \,. \tag{40.64}$$

40.3.4 Structure of the bicomplex

A well-defined morphism of variational bicomplexes should preserve the bigrading of forms, but this will clearly not be the case in general. Therefore, a 'projected pullback' can be introduced:

$$\Phi^\sharp : \Omega^{p,q}(J^\infty(E')) \to \Omega^{p,q}(J^\infty(E)) : \omega \mapsto \pi^{p,q}(\Phi^*\omega) \,. \tag{40.65}$$

[5]This is a generalization of Property 40.2.14. P is said to be formally skew-adjoint if there exists an $(m-1, q-1)$-form K such that

$$j^\infty X \lrcorner P(Y) + j^\infty Y \lrcorner P(X) = \mathrm{d}K \tag{40.61}$$

for all $X, Y \in \mathsf{Ev}(J^\infty(E))$.

Here, the projection $\pi^{p,q}$ is the projection of the variational bicomplex, not the one of jet bundles (which is denoted by subscripts).

Remark 40.3.22. The projection $\pi^{p,q} : \Omega^{p+q}(J^\infty(E)) \to \Omega^{p,q}(J^\infty(E))$ is defined by substituting $\mathbf{d}u^\alpha_I \longrightarrow \delta u^\alpha_I + u^\alpha_{\mu I}\mathrm{d}x^\mu$ and then projecting onto the correct horizontal and vertical degrees. Note that this does not preserve the order of forms due to the presence of the factor $u^\alpha_{\mu I}$.

The main argument for introducing the projected pullback is that functionals of the form of Property 40.3.20 only care about (m, q)-forms for a specific q (in particular **action functionals**, i.e. $q = 0$).

Formula 40.3.23 (Local Lagrangians). For local Lagrangians $\mathbf{L} \equiv L\,\mathsf{Vol}_{E'}$, the projected pullback acts as follows:

$$\Phi^\sharp \mathbf{L} = (L \circ \Phi)\,\det(D_\mu f^\mu)\,\mathsf{Vol}_E\,, \tag{40.66}$$

where $\pi'(\Phi[u]) = (f^\mu)$. So, one obtains the usual formula for pullbacks of top-dimensional forms, but with partial derivatives replaced by total derivatives.

An important property of the de Rham differential is its naturality (Property 34.4.9). The following property states the 'naturality' of the different operators on the variational bicomplex with respect to the projected pullback.

Property 40.3.24. Consider a morphism $\Phi : J^\infty(E) \to J^\infty(E')$.

- $\Phi^\sharp$ and δ commute if and only if Φ covers a morphism of the base manifolds.
- $\Phi^\sharp$ and d commute if and only if Φ^* is a contact transformation.
- $\Phi^\sharp$ commutes with both differentials if and only if it coincides with Φ^*

Furthermore, the projected pullback defines a contravariant functor on the subcategory on morphisms that satisfy at least one of the above properties.

The following property gives an infinitesimal analogue of the above considerations.

Property 40.3.25. Consider a vector field X on $J^\infty(E)$. Its Lie derivative will, in general, not respect the bigrading of the variational bicomplex (unless X is evolutionary) and, therefore, the 'projected Lie derivative' is introduced:

$$\mathcal{L}^\sharp_X : \Omega^{p,q}(J^\infty(E)) \to \Omega^{p,q}(J^\infty(E)) : \omega \mapsto \pi^{p,q}(\mathcal{L}_X\omega)\,. \tag{40.67}$$

This operator satisfies the following properties:

- It commutes with δ if and only if X is π_∞-related to a vector field on M.
- It commutes with d if and only if X is the prolongation of a generalized vector field on E.

Note that this does not simply follow from the previous property since X does not necessarily define a flow on $J^\infty(E)$.

Aside from the differentials on the variational bicomplex, one should also look at the structure of the functional complex $(\mathcal{F}^\bullet, \delta_V)$ in (40.56). It can be shown that requiring both $[I, \Phi^\sharp] = 0$ and $[\delta, \Phi^\sharp] = 0$ is very restrictive. Furthermore, a complete characterization of those morphisms Φ that satisfy only $[I, \Phi^\sharp] = 0$ is not fully understood. However, the infinitesimal version is easier to handle since it only involves linear equations:

Property 40.3.26. Let $n \in \mathbb{N}$ be the rank of E and consider a vector field X on $J^\infty(E)$.

- If $n = 1$, then $\mathcal{L}_X^\sharp$ commutes with I if and only if X is locally the prolongation of a generalized vector field on E of the form

$$Y = -\frac{\partial S}{\partial u_\mu}\partial_\mu + \left(S - u_\mu \frac{\partial S}{\partial u_\mu}\right)\partial_u\,, \tag{40.68}$$

  where S is a function on the first jet bundle $J^1(U)$.

- If $n > 1$, then $\mathcal{L}_X^\sharp$ commutes with I if and only if X is the prolongation of a vector field on E.

The next step is to define operators that do preserve the functional complex. To this end, the projection property of I is used:

$$\begin{aligned} \Phi^\natural &: \Omega^{m,q}(J^\infty(E)) \to \mathcal{F}^q(J^\infty(E')) : \omega \mapsto (I \circ \Phi^\sharp)\omega\,, \\ \mathcal{L}_X^\natural &: \Omega^{m,q}(J^\infty(E)) \to \mathcal{F}^q(J^\infty(E)) : \omega \mapsto (I \circ \mathcal{L}_X^\sharp)\omega\,. \end{aligned} \tag{40.69}$$

Property 40.3.27. These operators satisfy the following properties:

- If Φ is a contact transformation, then $\Phi^\natural$ commutes with both I and δ_V.
- If X is a generalized vector field on E, then $\mathcal{L}_{j^\infty X}^\natural$ commutes with both I and δ_V.
- $\Phi^\natural$ preserves locally variational forms.

The projected and functionally projected operators also satisfy the following relations.

Property 40.3.28 (Euler–Lagrange operator). Consider a local Lagrangian $\mathbf{L}$. If Φ is a contact transformation, then

$$\delta_{\mathrm{EL}}(\Phi^\sharp \mathbf{L}) = \Phi^\natural(\delta_{\mathrm{EL}}\mathbf{L})\,. \tag{40.70}$$

If X is a generalized vector field, then

$$\delta_{\mathrm{EL}}(\mathcal{L}_{j^\infty X}^\sharp \mathbf{L}) = \mathcal{L}_{j^\infty X}^\natural(\delta_{\mathrm{EL}}\mathbf{L})\,. \tag{40.71}$$

Formula 40.3.29 (Functionally projected Lie derivative). Let X be a generalized vector field on E and let $\omega \in \mathcal{F}^\bullet(J^\infty(E))$ be a functional form.

$$\mathcal{L}^\natural_{j^\infty X}\omega = \delta_V(j^\infty X_{\mathsf{ev}} \rightharpoonup \omega) + I(j^\infty X_{\mathsf{ev}} \rightharpoonup \delta_V\omega) \tag{40.72}$$

For projectable vector fields, one can replace the left-hand side by the ordinary Lie derivative $\mathcal{L}_{j^\infty X}\omega$.

In the remainder of this section, the homological properties of the variational bicomplex over a local chart (or equivalently, over a trivial bundle) will be studied. To prove the acyclicity of the various subcomplexes, the usual approach of finding a null-homotopy (Corollary 5.1.7) will be followed, i.e. a cochain map will be found $h : C_\bullet \to C_\bullet$ such that

$$\mathbb{1} = \mathsf{d} \circ h + h \circ \mathsf{d}\,. \tag{40.73}$$

Property 40.3.30 (Vertical complex is exact). Homotopy operators $h_V^{p,q} : \Omega^{p,q} \to \Omega^{p,q-1}$ for the vertical complex

$$0 \longrightarrow \Omega^p(M) \xrightarrow{\pi^*_\infty} \Omega^{p,0} \xrightarrow{\delta} \Omega^{p,1} \xrightarrow{\delta} \cdots \tag{40.74}$$

are given by the following formula

$$h_V^{p,q}(\omega) = \int_0^1 \frac{1}{t}\Phi^*_{\log t}(j^\infty R \rightharpoonup \omega)\,dt\,, \tag{40.75}$$

where $R := u^\alpha\partial_\alpha$ is the (vertically) radial vector field and $\Phi_\varepsilon : (x,u) \mapsto (x, e^\varepsilon u)$. It is not too difficult to check that, for source forms, this gives rise to Eq. (40.95).

The analogous statement for the horizontal complex is a bit more involved.

Property 40.3.31 (Augmented horizontal complex is exact). Homotopy operators $h_H^{p,q} : \Omega^{p,q} \to \Omega^{p-1,q}$ for the augmented horizontal complex

$$0 \longrightarrow \Omega^{0,q} \xrightarrow{\mathsf{d}} \Omega^{1,q} \xrightarrow{\mathsf{d}} \cdots \xrightarrow{\mathsf{d}} \Omega^{\dim(M),q} \xrightarrow{I} \mathcal{F}^q \xrightarrow{I} 0 \tag{40.76}$$

are given by the following formula

$$h_H^{p,q}(\omega) = \frac{1}{q}\sum_{|I|=0}^{k-1} \frac{|I|+1}{m-p+|I|+1} D_I\big(\delta u^\alpha \wedge F^{I\mu}_\alpha(D_\mu \rightharpoonup \omega)\big)\,, \tag{40.77}$$

where the F^I_α are the interior Euler operators (40.54).

Corollary 40.3.32 (Functional decomposition). The de Rham spaces on $J^\infty(E)$ admit the following decomposition:

$$\Omega^{p,q}(J^\infty(E)) = \mathsf{d}\Omega^{p-1,q}(J^\infty(E)) \oplus \mathcal{F}^q(J^\infty(E))\,, \tag{40.78}$$

where the functional part is obtained by applying I to a form.

Using the above properties, one can prove the acyclicity of the **Euler–Lagrange complex** $\mathcal{E}$ (again, over a local chart):

$$0 \longrightarrow \mathbb{R} \longrightarrow \Omega^{0,0} \xrightarrow{\mathsf{d}} \Omega^{1,0} \xrightarrow{\mathsf{d}} \cdots \xrightarrow{\mathsf{d}} \Omega^{\dim(M),0} \xrightarrow{\delta_{\mathrm{EL}}} \mathcal{F}^1 \xrightarrow{\delta_V} \mathcal{F}^2 \xrightarrow{\delta_V} \cdots . \tag{40.79}$$

Explicit formulas are not shown since these are too complicated for the current objective. See Anderson (1992) for a full account.

Remark 40.3.33 (Minimal solutions). Although the (local) exactness of the variational bicomplex is stated, it should be noted that this is not an optimal solution. Consider the example of locally variational source forms (see the next two sections), i.e. differential forms of the form $\delta_{\mathrm{EL}}\mathbf{L}$. From the form of the homotopy operator $\mathcal{F}^1 \to \Omega^{m,0}$, it should be clear that an order-k source form is mapped to an order-k Lagrangian. However, the Euler–Lagrange operator δ_{EL} will, in general, map order-l Lagrangians to order-$2l$ source forms. Hence, it can be seen that the homotopy operator will, in general, not give Lagrangians of minimal order.

40.3.5 Variational problems

The forms in $\mathcal{F}^1$ admit the following characterization.

Definition 40.3.34 (Source form). A differential form $\omega \in \Omega^{\dim(M),1}(J^\infty(E))$ such that the evaluation on a vector field only depends on the projection $(\pi_{\infty,0})_* X \in TE$. After pulling back along the prolongation map j^∞, this can be written as follows:

$$\Omega^{\dim(M),1}_{\text{source}}(E) := \delta C^\infty(E) \wedge \Omega^{\dim(M),0}(E) . \tag{40.80}$$

Locally, it can be written as

$$\omega = \omega_\alpha(x, u, u_I)\delta u^\alpha \wedge \mathsf{Vol} . \tag{40.81}$$

The PDEs associated to the source form ω are called **source equations**.

Remark. In fact, one can extend the above definition to all of $\Omega^{\bullet,1}(J^\infty(E))$. So, in general, $\mathcal{F}^1$ is only a subspace of the space of source forms.

Property 40.3.35 (First variational formula). In Property 40.3.14, it was shown that the Euler–Lagrange operator locally satisfies

$$\mathcal{L}_{j^\infty X}\mathbf{L} = j^\infty X \lrcorner\, \delta_{\mathrm{EL}}\mathbf{L} + \mathsf{d}(j^\infty X \lrcorner\, \gamma) \tag{40.82}$$

for some locally defined form γ that can be constructed from the Lie–Euler operators of $\mathbf{L}$. Using Cartan's magic formula and the fact that $\mathbf{L}$ is horizontal, this can be rewritten in terms of differentials. Moreover, using Corollary 40.3.32, one can find a globally defined form that still satisfies the same formula (with the disadvantage that it does not admit a canonical expression):

$$\delta\mathbf{L} = \delta_{\mathrm{EL}}\mathbf{L} + \mathsf{d}\sigma . \tag{40.83}$$

This formula is often called the first variational formula for the following reason. Given a local Lagrangian $\mathbf{L}$, one can look at solutions of the associated variational principle obtained by extremizing over perturbations of a field configuration. Such perturbations are generated by evolutionary vector fields and, hence, one can write the extremality condition as

$$\forall X \in \mathsf{Ev}(J^\infty(E)) : \mathcal{L}_{j^\infty X}\mathbf{L} = 0\,. \tag{40.84}$$

Equation (40.83) then becomes:

$$\delta_{\mathsf{EL}}\mathbf{L} + \mathsf{d}\sigma = 0\,. \tag{40.85}$$

So, up to boundary contributions, the first variational formula says that extremality of the local Lagrangian (globally) corresponds to the vanishing of the Euler–Lagrange operator.

Corollary 40.3.36 (Global variational formula). The variational formula can also be extended to all generalized vector fields:

$$\mathcal{L}^\sharp_{j^\infty X}\mathbf{L} = X_{\mathrm{ev}} \lrcorner\, \delta_{\mathrm{EL}}(\mathbf{L}) + \mathsf{d}\sigma \tag{40.86}$$

for some $\sigma \in \Omega^{\dim(M),0}(J^\infty(E))$.

Definition 40.3.37 (Lepage form). A $\dim(M)$-form $\rho \in \Omega^\bullet(J^\infty(E))$ such that

$$\pi^{\dim(M),0}(X \lrcorner\, \mathbf{d}\rho) = 0 \tag{40.87}$$

for all $\pi_{\infty,0}$-vertical vector fields X. Given a local Lagrangian $\mathbf{L} \in \Omega^{\dim(M),0}(J^\infty(E))$, $\mathbf{L}$ and ρ are said to be Lepage equivalent if $\pi^{\dim(M),0}\rho = \mathbf{L}$.

Property 40.3.38 (Lepage equivalent). Consider the first variational formula for a local Lagrangian $\mathbf{L}$. A Lepage equivalent is given by the form $\mathbf{L} + \gamma$.

Definition 40.3.39 (Distinguished symmetry). A distinguished (generalized) symmetry of a source form $\Delta \in \mathcal{F}^1(J^\infty(E))$ is a generalized vector field X on E such that

$$\mathcal{L}^\natural_{j^\infty X}\Delta = 0\,. \tag{40.88}$$

As with the formula above, one can replace the projected Lie derivative by an ordinary Lie derivative when X is projectable.

If the source form comes from a local Lagrangian, the above definition admits a specific case by Property 40.3.28 and the fact that δ_{EL} annihilates d-exact forms (Corollary 40.3.17).

Definition 40.3.40 (Bessel-Hagen symmetry). A Bessel-Hagen or **divergence** symmetry of a Euler–Lagrange form $\delta_{\mathrm{EL}}(\mathbf{L})$ is a generalized vector field X such that

$$\mathcal{L}^\sharp_{j^\infty X}\mathbf{L} = \mathsf{d}\eta \tag{40.89}$$

for some $\eta \in \Omega^{\dim(M),0}(J^\infty(E))$.

If this condition holds locally, distinguished symmetries and Bessel-Hagen symmetries coincide. However, if the Bessel-Hagen condition is required to hold globally, the Bessel-Hagen symmetries form only a subset of the distinguished symmetries.

Definition 40.3.41 (Local conservation law). A generator of a local conservation law of a source form $\Delta \in \mathcal{F}^1(J^\infty(E))$ is an evolutionary vector field X such that

$$\delta_{\mathrm{EL}}(j^\infty X \lrcorner \Delta) = 0 \tag{40.90}$$

or, again by 40.3.17,

$$X \lrcorner \Delta = \mathrm{d}\eta \tag{40.91}$$

for some local $\eta \in \Omega^{\dim(M),0}(J^\infty(E))$. Corollary 40.3.36 shows that to every global (generalized) symmetry, there corresponds a global conservation law.

Remark 40.3.42. Note that the Bessel-Hagen/divergence symmetries are symmetries of the source form, while the local symmetries coming from the first variational formula exist on the level of local Lagrangians.

Property 40.3.43 (Lie algebra of symmetries). Given a source form $\Delta \in \mathcal{F}^1(J^\infty(E))$, one can equip the vector space of generalized vector fields satisfying the following two conditions with the structure of a Lie algebra:

- They are distinguished symmetries.
- Their evolutionary part is a generator of local conservation laws.

Theorem 40.3.44 (Noether). *If $\Delta \in \mathcal{F}^1(J^\infty(E))$ is locally variational, a generalized vector field on E is a distinguished symmetry if and only if its evolutionary part is a generator of local conservation laws.*

40.3.6 Inverse problem

The inverse problem in the calculus of variations consists of determining when a given system of PDEs can be obtained from a variational problem, i.e. when a source form Δ can be written in the form $\delta_{\mathrm{EL}}\mathbf{L}$, these are said to be **locally variational**.

Helmholtz was the first to study the inverse problem, so the following sufficient conditions are named after him.

Property 40.3.45 (Helmholtz conditions). By applying the Euler–Lagrange operator δ_{EL} to its defining variational formula (Property 40.3.14) and using the (infinitesimal) naturality condition (40.51) and the fact that it annihilates d-exact forms (Corollary 40.3.17), it can be seen that a source form Δ can be obtained from a local Lagrangian if

$$\mathcal{L}_{j^\infty X}\Delta = \delta_{\mathrm{EL}}(X \lrcorner \Delta) \tag{40.92}$$

is satisfied for all evolutionary vector fields X. This can also be rewritten in terms of the Helmholtz operator δ_V (hence its name):[6]

$$\delta_V \Delta = 0\,. \tag{40.93}$$

Formula 40.3.46 (Local expression). Consider a source form that admits the local expression

$$\Delta = P_\alpha \delta u^\alpha \wedge \mathsf{Vol}\ .$$

The Helmholtz conditions can locally be expressed as follows:

$$(-1)^{|I|}\partial^I_\alpha P_\beta = E^I_\beta(P_\alpha)\,, \tag{40.94}$$

where the E^I_β are the Lie–Euler operators (40.48).

Example 40.3.47 (Volterra–Vainberg formula). If E is trivial and $\Delta = F_\alpha \delta u^\alpha \wedge \mathsf{Vol}$ satisfies the Helmholtz conditions, then

$$L := \int_0^1 u^\alpha F_\alpha[tu]\,dt \tag{40.95}$$

satisfies $\Delta = \delta_{\mathsf{EL}}\mathbf{L}$. If Δ is homogeneous of degree $k \in \mathbb{N}$ in u, this can be expressed as

$$\mathbf{L} = \frac{1}{k+1}\iota_R \Delta\,, \tag{40.96}$$

where $R := u^\alpha \partial_\alpha$ is the (vertically) radial vector field on E.

40.3.7 Finite jet bundles

A last subject that will be considered in this section is the restriction of the variational bicomplex to finite jet bundles. However, as is clear from the definition of the horizontal differential, forms of order $k \in \mathbb{N}$ are mapped to forms of order $k+1$. Therefore, attention will have to be restricted to a specific subcomplex of $\Omega^\bullet(J^\infty(E))$:

$$\Omega^\bullet_k(E) \subset \Omega^\bullet(J^{k+1}(E)) := \delta\text{-closure of } \Omega^\bullet(J^k(E))\,. \tag{40.97}$$

From the basic definitions and properties of the two differentials, it follows that $\Omega^\bullet_k$ is generated by $C^\infty(J^k(E))$, the horizontal basis $\{\mathrm{d}x^\mu\}_{\mu\leq\dim(M)}$ and the contact basis $\{\delta u^\alpha_I\}^{|I|\leq k}_{\mu\leq\dim(M)}$. The next step is to further restrict to a horizontally closed subcomplex. To this end, consider forms $\omega \in \Omega^{p,q}_k(E)$ of the form

$$\omega = \left[\mathrm{d}u^{\alpha_1}_{I_1} \wedge \ldots \wedge \mathrm{d}u^{\alpha_r}_{I_r} \wedge \mathrm{d}\delta u^{\beta_1}_{J_1} \wedge \ldots \wedge \mathrm{d}\delta u^{\beta_s}_{J_s}\right] \wedge f\mathrm{d}x^{\kappa_1} \wedge \ldots \wedge \mathrm{d}x^{\kappa_{p-r-s}}\,, \tag{40.98}$$

[6]This expression also immediately follows from Property 40.3.16.

where $|I_i| = |J_i| = k-1$ and $f \in C^\infty(J^{k-1}(E))$. It is immediately clear that the subcomplex of such forms is also horizontally closed. The factor in between square brackets can also be rewritten as follows:

$$J = \frac{1}{(r+s)!}\frac{D(u^{\alpha_1}_{I_1}, \dots, u^{\alpha_r}_{I_r}, \delta u^{\beta_1}_{J_1}, \dots, \delta u^{\beta_s}_{J_s})}{D(x^{\mu_1}, \dots, x^{\mu_r}, x^{\nu_1}, \dots, x^{\nu_s})} \mathrm{d}x^{\mu_1} \wedge \dots \wedge \mathrm{d}x^{\mu_r} \wedge \mathrm{d}x^{\nu_1} \wedge \dots \wedge \mathrm{d}x^{\nu_s} . \tag{40.99}$$

This factor has the form of a Jacobian determinant (with respect to total derivatives) and, as such, the subcomplex spanned by the above forms is called the **Jacobian complex** $\mathcal{J}^\bullet_k(E)$.

Property 40.3.48 (Alternative characterizations). The Jacobian complex can also be characterized as follows:

- Consider the projection $\pi^{\bullet,0} : \Omega^\bullet \to \Omega^{\bullet,0}$ (note that this maps forms in $\Omega^r(J^k(E))$ to forms in $\Omega^{r,0}_{k+1}(E)$ due to Remark 40.3.22).

$$\mathcal{J}^{p,q}_k(E) = \Omega^{p,q}(J^\infty(E)) \cap \delta\text{-closure of } \pi^{\bullet,0}\big(\Omega^\bullet(J^{k-1}(E))\big) . \tag{40.100}$$

- For $p < m$, the Jacobian complex contains those forms for which d does not increase the order:

$$\mathcal{J}^{p,q}_k(E) = \{\omega \in \Omega^{p,q}_k \mid \mathrm{d}\omega \in \Omega^{p+1,q}_k(E)\} . \tag{40.101}$$

- If $\omega \in \mathcal{J}^{p,q}_k(E)$, then ω is a polynomial in u^α_i's and δu^α_I's of degree $\leq r$ with $|I| = k$.

It can be shown that the Jacobian subcomplex is (locally) exact and that it respects the structure of the Euler–Lagrange complex.

Property 40.3.49 (Exactness). Let E be trivial. If $\delta_{\mathrm{EL}}\omega = 0$ for $\omega \in \Omega^{m,q=0}_k(E)$ or $I(\omega) = 0$ for $\omega \in \Omega^{m,q\geq 1}_k(E)$, then $\omega \in \mathcal{J}^{m,q}_k(E)$ and $\omega = \mathrm{d}\eta$ for $\eta \in \mathcal{J}^{m-1,q}_k(E)$.

Property 40.3.50 (Functional dependence of Lagrangians). If Δ is a locally variational source form of order $k \in \mathbb{N}$, it is polynomial of degree m in k^{th}-order derivatives of the u^α.

40.4 Partial differential equations

In this section, the content of Chapter 21 is generalized using the language of differential geometry.

40.4.1 Algebraic formulation

In this section, partial differential equations of the form

$$f(x, u, u_I) = 0 \tag{40.102}$$

are considered, where f is a smooth function. A partial differential equation is regarded as an algebraic equation involving derivatives and, hence, f can be interpreted as a function on the jet bundle $J^k(\mathbb{R}^m)$, where $k \in \mathbb{N}$ is the order of the PDE and $m \in \mathbb{N}$ is the number of independent variables.

In this framework, one can define a solution of the above PDE as a function ϕ satisfying $f \circ j^k\phi = 0$. This can be rephrased in a geometric way. Every PDE of order $k \in \mathbb{N}$ defines a subbundle Σ^0 of the finite jet bundle $J^k(E)$ and a solution is simply a section of Σ^0.

Remark 40.4.1. For every $l \in \mathbb{N}$, define the subspace

$$\Sigma^l := \{(x, u, u_I) \in J^{k+l}(E) \mid \forall |J| \leq l : D_J f(x, u, u_I) = 0\} = J^l(\Sigma^0) \cap J^{k+l}(E)\,, \tag{40.103}$$

i.e. the set of holonomic sections of $J^{k+l}(E)$ that are l^{th}-order tangent to Σ^0. A function $\phi \in \Gamma(E)$ is a solution if there exists some $l \in \mathbb{N}$ such that the image of $j^{k+l}\phi$ lies in Σ^l and, conversely, $j^{k+l}\phi$ lies in Σ^l for all $l \in \mathbb{N}$ if ϕ is a solution.

Definition 40.4.2 (Formal integrability). A PDE $f : \Sigma^0 \hookrightarrow J^k(E)$ is called formally integrable if

1. all prolongations Σ^l are smooth manifolds.
2. all projections $\Sigma^{l+1} \to \Sigma^l$ are smooth fibre bundles.

Property 40.4.3 (Diffiety). The leaves of the Cartan distribution on $J^\infty(E)$ are the graphs of infinity-prolongations $j^\infty\phi$ for local sections $\phi \in \Gamma(E)$. When restricting the distribution to the PDE Σ^0, the leaves are given by the graphs of the infinity-prolongations of (local) solutions.

A pair $(\Sigma, \mathcal{C}(\Sigma))$ consisting of a smooth manifold Σ and a finite-dimensional distribution $\mathcal{C}(\Sigma)$, such that Σ is locally isomorphic to the infinity-prolongation of a PDE and $\mathcal{C}(\Sigma)$ is locally isomorphic to the associated Cartan distribution, is called a **diffiety** (short for **differential variety**).

Remark 40.4.4. The reason for this terminology stems from the apparent similarity with algebraic varieties. A variety is (locally) defined by a set of algebraic equations together with all algebraic consequences, i.e. it is defined by the ideal generated by the algebraic equations. Similarly, a diffiety is (locally) defined by a set of differential equations together with all differential consequences, i.e. it is defined by the differential ideal generated by the differential equations.

Definition 40.4.5 (Exterior system). Consider a general PDE of the form (40.102). This PDE can be turned into a set of differential forms on the zero locus of f, i.e. Σ_0:

$$\theta := \mathrm{d}u - u_\mu \mathrm{d}x^\mu\,, \tag{40.104}$$

$$\vdots$$

These differential one-forms generate an ideal in $\Omega^\bullet(\Sigma^0)$ that represents the differential equation in that it relates the variables in the algebraic condition $f = 0$ by a set of differential relations. It is a Pfaffian system (Property 34.6.8) or, combined with the PDE $f = 0$ and its differential consequences, it gives a characteristic system (Definition 34.6.10). The associated distribution becomes integrable (sometimes called **completely integrable**) if and only if the ideal is a differential ideal by Frobrenius's theorem 34.6.7. The integral manifolds of the distribution then give a solution of the PDE.

Formula 40.4.6 (Lagrange–Charpit equations). Let $f = 0$ be a first-order PDE on a smooth manifold and consider the differential closure of the exterior system of $f = 0$:

$$\begin{cases} f &= 0\,, \\ \mathrm{d}f &= \left(\frac{\partial f}{\partial x^\mu} + u_\mu \frac{\partial f}{\partial u}\right)\mathrm{d}x^\mu + \frac{\partial f}{\partial u_\mu}\mathrm{d}u_\mu = 0\,, \\ \theta &= 0\,, \\ \mathrm{d}\theta &= 0\,. \end{cases} \tag{40.105}$$

The characteristic system of these equations is given by the PDE itself, together with the Lagrange–Charpit equations:

$$\begin{aligned} \frac{\mathrm{d}x^1}{\partial f/\partial u_1} = \cdots = \frac{\mathrm{d}x^n}{\partial f/\partial u_n} &= \frac{-\mathrm{d}u_1}{\partial f/\partial x^1 + u_1(\partial f/\partial u)} \\ &= \cdots = \frac{\mathrm{d}u_n}{\partial f/\partial x^n + u_n(\partial f/\partial u)} = \frac{\mathrm{d}u}{\sum_\mu u_\mu(\partial f/\partial u_\mu)}\,. \end{aligned} \tag{40.106}$$

Definition 40.4.7 (Monge cone). Consider the characteristic system above. The rank of the Pfaffian system is $2\dim(M)$. At every point $(x^\mu, u) \in M \times \mathbb{R}$, the solutions (x^μ, u, u_μ) of the PDE determine a tangent plane to M. The $\dim(M)$-parameter family of tangent planes obtained by varying the derivative coordinates u_μ gives rise to the Monge cone with apex (x^μ, u). A solution to the PDE is a hypersurface that is everywhere tangent to the Monge cones. This approach to finding a solution is called the **method of the characteristics**.

40.4.2 Symmetries

Property 40.4.8. A Lie group G with Lie algebra $\mathfrak{g}$ is the symmetry group of a (nondegenerate) system F of PDEs if and only if

$$F = 0 \implies j^\infty X(F) = 0 \tag{40.107}$$

for all generators $X \in \mathfrak{g}$.

@@ COMPLETE @@

40.4.3 Pseudogroups ♣

Example 40.4.9 (Diffeomorphism jet). Let M be a smooth manifold. Consider the set $\mathcal{D}^\omega(M)$ of local analytic diffeomorphisms $\phi : M \to M : z \mapsto \phi(z)$. The locality property turns this set into a (smooth) pseudogroup (Definition 6.2.21).

By the Inverse Function Theorem 34.1.10, one can define the diffeomorphism jet bundle $\mathcal{D}^r(M)$ as the subbundle of $J^r(M, M)$ for which

$$\det\left(\frac{\partial \phi^\alpha}{\partial z^\beta}\right) \neq 0\,.$$

It is also possible to endow this jet bundle with the structure of a groupoid (Definition 3.5.9). Using the source and target projections, one can check that two elements $f, g \in \mathcal{D}^r(M)$ can be multiplied if and only if $\pi_r(g) = \pi_{r,0}(f)$. The derivative coordinates can be found using the *Faá di Bruno formula*.

Furthermore, every pseudogroup $\mathcal{G} \subset \mathcal{D}^\omega$ induces a subbundle $\mathcal{G}^{(r)} \subset \mathcal{D}^{(r)}$. This structure gives rise to the following notions.

Definition 40.4.10 (Regular pseudogroup). Consider a smooth manifold M. Let $\mathcal{D}^\omega(M)$ be its diffeomorphism pseudogroup and let $\mathcal{G} \subset \mathcal{D}^\omega$ be another pseudogroup. If there exists a $k \in \mathbb{N}_0$, called the **order**, such that, for all $n \geq k$, the jets $\pi_n : \mathcal{G}^{(n)} \to M$ form an embedded submanifold of $\Pi_n : \mathcal{D}^{(n)} \to M$ and such that the jet projections $\pi_{n+1,n} : \mathcal{G}^{(n+1)} \to \mathcal{G}^{(n)}$ are fibrations, then $\mathcal{G}$ is called a regular pseudogroup.

Definition 40.4.11 (Lie pseudogroup). Let $\mathcal{G} \subset \mathcal{D}^\omega$ be a regular analytic pseudogroup of order $k \in \mathbb{N}$. If every local diffeomorphism $\phi \in \mathcal{D}^\omega$ satisfying $j^k\phi \in \mathcal{G}^{(k)}$ is also an element of $\mathcal{G}$, then $\mathcal{G}$ is called a Lie pseudogroup.

Property 40.4.12. Let $\mathcal{G}$ be a Lie pseudogroup of order $k \in \mathbb{N}$. The regularity condition implies that for all $n \geq k$ the jet bundle $\mathcal{G}^{(n)}$ is described by a set of nth-order PDEs

$$F\big(z, \phi^{(n)}\big) = 0\,. \tag{40.108}$$

The (local) solutions to these equations are exactly the analytic functions that have $(z_0, \phi_0^{(n)})$ as local coordinates of their n-jet at $z_0 \in M$.

The Lie condition on $\mathcal{G}$ implies that every solution to the system is, in fact, an element of $\mathcal{G}$. This system of equations is called the **determining system** of the Lie pseudogroup.

Definition 40.4.13 (Lie completion). Let $\mathcal{G}$ be a regular pseudogroup. The Lie completion $\overline{\mathcal{G}}$ of $\mathcal{G}$ is defined as the set of all (local) analytic diffeomorphisms solving the determining system of $\mathcal{G}$. This completion is itself a Lie pseudogroup. If $\mathcal{H}$ is a Lie pseudogroup, then $\overline{\mathcal{H}} = \mathcal{H}$.

@@ COMPLETE (IS THIS EVEN USEFUL?) @@

Chapter 41

K-theory ♣

The main reference for this chapter is Karoubi (1978).

In this chapter, all topological spaces are supposed to be both compact and Hausdorff, unless stated otherwise. This ensures that the complex of K-theories satisfies the Eilenberg–Steenrod axioms (Definition 7.5.1). In general, all statements will be given for an arbitrary base field $\mathfrak{K}$. When it is necessary to restrict to the specific cases $\mathfrak{K} = \mathbb{R}, \mathbb{C}$, this will be stated explicitly.

Chapter Contents

41.1 Preliminaries **779**

41.2 Topological K-theory **779**

41.2.1 Introduction 779

41.2.2 Classification 780

41.2.3 Negative degree 782

41.2.4 Bott periodicity 783

41.2.5 Clifford modules 785

41.2.6 Cohomology theory 786

41.2.7 Applications 786

41.3 Twisted K-theory **788**

41.4 Differential K-theory **788**

41.5 Algebraic K-theory **789**

41.5.1 Determinant 789

41.6 Operator K-theory **790**

41.1 Preliminaries

Definition 41.1.1 (Stable general linear group). Let R be a (unital) ring. For every two integers $m < n$, there exists a canonical inclusion $\mathsf{GL}(m, R) \hookrightarrow \mathsf{GL}(n, R)$ through extension (direct sum) by $\mathbb{1}_{n-m}$. This allows to define the stable general linear group as the direct limit

$$\mathsf{GL}(R) := \varinjlim_{n \in \mathbb{N}} \mathsf{GL}(n, R) . \tag{41.1}$$

Remark 41.1.2. A similar construction leads to the stable orthogonal and stable unitary groups O and U for $R = \mathbb{R}$ and $R = \mathbb{C}$, respectively. Sometimes, the notations $\mathsf{O}(\infty)$ and $\mathsf{U}(\infty)$ are also used.

41.2 Topological K-theory

41.2.1 Introduction

Definition 41.2.1 (K-theory). Let $\mathsf{Vect}(X)/\sim$ be the set of isomorphism classes of finite-dimensional vector bundles over a base space X. Because this set is well-behaved with respect to Whitney sums, the pair $(\mathsf{Vect}(X)/\sim, \oplus)$ is an Abelian monoid. The Grothendieck completion (Construction 2.2.6) of this monoid is called the (real) K-theory of X.

Notation 41.2.2. The K-theory of a space X is denoted by $K^0(X)$. If one wants to emphasize the base field $\mathfrak{K}$, the notation $K^0_{\mathfrak{K}}(X)$ is sometimes used, e.g. for complex vector bundles, one writes $K^0_{\mathbb{C}}(X)$. However, because the real and complex numbers are the most important choices, these two K-theories are simply denoted by $KO(X)$ and $K(X)$, respectively.

Example 41.2.3 (Point). Let $*$ be the one-point space. The K-theory $K^0(*)$ is isomorphic to the additive group of integers $(\mathbb{Z}, +)$.

Definition 41.2.4 (Virtual vector bundle). The elements of $K^0(X)$ are pairs $([E], [E'])$ that can formally be written as a difference $[E] - [E']$ of (isomorphism classes of) vector bundles. Such pairs are called virtual (vector) bundles.

Definition 41.2.5 (Virtual rank). The virtual rank of the virtual bundle $([E], [E'])$ is defined as follows:

$$\mathsf{rk}([E], [E']) := \mathsf{rk}(E) - \mathsf{rk}(E') . \tag{41.2}$$

Property 41.2.6. Property 34.2.15 implies that every virtual bundle has a representative of the form $[E] - [X \times \mathbb{R}^n]$ for some vector bundle E and integer $n \in \mathbb{N}$.

Definition 41.2.7 (Reduced K-theory). Let (X, x_0) be a pointed space. The inclusion $\{x_0\} \hookrightarrow X$ induces a group morphism $\rho : K^0(X) \to K^0(x_0)$ given by the restriction of virtual bundles to the basepoint x_0. The reduced K-theory $\widetilde{K}^0(X)$ is defined as $\ker(\rho)$.

Alternative Definition 41.2.8. The reduced K-theory $\widetilde{K}^0(X)$ can equivalently be defined as follows. Consider the stable isomorphism classes (Definition 34.2.16) of vector bundles over X. Under Whitney sums these form a commutative group $(\mathsf{Vect}(X)/\sim_{\text{stable}}, \oplus)$, which is (naturally) isomorphic to $\widetilde{K}^0(X)$.

The following construction is very similar to Construction 27.3.7. In fact, it is the one obtained for the Banach functor that restricts vector bundles to subspaces.

Definition 41.2.9 (Relative K-theory). Consider a space X and a closed subspace Y. Let $\mathscr{V}(X, Y)$ denote the set of triples (E, E', f) where E, E' are vector bundles over X and f is an isomorphism between the restrictions $E|_Y$ and $E'|_Y$. Elements in $\mathscr{V}(X, Y)$ are deemed isomorphic if there exist isomorphisms of vector bundles that make the 'obvious' diagram commute. The sum of such triples is defined elementwise. Let $\mathscr{E}(X, Y)$ denote the subset of $\mathscr{V}(X, Y)$ consisting of triples (F, F', g) where $F = F'$ and g is homotopic to $\mathbb{1}_{F|_Y}$ in $\mathsf{Aut}(F|_Y)$.

The relative K-theory $K^0(X, Y)$ is defined as the quotient of $\mathscr{V}(X, Y)$ by the following equivalence relation:

$$x \sim x' \iff \exists e, e' \in \mathscr{E}(X, Y) : x + e \cong_{\mathscr{V}} x' + e' . \tag{41.3}$$

Elements of relative K-theory are pairs of vector bundles over X that coincide on the subspace Y modulo a relation akin to that in the Grothendieck construction. It should also be clear that choosing $Y = \emptyset$ gives exactly the Grothendieck construction and, hence, $K^0(X, \emptyset) = K^0(X)$.

Property 41.2.10. Consider a space X with a closed subspace Y.

$$K^0(X, Y) \cong \ker\big(K^0(X) \to K^0(Y)\big) \tag{41.4}$$

Property 41.2.11 (Excision). Consider a space X together with a closed subspace Y. The relative K-theory is related to the reduced K-theory as follows:

$$K^0(X, Y) \cong \widetilde{K}^0(X/Y) . \tag{41.5}$$

41.2.2 Classification

Property 41.2.12 (Orthogonal groups). Consider the classifying space (Definition 35.2.1) of the orthogonal group $\mathsf{O}(n)$ and recall the Grassmannian $\mathsf{Gr}(n, \mathbb{R}^N)$ of n-dimensional subspaces in $\mathbb{R}^k$. There exists a canonical inclusion of Grassmannians:

$$\iota_k : \mathsf{Gr}(n, \mathbb{R}^k) \hookrightarrow \mathsf{Gr}(n, \mathbb{R}^{k+1}) : W \mapsto W . \tag{41.6}$$

By taking the direct limit of these inclusions, one obtains the infinite Grassmannian:

$$BO(n) \cong \varinjlim_{k\in\mathbb{N}} \mathsf{Gr}(n, \mathbb{R}^k)\,. \tag{41.7}$$

As the notation implies, it can be shown that this is the classifying space of $O(n)$.

There also exists a canonical inclusion of Grassmannians

$$\iota_{n,k} : \mathsf{Gr}(n, \mathbb{R}^k) \hookrightarrow \mathsf{Gr}(n+1, \mathbb{R}^{k+1}) : W \mapsto W \oplus \mathsf{span}\{e_{k+1}\}\,. \tag{41.8}$$

This in turn induces an inclusion $BO(n) \hookrightarrow BO(n+1)$ of classifying spaces. The direct limit of this system of inclusions is denoted by BO. It is the classifying space of the stable orthogonal group O.

Remark 41.2.13. A similar construction exists for the classifying spaces $BU(n)$ and BU.

Remark 41.2.14 (Kuiper's theorem). It should be noted that neither BO nor BU can be expressed as classifying spaces of automorphism groups of infinite-dimensional Hilbert spaces. This follows from *Kuiper's theorem*, which states that such groups are contractible and, hence, have vanishing homotopy groups (this clearly does not hold for the classifying spaces in this section).

Property 41.2.15 (Homotopy classification). For all spaces X, the K-theory can be represented as follows:

$$K^0_{\mathfrak{K}}(X) = [X, B\mathsf{GL}(\mathfrak{K}) \times \mathbb{Z}]\,. \tag{41.9}$$

When specializing to $\mathfrak{K} = \mathbb{R}, \mathbb{C}$, this gives

$$KO^0(X) = [X, BO \times \mathbb{Z}] \tag{41.10}$$

and

$$K^0(X) = [X, BU \times \mathbb{Z}] \tag{41.11}$$

due to homotopy invariance. Reduced K-theory can be obtained by considering basepoint-preserving homotopies, which, for connected spaces, gives $[X, B\mathsf{GL}(\mathfrak{K})]$.

Remark 41.2.16 (Noncompact spaces). When considering noncompact spaces, one can still use either the Grothendieck construction or the representable definition for topological K-theory. However, these will not coincide anymore even though there does exist an injection from Grothendieck K-theory to representable K-theory.

The following theorem should be compared to Remark 41.2.14 above (in fact, this theorem can be proven through *Kuiper's theorem*).

Theorem 41.2.17 (Atiyah–Jänich). *The space of Fredholm operators (Definition 26.1.37) on a separable, infinite-dimensional Hilbert space forms a classifying space for K-theory.*

41.2.3 Negative degree

Definition 41.2.18 (K^{-1}). For every field $\mathfrak{K}$, the functor K^{-1} is defined as follows (this can again be obtained as a property when using a different definition):

$$K^{-1}_{\mathfrak{K}}(X,Y) := [X/Y, \mathsf{GL}(\mathfrak{K})]_* \,, \tag{41.12}$$

where the asterisk denotes the fact that only basepoint-preserving homotopies are considered. $K^{-1}(X)$ can be obtained by considering $Y = \emptyset$ and recalling Eq. (6.14):

$$K^{-1}_{\mathfrak{K}}(X) := [X, \mathsf{GL}(\mathfrak{K})] \,. \tag{41.13}$$

For $\mathfrak{K} = \mathbb{R}, \mathbb{C}$, one can use homotopy invariance to obtain

$$\begin{aligned} KO^{-1}(X) &:= [X, \mathsf{O}] \,, \\ K^{-1}(X) &:= [X, \mathsf{U}] \,. \end{aligned} \tag{41.14}$$

To define lower degree groups, it will be useful to extend K-theory to locally compact spaces.

Definition 41.2.19 (K-theory of locally compact spaces). Let X be a locally compact space and denote its one-point compactification (Definition 6.5.30) by $\widehat{X}$. The groups $K^0(X)$ and $K^{-1}(X)$ are defined as follows:

$$\begin{aligned} K^0(X) &:= \ker\big(K^0(\widehat{X}) \to K^0(\{\infty\})\big) \,, \\ K^{-1}(X) &:= \ker\big(K^{-1}(\widehat{X}) \to K^{-1}(\{\infty\})\big) \,. \end{aligned} \tag{41.15}$$

So, the K-theory of a locally compact space is defined as the reduced K-theory of its one-point compactification.

Corollary 41.2.20 (Relative K-theory and complements). Consider a space X with a closed subspace Y. One can identify $(X/Y)\backslash\{y_0\}$ with $X\backslash Y$ and, hence, one obtains

$$K^0(X\backslash Y) = \widetilde{K}^0(X/Y) \,. \tag{41.16}$$

When combined with the excision property, this gives a result akin to ordinary (singular) cohomology, where the relative cocycles were those defined on the complement $X\backslash Y$:

$$K^0(X,Y) \cong K^0(X\backslash Y) \,. \tag{41.17}$$

Definition 41.2.21 (K^{-n}). Lower-degree relative K-groups are defined as follows:

$$K^{-n}(X,Y) := K^0\big((X\backslash Y) \times \mathbb{R}^n\big) \,. \tag{41.18}$$

By taking $Y = \emptyset$, one obtains the groups $K^{-n}(X)$ as K^0-groups of trivial line bundles:

$$K^{-n}(X) := K^0(X \times \mathbb{R}^n) \,. \tag{41.19}$$

Before relating this to reduced K-theory, the reader should be aware of a possible confusion. Homotopy invariance of K-theory would seem to imply that the above definition is senseless, since $X \cong X \times \mathbb{R}^n$ in the homotopy category. However, $X \times \mathbb{R}^n$ is not compact (even if X is) and, hence, one should work with Definition 41.2.19.

It can be shown through a series of homeomorphisms that the above definition is equivalent to the following one:

$$K^{-n}(X, Y) \cong \widetilde{K}^0\big(\Sigma^n(X/Y)\big)\,, \tag{41.20}$$

where Σ denotes the reduced suspension functor 6.3.7. As such, the reduced suspension functor gives a way to move down in the tower of K-groups.

41.2.4 Bott periodicity

Definition 41.2.22 (Cup product). First, generalize the Whitney sum and tensor product constructions to vector bundles over different base spaces. Let E, E' be vector bundles over the base spaces X, X' and consider the projection maps $\pi : X \times X' \to X$ and $\pi' : X \times X' \to X'$. The exterior sum $E \oplus E' \to X \times X'$ is defined as the Whitney sum $\pi^*(E) \oplus \pi'^*(E')$. Analogously, the exterior product bundle is defined as the tensor product $\pi^*(E) \otimes \pi'^*(E')$. Fibrewise, this is just the ordinary direct sum and tensor product of vector spaces.

The exterior product induces a bilinear map on K-theory as follows. From Definition 41.2.4, it follows that every element $x \in K^0(X)$ can be written as a formal difference $[E] - [E']$ of vector bundles over X. Using this decomposition, one can define the cup product $x \cup y$ through the following formula:

$$([E] - [E']) \cup ([F] - [F']) := [E \otimes F] + [E' \otimes F'] - [E \otimes F'] - [E' \otimes F]\,. \tag{41.21}$$

This definition can now be extended to locally compact spaces. Every element of $K^0(Y)$, for Y locally compact, defines an element in $K^0(\widehat{Y})$ and for such elements, the cup product was just defined. By restricting to $Y \times Y'$, one can then obtain an element of $K^0(Y \times Y')$:

$$K^0(Y) \times K^0(Y') \xrightarrow{\iota \times \iota'} K^0(\widehat{Y}) \times K^0(\widehat{Y'}) \xrightarrow{\cup} K^0(\widehat{Y} \times \widehat{Y'}) \xrightarrow{\text{res}} K^0(Y \times Y')\,, \tag{41.22}$$

where ι, ι' are the inclusions induced by Eq. (41.15).

Property 41.2.23 (Ring structure). By precomposing with the diagonal morphism

$$K^0(X \times X) \to K^0(X)\,, \tag{41.23}$$

the set $K^0(X)$ can be endowed with a commutative ring structure (at least over commutative fields such as $\mathbb{R}, \mathbb{C}$).

Property 41.2.24. It can immediately be seen that the cup product on K^0 also defines a bilinear operation $K^{-m}(X) \times K^{-n}(X) \to K^{-m-n}(X)$. Furthermore, as above, this induces a multiplicative structure on the complex $K^\bullet(X) := \bigoplus_{n=0}^{+\infty} K^{-n}(X)$. This multiplication can be shown to endow the K-complex with the structure of a graded-commutative algebra (Definition 23.5.5).

By using the isomorphism (41.17), one can also extend the cup product to an operation on relative K-theory:

$$K^{-m}(X,Y) \to K^{-n}(X',Y') \to K^{-m-n}(X \times X', X \times Y' \cup X' \times Y)\,. \tag{41.24}$$

Notation 41.2.25 (Bott element). Consider the complex relative K-group

$$K^0_{\mathbb{C}}(D^2,S^1) \cong \widetilde{K}^0_{\mathbb{C}}(S^2) \cong K^0_{\mathbb{C}}(\mathbb{R}^2) \tag{41.25}$$

of the unit disk with respect to its boundary. The Bott element is the K-theory element represented by the triple $\big(D^2 \times \mathbb{C}^2, D^2 \times \mathbb{C}^2, \alpha : (x,v) \mapsto xv\big)$. In terms of virtual vector bundles, it is represented by the difference of the trivial line bundle and the tautological line bundle on the 2-sphere.

Theorem 41.2.26 (Complex Bott periodicity). *The cup product with the Bott element gives an isomorphism $K^{-n}(X,Y) \cong K^{-n}(X \times D^2, X \times S^1 \cup Y \times D^2) \cong K^{-n-2}(X,Y)$. This also implies that cupping with this element gives an isomorphism $K^0(X) \cong K^0(X \times \mathbb{R}^2)$.*

Corollary 41.2.27. Applying Bott periodicity to the case $X = *, Y = \emptyset$ and comparing to Example 41.2.3, one obtains $K^0(D^2,S^1) \cong \mathbb{Z}$. One can also conclude that the Bott element is a generator of $K^0(D^2,S^1)$.

Corollary 41.2.28 (Spheres). Bott periodicity and Eq. (41.20) also allow to compute the reduced K-theory of spheres:

$$\widetilde{K}^0(S^n) = \begin{cases} 0 & \text{if } n \text{ is odd}\,, \\ \mathbb{Z} & \text{if } n \text{ is even}\,. \end{cases} \tag{41.26}$$

For even n, one can see that the generator is given by $\beta^{n/2}$, with β the Bott element.

Property 41.2.29 (Homotopy groups of unitary group). Property 34.2.4 can be generalized to the stable linear group to obtain an isomorphism $\widetilde{K}^0_{\mathfrak{K}}(S^n) \cong \pi_{n-1}\big(\mathsf{GL}(\mathfrak{K})\big)$. By specializing to the complex numbers, recalling that $\mathsf{GL}(m,\mathbb{C})$ deformation retracts onto $\mathsf{U}(m)$ and applying Bott periodicity, one obtains that the homotopy groups of the stable unitary group are mod 2-periodic. Furthermore, by using the long exact sequence induced by the fibration

$$\mathsf{U}(n) \to \mathsf{U}(n+1) \to S^{2n+1}\,, \tag{41.27}$$

one can show that, for $n > i/2$, the homotopy groups $\pi_i\big(\mathsf{U}(n)\big)$ satisfy the same periodicity relation.

Theorem 41.2.30 (Real Bott periodicity). *The cup product with the real Bott element, i.e. the generator of $KO^{-8}(*) \cong \mathbb{Z}$, gives an isomorphism*

$$KO^{-n}(X,Y) \to KO^{-n-8}(X,Y)\,. \tag{41.28}$$

Theorem 41.2.31 (Weak Bott periodicity). *The following spaces are homotopy equivalent:*[1]

$$\begin{aligned} \mathsf{GL}(\mathbb{R}) &\sim \Omega^8\,\mathsf{GL}(\mathbb{R}) \\ \mathsf{O} &\sim \Omega^8\mathsf{O} \\ \mathsf{U} &\sim \Omega(\mathbb{Z}\times B\mathsf{U}) \\ \mathbb{Z}\times B\mathsf{U} &\sim \Omega\mathsf{U}\,. \end{aligned} \tag{41.29}$$

Corollary 41.2.32. Through Eckmann–Hilton duality, this implies the periodicity of the homotopy groups of the stable orthogonal and unitary groups (cf. Property 41.2.29).

Property 41.2.33. One can also relate real and quaternionic K-theory through the following homotopy equivalences:

$$\begin{aligned} \mathbb{Z}\times B\mathsf{GL}(\mathbb{R}) &\sim \Omega^4(B\mathsf{GL}(\mathbb{H}))\,, \\ \mathbb{Z}\times B\mathsf{GL}(\mathbb{H}) &\sim \Omega^4(B\mathsf{GL}(\mathbb{R}))\,. \end{aligned} \tag{41.30}$$

Remark 41.2.34 (Positive degree). Bott periodicity allows to define K-groups in positive degree.

41.2.5 Clifford modules

One can restate the previous sections in terms of Clifford modules (Definition 36.3.17), i.e. vector bundles for which the fibres carry a representation of a Clifford algebra. The content of Section 27.3 will be used here.

From Construction 27.3.7 and Example 27.3.9, it should be clear that what was called $K^0(X)$ is in fact equivalent to $K^{0,0}\big(\mathbf{Vect}(X)\big)$. In a similar vein, one can prove that $K^{-1}(X)$ is equivalent to $K^{1,0}\big(\mathbf{Vect}(X)\big)$. This relation is, in fact, generalizable to all values of $p,q \in \mathbb{N}$.

Property 41.2.35. The $K^0(X)$-modules $K^{p,q}\big(\mathbf{Vect}(X)\big)$ and $K^{q-p}(X)$ are isomorphic.

Property 27.3.6 then implies the Bott periodicity for the groups $K^{q-p}(X)$. For complex K-theory, one can use Remark 27.3.10. The most important takeaway for this section is that one can rephrase K-theory in terms of Clifford modules and canonically induced functors between them.

[1] For an extensive lis,t see Karoubi (1978).

41.2.6 Cohomology theory

Property 41.2.36 (Excision). It can be shown that the excision property 41.2.11 holds at every degree $n \in \mathbb{Z}$:

$$K^n(X/Y, \{y_0\}) \cong K^n(X, Y)\,. \tag{41.31}$$

More generally, this can be stated as

$$K^n(X \backslash U, Y \backslash U) \cong K^n(X, Y)\,, \tag{41.32}$$

where $\overline{U} \subset Y^\circ$.

Property 41.2.37 (Homotopy invariance). Homotopic maps induce coinciding morphisms in K-theory at every degree. In particular, this implies that homotopy equivalences induce isomorphisms in K-theory.

Remark 41.2.38 (Generalized cohomology). The above properties imply that the complex of K-groups satisfies the Eilenberg–Steenrod axioms (Definition 7.5.1) except for the dimension axiom. As such, it is a generalized cohomology theory.

41.2.7 Applications

@@ ALSO REFER TO PHYSICS (topological phases, course notes Jutho) @@

Property 41.2.39 (Degree). Recall Definition 7.2.38 of the degree of a function in algebraic topology. In a similar way, one can assign to every continuous function $f : S^n \to S^n$ a degree through its induced action on K-theory. It can be shown that the topological degree and the K-theoretic degree coincide.

One can also extend this to **multidegrees**. For example, in the case of bidegree, one considers continuous functions $\mu : S^n \times S^n \to S^n$. The bidegree $(p, q) \in \mathbb{N} \times \mathbb{N}$ of μ is defined as the pair of degrees of the maps $x \mapsto \mu(x, x_0)$ and $y \mapsto (x_0, y)$ for a fixed basepoint $x_0 \in S^n$.

Definition 41.2.40 (H-space). A sphere S^n that admits a continuous function $\mu : S^n \times S^n \to S^n$ of bidegree $(1, 1)$. More generally (e.g. the H-structure on loop groups in Remark 7.1.13), it is a pointed topological space X that admits a function $\mu : X \times X \to X$ for which the pointwise maps are homotopic to the identity.

Property 41.2.41 (Puppe sequence). The mapping cone C_f from Construction 7.1.68 fits in an exact sequence:

$$X \longrightarrow Y \longrightarrow C_f \longrightarrow \Sigma X \longrightarrow \Sigma Y\,. \tag{41.33}$$

Reduced K-theory maps this to an induced exact sequence (whenever the reduced K-theory is defined for X, Y):

$$\widetilde{K}^0(\Sigma X) \longrightarrow \widetilde{K}^0(\Sigma Y) \longrightarrow \widetilde{K}^0(C_f) \longrightarrow \widetilde{K}^0(X) \longrightarrow \widetilde{K}^0(Y)\,. \tag{41.34}$$

The Puppe sequence in K-theory allows to prove an important theorem.

Definition 41.2.42 (Hopf invariant). The Puppe sequence for $f : S^{2n-1} \to S^n$ with n even implies that $\widetilde{K}^0(C_f) \cong \mathbb{Z} \oplus \mathbb{Z}$, where the generators are induced by the Bott elements of the spheres. The relation $\beta_{2n} = \beta_n \cup \beta_n$ of Bott elements gives an induced relation $a^2 = \lambda b$ of generators in $\widetilde{K}^0(C_f)$. The integer λ is called the Hopf invariant of f.[2]

One can show that every 'multiplication' map $\mu : S^{n-1} \times S^{n-1} \to S^{n-1}$ of bidegree (p, q) induces a map $S^{2n-1} \to S^n$ of Hopf invariant pq.

Theorem 41.2.43 (Atiyah–Adams). *Let n be even. If a continuous function $S^{2n-1} \to S^n$ has odd Hopf invariant, then $n = 2, 4$ or 8.*

Corollary 41.2.44. A sphere S^{n-1} can only admit a multiplication map of bidegree $(1, 1)$ or, equivalently, admit an H-structure, if $n = 2, 4$ or 8. (The case S^0 can be proven in a different way.)

Using Chern–Weil theory (Section 35.6.1), one can define a morphism to ordinary cohomology.

Construction 41.2.45 (Chern character). The Chern character of vector bundles (Definition 35.6.8) induces a Chern character on K-theory:

$$\mathsf{ch} : K(X) \mapsto H^\bullet(M) : [E] - [F] \mapsto \mathsf{ch}(\nabla^E) - \mathsf{ch}(\nabla^F)\,, \tag{41.35}$$

where ∇^E, ∇^F are connections on E and F, respectively. This morphism is a ring morphism (with respect to Whitney sums and tensor products). For relative K-theory classes $[E, F, \alpha] \in K(M, N)$, one should pick connections that are related by $\nabla^E|_N = \alpha^*(\nabla^F|_N)$.

Theorem 41.2.46 (Atiyah–Hirzebruch). *Consider a base space X. The Chern character induces the following isomorphisms:*

$$\begin{aligned} K^0_{\mathbb{C}}(X) \otimes \mathbb{Q} &\cong \bigoplus_{i \in \mathbb{Z}} H^{2i}(X; \mathbb{Q})\,, \\ K^1_{\mathbb{C}}(X) \otimes \mathbb{Q} &\cong \bigoplus_{i \in \mathbb{Z}} H^{2i+1}(X; \mathbb{Q})\,. \end{aligned} \tag{41.36}$$

For noncompact spaces, one needs to work with rational K-theory.

Property 41.2.47 (KK-theory). Topological K-theory and Kasparov K-theory (Definition 26.2.53) are related as follows:

$$\begin{aligned} KK\big(\mathbb{C}, C(X)\big) &\cong K^0(X)\,, \\ KK\big(C(\mathbb{R}), C(X)\big) &\cong K^1(X)\,. \end{aligned} \tag{41.37}$$

[2]The choice of generator a corresponding to β_n is not relevant due to the relations $b^2 = ab = 0$ obtained by dimensional arguments.

41.3 Twisted K-theory

The most straightforward way to introduce twisted K-theory is similar to that of twisted de Rham theory (Definition 34.9.11). Namely, by introducing a degree-3 integral cohomology class or, equivalently, by twisting by a $\mathsf{U}(1)$-*bundle gerbe* (see Chapter 44).

One way to obtain such a twist is the following approach. By the Atiyah–Jänich theorem, K-theory is classified by Fredholm operators on a separable infinite-dimensional Hilbert space $\mathcal{H}$. Now, consider a fibre bundle $\pi : P \to M$ with typical fibre the projective space $\mathcal{H}\mathbb{P}$. Locally, such a bundle can be obtained as the (fibrewise) projectivization of a Hilbert bundle. By lifting the $\mathsf{PGL}(\mathbb{C})$-valued transition functions to $\mathsf{GL}(\mathbb{C})$-valued functions, one obtains a $\mathsf{U}(1)$-valued 3-cocycle $f_{ijk} := \tilde{g}_{ij}\tilde{g}_{jk}\tilde{g}_{ki}$. This defines a class in $\check{H}^2(M;\mathsf{U}(1)) \cong H^3(M;\mathbb{Z})$. Isomorphism classes of bundles of projective spaces are in bijection with such cohomology classes and, if the class vanishes, the bundle can be globally obtained as the fibrewise (projectivization) of a Hilbert bundle. Given that the projective unitary group also acts on Fredholm operators (by conjugation), one can, for every class $\alpha \in H^3(M;\mathbb{Z})$, construct a bundle of Fredholm operators $\mathsf{Fred}(\mathcal{H}) \hookrightarrow \mathsf{Fred}(P) \to M$.

Definition 41.3.1 (Twisted K-theory). Consider a cohomology class $\alpha \in H^3(M;\mathbb{Z})$ together with its associated bundle of Fredholm operators $\mathsf{Fred}(P) \to M$. The α-twisted K-theory on M is defined as the set of homotopy classes of sections of this bundle:

$$K^0_\alpha(M) := \Gamma\big(\mathsf{Fred}(P)\big)/\sim . \tag{41.38}$$

41.4 Differential K-theory

Just as in the case of ordinary cohomology (Section 35.7), one can also generalize ordinary K-theory to a differential cohomology theory by incorporating connection data. Grothendieck K-theory makes use of (virtual) vector bundles and, therefore, the most natural way to define differential K-theory would be through (virtual) vector bundles with connection. This is the approach of *Simons* and *Sullivan*.

Definition 41.4.1 (Simons–Sullivan model). A vector bundle equipped with an equivalence class of connections, where two connections are equivalent if their Chern–Simons form is exact, is called a **structured bundle**. Two structured bundles are isomorphic if there exists a vector bundle isomorphism that identifies the connections.

Consider the set of $\mathsf{Struct}(M)$ of isomorphism classes of structured bundles over M. This set becomes a commutative monoid under Whitney sums. The Grothendieck completion (Construction 2.2.6) of $\mathsf{Struct}(M)$ gives a model for differential K-theory:

$$\widehat{K}^0(M) := G\big(\mathsf{Struct}(M)\big) . \tag{41.39}$$

@@ COMPLETE @@

41.5 Algebraic K-theory

41.5.1 Determinant

Over noncommutative rings R, the determinant of a matrix is not as easily defined as over commutative rings. For example, in the 2×2-case one could choose either $ad - bc$ or $da - bc$ (or any other permutation). There exists no canonical choice. To fix this issue, one can take a look at the most important properties of the determinant map:

- It is invariant under elementary row/column operations (item 3 of Property 22.4.39).
- It is invariant under augmentation by the identity, i.e. under the transformation $A \mapsto A \oplus \mathbb{1}$.

To implement the second property, it will be necessary to move from the finite-dimensional general linear group $\mathsf{GL}(n, R)$ to its stable version (Definition 41.1.1). On this group, one can define an equivalence relation by saying that two matrices are equivalent if they belong to the same coset with respect to the subgroup $E(R)$ generated by the elementary matrices 22.4.51. It can also be shown that $E(R)$ is equal to the commutator subgroup $[\mathsf{GL}(R), \mathsf{GL}(R)]$.

The determinant map is then abstractly defined as the Abelianization map.

Definition 41.5.1 (K_1). The first algebraic K-group of a ring R is defined as the Abelianization of its stable linear group (Property 2.2.11):

$$K_1(R) := \mathsf{GL}(R)/[\mathsf{GL}(R), \mathsf{GL}(R)]\,. \tag{41.40}$$

The quotient map $\pi : \mathsf{GL}(R) \to K_1(R)$ is called the **determinant map**.

To obtain lower K-groups, a 'suspension functor' needs to be defined.

Definition 41.5.2 (Suspension). Let R be a ring. Denote the infinite matrix ring over R by $M(R)$, i.e. the set of matrices with a finite number of nonzero entries in each row and column. This ring contains an ideal $M_{\text{fin}}(R)$ generated by all matrices that are zero outside a block of finite size. The suspension of R is then defined as follows:

$$\Sigma R := M(R)/M_{\text{fin}}(R)\,. \tag{41.41}$$

Definition 41.5.3 (Lower K-groups). For all integers $n \geq 1$, one defines the K-groups as follows:

$$K_n(R) := K_{1-n}(\Sigma^n R)\,. \tag{41.42}$$

Example 41.5.4 (K_0). It can be shown that $K_0(R)$ corresponds to the Grothendieck group associated to the monoid of finitely generated projective R-modules. The relation to topological K-theory is then given by the Serre–Swan theorem 34.2.7:

$$K_0\big(C(X)\big) \cong K^0(X). \tag{41.43}$$

41.6 Operator K-theory

Definition 41.6.1 (K_0). Let A be a unital C^*-algebra. The group $K_0(A)$ is defined as the Grothendieck completion of the monoid with the projections in $M_n(A)$ for all $n \in \mathbb{N}$ as generators and the relations

1. $[p \oplus q] = [p] + [q]$.
2. $[0] = 0$.
3. If p, q are connected by a continuous path of projections, then $[p] = [q]$.

The first two conditions imply, in particular, that the embedding of a matrix as the upper-left part of a larger matrix leaves the class unchanged.

Remark 41.6.2. One could replace homotopy equivalence by Murray–Von Neumann equivalence in this definition, since two projections p and q are Murray–Von Neumann-equivalent if and only if the matrices

$$\begin{pmatrix} p & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} q & 0 \\ 0 & 0 \end{pmatrix} \tag{41.44}$$

are homotopy-equivalent.

Property 41.6.3 (Topological K-theory). If X is a compact Hausdorff space, then

$$K^0(X) \cong K_0\big(C(X)\big)\,. \tag{41.45}$$

This isomorphism is induced by the Serre–Swan theorem 34.2.7.

Definition 41.6.4 (K_1). Let A be a unital C^*-algebra. The group $K_1(A)$ is defined as the Grothendieck completion of the monoid with the unitaries in $M_n(A)$ for all $n \in \mathbb{N}$ as generators and the relations

1. $[p \oplus q] = [p] + [q]$.
2. $[\mathbb{1}] = 0$.
3. If p, q are connected by a continuous path of unitaries, then $[p] = [q]$.

The first two conditions imply, in particular, that the embedding of a matrix as the upper-left part of a larger matrix leaves the class unchanged.

Property 41.6.5 (Bott periodicity). First, define the suspension of a C^*-algebra A as follows:

$$\Sigma A := \{ f \in C([0,1], A) \mid f(0) = f(1) = 0 \} . \tag{41.46}$$

Then, $K_1(A) \cong K_0(\Sigma A)$. Moreover, one defines all higher homology groups recursively as[3]

$$K_n(A) \cong K_{n-1}(\Sigma A) . \tag{41.47}$$

Property 41.6.6 (KK-theory). If A is a C^*-algebra, then operator K-theory and Kasparov K-theory (Definition 26.2.53) are related as follows:

$$\begin{aligned} KK(\mathbb{C}, A) &\cong K_0(A) , \\ KK\big(C(\mathbb{R}), A\big) &\cong K_1(A) . \end{aligned} \tag{41.48}$$

Property 41.2.47 is then simply a consequence of this isomorphism by Property 41.6.3.

[3]There exists a more general definition of operator K-theory for which this relation is a property.

Chapter 42

Noncommutative Geometry

This chapter heavily uses the concepts introduced in Chapter 28.

Chapter Contents

42.1 Noncommutative measure theory 792
42.2 Quantum geometry 795
42.3 Spectral geometry 796
42.4 Dunkl operators 800

42.1 Noncommutative measure theory

In this section, the content of Chapters 18 and 45 is generalized to noncommutative spaces using the content of Chapter 26. A good reference is Segal (1965).

The start is the following observation.

Property 42.1.1. Let X be a locally compact Hausdorff space and consider its algebra of functions with compact support $C_c(X)$. By the Riesz–Markov theorem 19.1.6, the positive linear functionals on this algebra correspond to Radon measures on X. The subspace of states then correspond exactly to the probability measures on X. Under this identification, the pure states correspond to Dirac measures (evaluation functionals).

This property can be generalized to characterize arbitrary probability spaces.

Property 42.1.2. Let A be a unital commutative algebra and consider a functional E on A. Assume, moreover, that the following conditions hold:

1. **Positivity**: $E(a^2) \geq 0$.

2. **Nondegeneracy**: $E(a^2) = 0 \iff a = 0$.

3. **Boudnedness**: For any $b \in A$, there exists a constant $\lambda_b \in \mathbb{R}$ such that

$$E(a^2 b) = \lambda_b E(a^2) \tag{42.1}$$

for all $a \in A$.

Then, (A, E) is isomorphic to a dense subalgebra of $L^\infty(M)$ for some finite measure space M, where the integral operator is given by E.

The following result extends the diagonalization property for self-adjoint operators to the infinite-dimensional setting.

Corollary 42.1.3 (Diagonalization). For any bounded self-adjoint operator Q on a Hilbert space $\mathcal{H}$, there exists a unique projection-valued measure μ on $\mathbb{R}$ such that

$$Q = \int_{\mathbb{R}} \lambda \, d\mu \,. \tag{42.2}$$

Moreover, μ vanishes outside the spectral radius of Q.

Theorem 42.1.4 (Gleason). *Consider a **finitely additive measure**[1] on the von Neumann algebra $\mathcal{B}(\mathcal{H})$, i.e. a function μ on the set of orthogonal projections in $\mathcal{B}(\mathcal{H})$ such that*

$$\mu\left(\sum_{i=1}^{n} p_i\right) = \sum_{i=1}^{n} \mu(p_i) \tag{42.3}$$

for all $n \in \mathbb{N}$. If $\dim(\mathcal{H}) \neq 2$, each finitely additive measure can be uniquely extended to a state on $\mathcal{B}(\mathcal{H})$. Moreover, every σ-additive measure can be uniquely extended to a normal state.

Conversely, the restriction of every state (resp. normal state) to the set of orthogonal projections induces a finitely additive (resp. σ-additive) measure.

By extending Property 42.1.2 above, a noncommutative integration theory can be obtained.

Definition 42.1.5 (Noncommutative integration algebra). A unital $*$-algebra A with a functional E on A for which the following conditions hold:

1. **Positivity**: $E(a^*a) \geq 0$.

2. **Normality**: $E(a^*a) = E(aa^*)$

3. **Boudnedness**: For any $b \in A$, there exists a constant $\lambda_b \in \mathbb{R}$ such that

$$E(a^* b a) = \lambda_b E(a^* a) \tag{42.4}$$

for all $a \in A$.

[1] If this also holds for countable sequences, it is called **σ-additive** in analogy with Definition 18.1.1.

Definition 42.1.6 (Distribution). Consider a locally convex vector space (Definition 25.3.8) V. A **predistribution** on V is a linear map from V^* to the space of **measurables** on a finite measure space (Ω, Σ, μ), i.e. the space of equivalence classes of measurable functions modulo null functions.

A distribution on V is an equivalence class of predistributions, where two distributions d, d' are deemed equivalent if

$$\big(d(x_1), \dots, d(x_n)\big) \overset{d}{=} \big(d'(x_1), \dots, d'(x_n)\big) \tag{42.5}$$

for all $n \in \mathbb{N}$ and random variables $\{x_1, \dots, x_n\}$.

Remark 42.1.7. For finite-dimensional V, this definition is equivalent to the definition of ordinary finite measures. For, infinite-dimensional V, however, distributions (as defined here) are strictly more general. Distributions induced by ordinary measures will be called **strict**.

Example 42.1.8 (Normal distribution). Let $\mathcal{H}$ be a (real) Hilbert space and consider a bounded, positive operator C on $\mathcal{H}$. There exists a unique distribution on $\mathcal{H}$ such that:

- $\big(d(x_1), \dots, d(x_n)\big)$ are jointly normally distributed for all $n \in \mathbb{N}$ and $x_1, \dots, x_n \in \mathcal{H}$.
- For all $x \in \mathcal{H}$: $E\big(d(x)\big) = 0$.
- For all $x, y \in \mathcal{H}$: $E\big(d(x)d(y)\big) = \langle Cx \mid y\rangle$.

Such a distribution is strict if and only if $C \in \mathcal{B}_1(\mathcal{H})$. An important example of nonstrict distributions induced by self-adjoint operators are the **isonormal distributions**, where $C = \lambda \mathbb{1}_{\mathcal{H}}$.

The relation between (noncommutative) integration algebras and measure theory is given by the following property.

Property 42.1.9. Consider an integration algebra (A, E) such that $E(1) = 1$, where A is given by the space of all **tame functions** on a TVS V, i.e. all functions $f : V \to \mathbb{R}$ such that

$$f(x) = F\big(\lambda_1(x), \dots, \lambda_n(x)\big) \tag{42.6}$$

for some $n \in \mathbb{N}$, $\{\lambda_1, \dots, \lambda_n\} \subset V^*$ and Borel-measurable $F : \mathbb{R}^n \to \mathbb{R}$. There exists a (pre)distribution d on V with respect to some measure space (Ω, Σ, μ) such that:

- d extends to a $*$-morphism from A mapping a tame function f to

$$d(f) := F\big(d(\lambda_1), \dots, d(\lambda_n)\big)\,. \tag{42.7}$$

- $E(f) = \displaystyle\int_\Omega d(f)\, d\mu$.

Moreover, Ω can be chosen such that $d(A)$ is dense in $L^p(\Omega)$. Up to isomorphism, this choice is unique.

Definition 42.1.10 (Random variable). Consider a predistribution d (with respect to a measure space Ω) extended to a $*$-morphism on the space A of tame functions on some TVS V. The space of random variables $M(V, d)$ is defined as the minimal subspace of measurables on Ω that contains $d(A)$ and is closed with respect to (sequential) pointwise a.e. limits.

42.2 Quantum geometry

Definition 42.2.1 (Quantum metric). Consider an FODC (A, Γ) as in Definition 28.4.1. A **generalized inner-product** is a bimodule morphism $\langle \cdot \mid \cdot \rangle : \Gamma \otimes \Gamma \to A$. A (quantum or **generalized**) metric with respect to an inner product $\langle \cdot \mid \cdot \rangle$ is an element $g \in \Gamma \otimes \Gamma$ such that

$$\langle \omega \mid \cdot \rangle \otimes \mathbb{1}(g) = \omega = \mathbb{1} \otimes \langle \cdot \mid \omega \rangle(g) \tag{42.8}$$

for all $\omega \in \Gamma$. This condition represents the invertibility of the metric.

Definition 42.2.2 (Connection). Let $(\Omega^\bullet, \mathsf{d})$ be a differential calculus on an associative algebra A. A connection on a left A-module Γ is a linear map

$$\nabla : \Gamma \to \Omega^1 \otimes_A \Gamma \tag{42.9}$$

such that

$$\nabla(a\omega) = \mathsf{d}a \otimes_A \omega + a\nabla\omega\,. \tag{42.10}$$

By the graded Leibniz rule, this extends to a linear map on all of $\Omega^\bullet \otimes_A \Gamma$. The **curvature** of ∇ is given by ∇^2.

If $\Gamma = \Omega^1$, one speaks of a linear connection. In this case, one can also define the **torsion** of the connection:

$$T := \mathsf{d} - \pi \circ \nabla\,, \tag{42.11}$$

where $\pi : \Omega^1 \otimes_A \Omega^1 \to \Omega^2$ is the canonical projection. Although notions such as the Ricci tensor do not extend to the noncommutative setting, both the Cartan and Bianchi identities do.

Property 42.2.3 (Existence). A connection on a finite-rank module P for the universal differential calculus $\Omega^\bullet(A)$ exists if and only if the module is projective. An explicit form is given by

$$\nabla = p \circ \mathsf{d}\,, \tag{42.12}$$

where $p : A \to P$ is the canonical projection.

Example 42.2.4. Let A be a C^*-algebra (Definition 26.2.6). By the Gel'fand–Naimark theorem 26.2.42, there exists a representation $\rho : A \to \mathcal{B}(\mathcal{H})$ on a Hilbert space $\mathcal{H}$. If there exists an operator $D \in \mathcal{B}(\mathcal{H})$ such that $[D, \rho(A)] \subseteq \mathcal{B}(\mathcal{H})$, there exists an induced representation of the universal exterior algebra (Property 28.4.7):

$$\tilde{\rho} : \Omega^\bullet(A) \to \mathcal{B}(\mathcal{H}) : a_0 \mathrm{d}a_1 \otimes \cdots \otimes \mathrm{d}a_n \mapsto \rho(a_0)[D, \rho(a_1)] \cdots [D, \rho(a_n)] \,. \tag{42.13}$$

The problem here is that $\tilde{\rho}(a) = 0$ need not imply that $\tilde{\rho}(\mathrm{d}a) = 0$. One can, however, obtain a smaller differential calculus by modding out the differential ideal generated by the kernel of this action:

$$\widetilde{\Omega}^\bullet(A) := \Omega^\bullet(A) / \{\ker(\tilde{\rho}) + \mathrm{d}\ker(\tilde{\rho})\} \,. \tag{42.14}$$

Connes proved that if $A = C^\infty(M)$ for some Riemannian manifold M, with D the associated Dirac operator, then $\widetilde{\Omega}^\bullet(A)$ is the exterior complex of differential forms. For this reason, the above differential calculus is often called **Connes' differential calculus**.

Formula 42.2.5. Recall Property 42.2.3 above. If P is finite rank and projective, there exists a canonical connection $p \circ \mathrm{d}$ with respect to the universal differential calculus. By postcomposition with the projection to Connes' differential calculus, one obtains a new connection ∇_0. For any A-module morphism $\alpha : P \to P \otimes_A \widetilde{\Omega}^1(A)$, one obtains a new connection:

$$\nabla := \nabla_0 + \alpha \,. \tag{42.15}$$

The morphism α is often called a (noncommutative) **gauge field**.

42.3 Spectral geometry

In this section, the work of *Connes* and others is introduced. Here, the structure of a (compact) Riemannian manifold is completely encoded in the algebraic data of a commutative algebra and a linear 'Dirac' operator. By generalizing the algebraic properties, one can obtain noncommutative geometries.

Definition 42.3.1 (Spectral triple). A triple $(A, \mathcal{H}, D)$ where

1. $\mathcal{H}$ is a separable Hilbert space,
2. A is a $*$-subalgebra of $\mathcal{B}(\mathcal{H})$ or, more generally, a $*$-algebra with a faithful $*$-representation $\pi : A \to \mathcal{B}(\mathcal{H})$, and
3. D is an (unbounded) self-adjoint operator on $\mathcal{H}$

that satisfy the following conditions:

1. **Closure**: $[D, A] \subseteq \mathcal{B}(\mathcal{H})$.

2. **Compact resolvent**: The resolvent $(\pi(a)D)_\lambda$ is compact for all $a \in A, \lambda \in \mathbb{C}\backslash\mathbb{R}$. If A is unital, this only has to be checked for D itself.

3. **Chirality**[2]: There exists a grading γ, turning $\mathcal{H}$ into a super-Hilbert space, such that

$$[\gamma, A] = 0 \qquad \{\gamma, D\}_+ = 0\,. \tag{42.16}$$

Remark 42.3.2 (Connes axioms). The conditions in the previous definition are often extended by the following list of 8 axioms that generalize the behaviour of the 'ordinary' Dirac operator:

1. **Finite summability**: The operator $|D|^{-1}$, where $|D|$ is obtained by a polar decomposition, is compact. Moreover, one requires that there exists a number $p \in \mathbb{N}$ such that the spectrum goes as $O(n^{-1/p})$. Equivalently, there exists an integer $n \in \mathbb{N}$ such that the Dixmier trace (Definition 26.1.63) of $|D|^{-n}$ is finite and nonzero. The smallest such integer is called the **dimension** of the spectral triple.

2. **First-order**: $[[D, A], A] = 0$.

3. **Strong regularity**: On $\mathcal{B}(\mathcal{H})$, define the operator

$$\delta : T \mapsto [|D|, T] \tag{42.17}$$

and consider the spaces

$$\mathcal{H}^\infty := \bigcap_{i=1}^{+\infty} \mathsf{dom}(D^i) \qquad \mathcal{B}^\infty(\mathcal{H}) := \bigcap_{i=1}^{+\infty} \mathsf{dom}(\delta^i)\,. \tag{42.18}$$

The spectral triple is said to be strongly regular if A, $[D, A]$ and $\mathsf{End}(\mathcal{H}^\infty)$ lie in $\mathcal{B}^\infty(\mathcal{H})$. If only the first two algebras satisfy this property, then the spectral triple is only said to be **regular**. Restricting to this subspace is similar to restricting to Sobolev spaces in the study of PDEs.

4. **Orientability**: Recall the definition of Hochschild homology (Definition 5.6.5). By regularity, there exists an induced representation

$$\pi_H : HC_n(A) \to \mathcal{B}(\mathcal{H}) : a_0 \otimes \cdots \otimes a_n \mapsto a_0[D, a_1] \cdots [D, a_n]\,. \tag{42.19}$$

The spectral triple is said to be orientable if there exists an antisymmetric Hochschild p-cycle χ, where $p \in \mathbb{N}$ is the dimension of the spectral triple, such that $\gamma = \pi_H(\chi)$. This is sometimes called the **noncommutative volume form**. (For odd spectral triples, one requires the image of this Hochschild cycle to be 1.)

[2]One sometimes makes a difference between **even** and **odd** spectral triples based on the validity of this axiom.

5. **Charge conjugation**: There exists an antiunitary operator $C \in \mathsf{End}(\mathcal{H})$ such that

$$C^2 = \pm 1 \qquad CD = \pm DC \qquad C\gamma = \pm \gamma C\,, \tag{42.20}$$

where the three signs depend on the dimension of the spectral triple. Moreover, one requires that $a = Ca^*C^{-1}$ for all $a \in A$. The dependence on the dimension is given by the following table (dimensions mod 8):[3]

n	0	1	2	3	4	5	6	7
C^2	$+$	$+$	$-$	$-$	$-$	$-$	$+$	$+$
D	$-$	$+$	$-$	$-$	$-$	$+$	$-$	$-$
γ	$-$		$+$		$-$		$+$	

6. **Irreducibility**: There exists no nontrivial projection on $\mathcal{H}$ that commutes with A, D, C or γ.

7. **Finiteness**: $\mathcal{H}^\infty$ is a finitely generated, projective (right) A-module.

8. **Absolute continuity**: Consider two elements $v, w \in \mathcal{H}^\infty$ and an operator $a \in A$.

$$\langle v \mid w \rangle_{\mathcal{H}} = \int \langle v \mid w \rangle_A\,, \tag{42.21}$$

where

$$\int : A \to \mathbb{C} : a \mapsto \mathrm{tr}_{\mathfrak{D}}(a|D|^{-p}) \tag{42.22}$$

is the 'noncommutative integral' induced by the Dixmier trace (Definition 26.1.63) ($a|D|^{-p}$ is measurable for all $a \in A$) and

$$\langle v \mid w \rangle_A := \sum_{i,j} v_i^* p_{ij} w_j \tag{42.23}$$

for p the projection characterizing $\mathcal{H}^\infty$. This condition relates the inner product on $\mathcal{H}$ and the (pre-)Hilbert A-module structure on $\mathcal{H}^\infty$.

@@ COMPLETE @@

Example 42.3.3 (Spin manifolds). The above 8 axioms admit a reasonable interpretation in the case of a spin manifold M, its algebra of smooth functions $C^\infty(M)$ and the Dirac operator (Definition 36.3.19) on its spinor bundle.

- The Hilbert space $\mathcal{H}^\infty$ is given by the square-integrable sections of the spinor bundle on M.
- The action of the Dirac operator on a smooth function is again a smooth function and, hence, commutes with all of $C^\infty(M)$.

[3] This is related to the notion of KO-dimension in (real) K-theory.

- The noncommutative volume form is given by the chirality operator $\gamma \sim \prod_{i=1}^{p} \gamma^i$, where γ^i are the Clifford generators.
- Irreducibility corresponds to M being connected.
- By the Serre–Swan theorem 34.2.7, $\mathcal{H}^\infty$ being a finitely generated, projective module implies that it is a vector bundle over M.
- The absolute continuity simply states that the restriction of the inner product to $\mathcal{H}^\infty$ is given by first taking the fibrewise pairing of a spinor and a cospinor and then integrating the resulting function over all of M.

@@ COMPLETE @@

Example 42.3.4 (Hodge theory). Another example of a spectral triple, where the underlying manifold does not have to be spin, is given by Hodge theory. On every Riemannian manifold, there exists a first-order operator

$$D := \mathrm{d} + \delta \,. \tag{42.24}$$

This is the Hodge–de Rham operator (Definition 36.1.5).

Construction 42.3.5 (Connes' reconstruction theorem). Let $(A, \mathcal{H}, D)$ be a commutative spectral triple, i.e. a spectral triple for which A is commutative. Let $p \in \mathbb{N}$ be the dimension of the spectral triple. If the spectral triple satisfies the 8 axioms above, the following properties hold:

- The Gel'fand spectra of A and of its (norm-)closure $\overline{A}$ coincide. Denote this set by M.
- There exists a cover $\{U_i\}_{i\in I}$ of M such that for every covering set U_i there exist p self-adjoint elements $\{x_i^\mu\}_{\mu\leq p} \subset A$ such that the map

$$\hat{x}_i := (\hat{x}_i^1, \dots, \hat{x}_i^p) : U_i \to \mathbb{R}^p \,, \tag{42.25}$$

  where $\hat{x}$ denotes the Gel'fand transformation of x, is a local homeomorphism.
- There exists a smooth family $\tau : \mathbb{R}^p \to \mathrm{Aut}(A) : v \mapsto \tau_v$, with $\tau_0 = \mathbb{1}_A$, such that for all $\chi \in M$ the function $\chi \circ \tau : \mathbb{R}^p \to M$ is a homeomorphism on some neighbourhood of 0 and for all $i \in I$ the function $\hat{x}_i \circ \chi \circ \tau : \mathbb{R}^p \to \mathbb{R}^p$ is a local diffeomorphism.
- If $\{a_1, \dots, a_n\} \subset A$ are self-adjoint and $f : \mathbb{R}^p \to \mathbb{C}$ is smooth, then the image of $f|_\Delta$, where Δ is the joint spectrum of $\{a_1, \dots, a_n\}$, under the Gel'fand isomorphism is again an element of A. It will be denoted by $f(a_1, \dots, a_n)$.
- There exists a 'smooth' partition of unity $\{\psi_j\}_{j\in J\subset I} \subset A$ subordinate to any cover of M.

Then M is a smooth, compact p-dimensional manifold and $A \cong C^\infty(M)$ as $*$-algebras.

The coordinate functions on M can be derived from the noncommutative volume form $\chi \in HC_n(A)$.

@@ COMPLETE @@

Formula 42.3.6 (Distance). Let $(A, \mathcal{H}, D)$ be a spectral triple such that

$$\{a \in A \backslash \mathbb{C} \mid \|[D, a]\| \leq 1\} \tag{42.26}$$

is norm-bounded in $\overline{A} \backslash \mathbb{C}$. Then

$$d(\phi, \psi) := \sup\left\{|\phi(a) - \psi(a)| \mid a \in A : \|[D, a]\| \leq 1\right\} \tag{42.27}$$

defines a metric on the space of pure states on $\overline{A}$. This is a generalization of the *Monge–Kantorovich distance* on probability measures.

Remark 42.3.7. For the canonical spectral triple on a compact spin manifold, the above distance coincides with the geodesic distance.

Property 42.3.8 (Fredholm modules). When $(A, \mathcal{H}, D)$ is a spectral triple, the operator $D|D|^{-1}$ satisfies the properties of a Fredholm module (Definition 26.2.49). This module is even (resp. odd) when the spectral triple is even (resp. odd).

42.4 Dunkl operators

Let G be a Coxeter group (Definition 32.4.48) with root system Φ over the space $\mathbb{R}^n$. A function $m : \Phi \to \mathbb{C}$ is called a **multiplicity (function)** if $m(\alpha) = m(\beta)$ whevener σ_α and σ_β are conjugate elements in G.

Definition 42.4.1 (Dunkl operator). Let m be a multiplicity on G. The Dunkl operators $T_i : C^1(\mathbb{R}^n) \to C^0(\mathbb{R}^n)$ are defined as follows:

$$T_i f(x) := \partial_i f(x) + \sum_{\alpha \in \Phi^+} m(\alpha) \frac{f(x) - f(\sigma_\alpha x)}{\langle \alpha \mid x \rangle} \langle \alpha \mid e_i \rangle \,, \tag{42.28}$$

where e_i is the ith basis element of $\mathbb{R}^n$.

Property 42.4.2 (Commutativity). For all $i, j \leq n$ the following property holds:

$$T_i T_j f = T_j T_i f \,. \tag{42.29}$$

Chapter 43

Synthetic Differential Geometry ♣

Chapter Contents

43.1 Neighbourhoods **801**

43.2 Affine connections **802**

43.3 Euclidean geometry **802**

43.3.1 Infinitesimal elements 802

43.3.2 Calculus 803

43.1 Neighbourhoods

Definition 43.1.1 (Neighbourhood relation). A reflexive and symmetric relation with the additional property that the morphisms in the category under consideration preserve this relationship.

Example 43.1.2 (Monad). Let M be a set. Given a neighbourhood relation $\sim$ on M, the (first order) monad around $x \in M$ is defined as

$$\underline{\mathfrak{M}}(x) := \{y \in M \mid y \sim x\}\,. \tag{43.1}$$

Definition 43.1.3 (Infinitesimal simplex). An infinitesimal k-simplex with respect to a neighbour relation $\sim$ is a collection of $k+1$ points $\{x_i\}_{i \leq k}$ such that $x_i \sim x_j$ for all $i, j \leq k$.

Definition 43.1.4 (Geometric distribution). Let M be a set equipped with a neighbourhood relation $\sim$. A (geometric) distribution on M is a reflexive symmetric refinement $\approx$ of $\sim$. A distribution is said to be **involutive** if

$$(x \approx y) \wedge (y \approx z) \wedge (x \sim z) \implies x \approx z \tag{43.2}$$

for all $x, y, z \in M$.

Definition 43.1.5 (Integral subset). Let M be a set equipped with a neighbourhood relation $\sim$ and an associated distribution $\approx$. A subset $N \subseteq M$ is said to be integral with respect to $\approx$ if $\approx$ and $\sim$ coincide on N.

Theorem 43.1.6 (Frobenius' theorem). *An involutive distribution admits maximal connected integral subsets. These subsets are called the* ***leaves*** *of the distribution.*

43.2 Affine connections

Definition 43.2.1 (Affine connection). An affine connection is a map $\lambda : M \times M \times M \to M$ that for every three points $x, y, z \in M$ such that $y \sim x$ and $z \sim x$ gives a point $w \in M$ such that $w \sim z$ and $w \sim y$. Graphically, this is given by the completion of a span as shown in Diagram 43.1.

$$\begin{tikzcd} & & w \\ z \arrow[urr, dashed] & & y \arrow[u, dashed] \\ x \arrow[u] \arrow[urr] & & \end{tikzcd}$$

Figure 43.1: Connection in synthetic theories.

Remark. By looking at these diagrams, the concept of parallel transport can be made a lot more intuitive than in classic differential geometry. Diagram 43.1, for example, shows the parallel transport of the point z along xy.

Definition 43.2.2 (Symmetric connection). An affine connection λ is said to be symmetric or **torsion free** if $\lambda(x, y, z) = \lambda(x, z, y)$.

Definition 43.2.3 (Flat connection). An affine connection λ is said to be flat or **curvature free** if parallel transporting a point around an infinitesimal 2-simplex leaves the point invariant.

Definition 43.2.4 (Curvature). Let λ be an affine connection on M. The curvature of λ is the map $\mathcal{R}$ that assigns to every infinitesimal 2-simplex $\{x_0, x_1, x_2\}$ the automorphism

$$\underline{\mathfrak{M}}(x_0) \to \underline{\mathfrak{M}}(x_0) : z \mapsto \text{result of parallel transporting z around } \{x_0, x_1, x_2\}\,. \tag{43.3}$$

Definition 43.2.5 (Geodesic). A subset $S \subseteq M$ stable under the affine connection λ.

43.3 Euclidean geometry

43.3.1 Infinitesimal elements

Definition 43.3.1 (Infinitesimal line). Consider the line R. By choosing two points, 0 and 1, on R it can be given the structure of a commutative ring. The infinitesimal line

is defined as the following set:

$$\Delta := \{x \in R \mid x^2 = 0\}\,. \tag{43.4}$$

A neighbourhood relation on R is induced by setting $\underline{\mathfrak{M}}(0) = \Delta$.

Remark 43.3.2. If the Euclidean doctrine would be followed, where $R = \mathbb{R}$, this set would be $\{0\}$. However, by not requiring R to be a field, a larger set is obtained.

Axiom 43.1 (Kock–Lawvere). For every map $f : \Delta \to R$, there exists a unique element $b \in R$, called the **slope** of f, such that

$$f(d) = f(0) + d \cdot b \tag{43.5}$$

for all $d \in \Delta$.

Corollary 43.3.3. The function $\alpha : R \times R \to R^\Delta : (a, b) \mapsto (f : d \mapsto a + d \cdot b)$ is invertible and, hence, is an isomorphism[1].

Corollary 43.3.4. Let $a, b \in R$. If $d \cdot a = d \cdot b$ for all $d \in \Delta$, then $a = b$.

Notation 43.3.5. Analogous to the infinitesimal line, one can define the sets $\mathbb{D}^k$ in the following way:

$$\mathbb{D}^k := \{x \in R \mid x^{k+1} = 0\}\,. \tag{43.6}$$

With this notation one has $\Delta \equiv \mathbb{D}^1$.

43.3.2 Calculus

Formula 43.3.6 (Taylor expansion). The Kock–Lawvere axiom implies the following exact Taylor expansion:

$$f(x + d) = f(x) + d \cdot f'(x)\,, \tag{43.7}$$

where $f'(x)$ can be interpreted as the derivative of f at the point $x \in R$.

Remark. If f also depends on additional parameters in R, one can define the partial derivatives in a similar fashion.

Property 43.3.7. The Kock–Lawvere axiom implies that the derivative is linear and satisfies Leibniz's rule.

Using the sets $\mathbb{D}^k$, one can derive higher-order expansions. To this end, Axiom 43.1 is generalized.

[1] If one equips the set $R \times R$ with the multiplication rule $(a, b) \cdot (a', b') = (a \cdot a', a \cdot b' + a' \cdot b)$, this becomes an R-algebra isomorphism.

Axiom 43.2. For every $g : \mathbb{D}^k \to R$, there exist unique elements $\{b_1, \dots, b_k\}$ in R such that

$$f(d) = f(0) + \sum_j d^j \cdot b_j \tag{43.8}$$

for all $d \in \mathbb{D}^k$.

Corollary 43.3.8. Let $f : R \to R$ and $d \in D_k$.

$$f(x + d) = f(x) + d \cdot f'(x) + \cdots + \frac{d^k}{k!} \cdot f^{(k)}(x) \tag{43.9}$$

Chapter 44

Higher-dimensional Geometry ♣

In this chapter, certain constructions and theorems introduced in the previous chapters are generalized to the setting of supergeometries and higher categories. As such, it can been as an analogue to Chapter 4 for (differential) geometry.

The main references are Baez and Schreiber (2005); Schreiber (2005). The section on spectral geometry is based on Sanders (2012), whereas that on smooth spaces is inspired by Baez and Hoffnung (2011). For an introduction to (higher) category theory, see Chapter 3. Section 4.9.2 gives a different approach to the higher-dimensional analogues of Lie algebras. For an introduction to supergeometry, see Cattaneo and Schätz (2011). The notion of cohesion was introduced in Lawvere (1994).

@@ CITE MORE (e.g. BARTELS, ...) @@

Chapter Contents

44.1 Graded manifolds **806**

44.2 Infinite-dimensional geometry **808**

44.3 Cohesion **809**

44.4 Smooth spaces **813**

44.4.1 Frölicher spaces 814

44.4.2 Smooth sets 815

44.4.3 Smooth algebras 819

44.4.4 Supergeometry 819

44.5 Higher Lie theory **824**

44.6 Gauge theory ♣ **824**

44.6.1 Lie groupoids 825

44.7 Weak groups **826**

44.7.1 Spaces 828
44.8 2-Bundles **829**
44.9 Space and quantity **831**

44.1 Graded manifolds

In this section, some of the notions from Part VI will be generalized to supermanifolds and general graded manifolds. The general notation (x^i) will be used for the collection of both even and odd coordinates.

Example 44.1.1 (Supermanifold). A super smooth set in the form of a locally ringed space $(M, \mathcal{A})$ that is locally isomorphic to a super Euclidean space, i.e. $\mathcal{A}$ is locally given by $C^\infty(U) \otimes \Lambda^\bullet \mathbb{R}^n$ for some $n \in \mathbb{N}$ and open $U \subseteq M$. More generally, a **graded manifold** is a locally ringed space modelled on $\left(\mathbb{R}^m, C^\infty(\mathbb{R}^m) \otimes \mathsf{Sym}(V^*)\right)$ for a graded vector space V. (A supermanifold can be recovered by taking $V = \mathbb{R}^{0|n}$.)

Theorem 44.1.2 (Batchelor). *Let $(M, \mathcal{A})$ be an $\mathbb{N}$-graded manifold. There exists a vector bundle $E \to M$ such that $\mathcal{A}$ is isomorphic to the structure sheaf $\mathcal{O}(\Lambda^\bullet E)$, i.e. $\mathcal{A}$ is locally given by $\mathsf{Sym}^\bullet(\Lambda^\bullet E^*)$. If $(M, \mathcal{A})$ is a supermanifold, there exists a vector bundle $E \to M$ such that $\mathcal{A}$ is locally given by $\Lambda^\bullet E^*$.*

Example 44.1.3 (Odd tangent bundle). Consider a smooth manifold M. By the associated bundle construction and functoriality, any endofunctor on **Set** (or any concrete category) can turn an existing bundle into a new one. Starting from the tangent bundle TM and applying the parity functor (Definition 23.5.7) gives rise to the odd tangent bundle ΠTM. As a locally ringed space, it is given by the de Rham sheaf $\Omega^\bullet$.

More generally, for any vector bundle $E \to B$, parity reversal gives a supermanifold with structure sheaf $\mathcal{O}_{\Pi E} := \Lambda^\bullet_{C^\infty(B)} \Gamma_{E^*}$.

Definition 44.1.4 (Vector field). A graded vector field of degree $k \in \mathbb{N}$ is a degree-k derivation on $C^\infty(M)$. The integer k is called the **degree**.

Definition 44.1.5 (Cohomological vector field). A graded vector field X of degree 1 that satisfies $[X, X] = 0$. Every degree-1 graded vector field satisfies

$$[X, X] = 2X \circ X\,. \tag{44.1}$$

This implies that every cohomological vector field defines a coboundary operator on $C^\infty(M)$. A graded manifold equipped with a cohomological vector field is called a **differential-graded manifold** (dg-manifold).

Example 44.1.6 (de Rham differential). Consider the de Rham complex $\Omega^\bullet(M)$ with differential d. This differential corresponds to a cohomological vector field Q on ΠTM,

locally given by

$$Q := \sum_{i=1}^{n} \mathrm{d}x^i \partial_i \,. \tag{44.2}$$

Note that the differentials $\mathrm{d}x^i$ are here to be regarded as coordinate functions on ΠTM. Moreover, the order of the two factors in the definition of Q is important!

Definition 44.1.7 (Degree). Let M be a graded manifold. The degree of a homogeneous element of $\Omega^\bullet(M)$ is defined as the difference of its graded degree and its form degree.

Property 44.1.8 (Euler vector field). Consider the graded vector field

$$E := \sum_{i=1}^{n} \deg(x^i) x^i \partial_i \,. \tag{44.3}$$

The Lie derivative $\mathcal{L}_E$, defined through the Cartan formula

$$\mathcal{L}_E := \iota_E \mathrm{d} + (-1)^{\deg(E)} \mathrm{d} \iota_E \,, \tag{44.4}$$

acts on homogeneous forms by multiplication by their degree.

Definition 44.1.9 (Poisson manifold). Consider a degree-k symplectic form ω. This form induces a Poisson structure on the algebra $C^\infty(M)$ as follows:

$$\{f, g\} := (\partial_i^R f) \omega^{ij} (\partial_j^L g) \,. \tag{44.5}$$

It is not hard to check that this operation is graded commutative. As in Definition 37.1.3, given a symplectic form, a Hamiltonian vector field can be defined for any smooth function $H \in C^\infty(M)$:

$$\omega(X_H, \cdot) = -\mathrm{d}H(\cdot) \,. \tag{44.6}$$

Property 44.1.10. Every closed differential form of degree $k \neq 0$ is exact. More generally, the de Rham cohomology of a graded manifold is isomorphic to the de Rham cohomology of its body. This, for example, implies that a degree-l symplectic vector field X is Hamiltonian with respect to a degree-k symplectic form if $k + l \neq 0$.

Corollary 44.1.11 (dg-symplectic manifold). Consider a Hamiltonian cohomological vector field X. There exists a Hamiltonian function H such that

$$Xf = \{H, f\} \tag{44.7}$$

for all $f \in C^\infty(M)$. If the symplectic form has degree $k \in \mathbb{N}$, the function H can be chosen to be of degree $k + 1$ and, accordingly, $\{H, H\}$ will be of degree $k + 2$. Now, the identity $[X, X] = 0$ also implies that $\{H, H\}$ is a constant and, since all constants are of degree 0, it follows that

$$\{H, H\} = 0 \tag{44.8}$$

whenever $k \neq -2$. This equation is often called the **classical master equation**. A graded manifold equipped with both a symplectic form and a symplectic cohomological vector field is called a **differential-graded symplectic manifold**.

If ω is of degree 1, it was shown by *Schwarz* that (M, ω) is symplectomorphic to ΠT^*M such that the Poisson bracket is mapped to the Schouten–Nijenhuis bracket (Definition 34.3.24) and the Hamiltonian is mapped to a Poisson bivector field exactly if it satisfies the master equation.

44.2 Infinite-dimensional geometry

In many situations, when considering function spaces, the objects under consideration do not form a finite-dimensional manifold. However, with some care, one can omit this size condition. In Chapter 25, it was shown how calculus could be extended from $\mathbb{R}^n$ to infinite-dimensional vector spaces. Here, geometry is extended to that setting.

The first approach uses a locally convex TVS as local model space.

Definition 44.2.1 (E-manifold). Let E be a locally convex TVS. A Hausdorff space is called an E-manifold if there exists an atlas of charts (U, φ), where $\varphi : U \to \varphi(U) \subseteq E$ is a homeomorphism and the transition maps are Gateaux-smooth.

Definition 44.2.2 (Kinematic tangent bundle). Let M be an E-manifold with a smooth atlas $\{(U_i, \varphi_i)\}_{i \in I}$. The kinematic tangent bundle of M is defined as the quotient of

$$\bigsqcup_{i \in I} U_i \times E \tag{44.9}$$

by the equivalence relations[1] $(x, v) \sim \big(x, \mathrm{d}\psi_{ji}(\varphi_i(x); v)\big)$, where, as in the finite-dimensional case, $\psi_{ji} : U_i \cap U_j \to \mathrm{Aut}(E)$ are the transition functions.

Remark 44.2.3. For infinite-dimensional E, this tangent bundle is not isomorphic to the definition in terms of derivations. The above construction is 'kinematical' because the pair (x, v) represent a vector tangent to a curve at the point $x \in M$.

Since infinite-dimensional vector spaces are, in general, not reflexive, simply defining the cotangent bundle to be the fibrewise dual of the kinematic tangent bundle would lead to even more size issues. *Kriegl* and *Michor* have shown that one can cook up to 12 sensible definitions of a cotangent bundle (this also includes 'operational' definitions using derivations). However, only one of these definitions is well behaved with respect to Lie derivatives, exterior derivatives and pullbacks. Luckily, this is also the most widely used definition in the finite-dimensional setting.

[1] Here, d denotes the Gateaux differential and not the de Rham differential.

Definition 44.2.4 (Kinematic cotangent bundle). Let M be an E-manifold. Consider the set of bounded, alternating linear maps $E^{\times k} \to \mathbb{R}$. This lifts to a vector bundle $L^k_{\text{alt}}(TM, M \times \mathbb{R})$.

44.3 Cohesion

In this section, the terminology (Grothendieck) topos **over** a topos $\mathcal{S}$ (or $\mathcal{S}$-topos) will mean a topos equipped with a geometric morphism to $\mathcal{S}$. Moreover, everything will be stated in terms of ∞-topoi, unless stated otherwise. All notions such as adjoints, limits, etc. should be understood in this ∞-sense. The language of Section 15.6.2 is adopted.

Definition 44.3.1 (Local topos). An $\mathcal{S}$-topos $\mathbf{H}$ is said to be ($\mathcal{S}$-)local if the geometric morphism $(f^* \dashv f_*) : \mathbf{H} \leftrightarrows \mathcal{S}$ admits a right adjoint $f^!$ such that one of the following equivalent statements holds:

- $f^!$ is fully faithful,
- f^* is fully faithful,
- $*$ $f^!$ is an $\mathcal{S}$-indexed functor (Definition 14.1.15), or
- $*$ $f^!$ is Cartesian closed (Definition 4.1.23),

where the starred items hold in the 1-categorical setting. If $\mathcal{S} = \mathbf{Set}$ or $\mathcal{S} = \infty\mathbf{Grpd}$, the conditions are automatically satisfied. The triple of functors $f^* \dashv f_* \dashv f^!$ is also called a **local geometric morphism**.

The right adjoint is sometimes called the **codiscrete object functor** coDisc (in fact, this terminology is applied more generally when $\mathbf{H}$ is just any category). If this functor exists, $\mathbf{H}$ is said to have **codiscrete objects**.

Property 44.3.2. A topos is local if and only if $*$ is tiny (Definition 3.4.48).

Definition 44.3.3 (Locally connected topos). An object in a category is said to be **connected** if its representable functor preserves finite coproducts. A topos is said to be **locally connected** if all objects can be written as coproducts of connected objects. This defines a functor

$$\Pi_0 : \mathbf{H} \to \mathbf{Set} : \bigsqcup_{i \in I} X_i \mapsto I\,, \tag{44.10}$$

left-adjoint to the **discrete object functor** Disc, the inverse image part of the global section geometric morphism. This functor is suitably called the **connected components functor**.

Property 44.3.4. A topos is locally connected if and only if its global section geometric morphism is essential. More generally, an $\mathcal{S}$-topos is said to be **locally connected** if its

associated geometric morphism is essential and the left adjoint is $\mathcal{S}$-indexed. In the case of $(\infty, 1)$-topoi, the image of the functor Π_0 is called the **fundamental ∞-groupoid**.

Definition 44.3.5 (Connected topos). A topos is said to be **connected** if the inverse image part of the associated geometric morphism is fully faithful. For sheaf topoi over a topological space X, this is exactly equivalent to X being connected.

For locally connected topoi, this amounts to the property that the left adjoint in its adjoint triple preserves the terminal object. Furthermore, a locally connected topos is said to be **strongly connected** if the left adjoint in its adjoint triple preserves finite products (in particular, turning it into a connected topos).

Property 44.3.6. Every local topos is connected.

Definition 44.3.7 (Cohesive topos). A local, strongly connected topos. This implies the existence of an adjoint quadruple $\Pi_0 \dashv \mathsf{Disc} \dashv \Gamma \dashv \mathsf{coDisc}$, where both Disc and coDisc are fully faithful:

$$\begin{tikzcd} \mathbf{H} \arrow[r, shift left=4, "\Pi_0"] \arrow[r, shift left=2, "\Gamma"] & \infty\mathbf{Grpd} \arrow[l, hook', shift right=2, "\mathsf{Disc}"'] \arrow[l, hook', shift right=4, "\mathsf{coDisc}"'] \end{tikzcd} \tag{44.11}$$

Property 44.3.8 (Cohesive modalities). The adjoint quadruple of a cohesive topos induces an adjoint triple of modalities (Definition 3.3.27):

$$(\textstyle\int \dashv \flat \dashv \sharp) := (\mathsf{Disc} \circ \Pi_0 \dashv \mathsf{Disc} \circ \Gamma \dashv \mathsf{coDisc} \circ \Gamma)\,. \tag{44.12}$$

These are respectively called the **shape**, **flat** and **sharp** modalities. The modal types of the flat and sharp modalities are called the **discrete** and **codiscrete objects**, respectively.

Remark 44.3.9 (Unity of opposites). Recall Example 15.6.10. A similar remark holds for the adjunction $\flat \dashv \sharp$. The induced unity of opposites

$$\mathsf{quantity}_X : \flat X \to X \to \sharp X \tag{44.13}$$

represents *Cantor*'s '*Kardinalen*' or *Hegel*'s '*Quantität*', which is the unity of continuity and discreteness.

Definition 44.3.10 (Concrete object). Consider a cohesive topos $\mathbf{H}$. An object $X \in \mathsf{ob}(\mathbf{H})$ is said to be concrete if $\eta^\sharp_X$ is monic.

Property 44.3.11 (Nullstellensatz). Consider a cohesive topos $\mathbf{H}$. The adjoint quadruple induces the following unity of opposites:

$$\mathsf{ptp} : \flat \Rightarrow \textstyle\int := \varepsilon^{\int} \circ \eta^{\flat}\,.$$

This is called the **points-to-pieces** transform, representing the unity between 'repulsion' and 'attraction' (or cohesion). Sometimes, the points-to-pieces transform is also considered to be the adjunct $\Gamma \Rightarrow \Pi_0$. If this transformation is an epi for all objects $X \in \mathsf{ob}(\mathbf{H})$, the Nullstellensatz is said to hold (or pieces are said to have points).

Property 44.3.12. In a cohesive topos $\mathbf{H}$, the following statements are equivalent:

- The Nullstellensatz holds.
- Discrete objects are concrete, i.e. $\eta^{\sharp}_{\flat X}$ is a mono for all $X \in \mathsf{ob}(\mathbf{H})$.
- $\flat \dashv \sharp$ exhibits Aufhebung (Definition 15.6.9) with respect to $\emptyset \dashv *$, i.e. the initial object is codiscrete.

Moreover, if $\mathbf{H}$ is the sheaf topos over a *cohesive site* $\mathbf{C}$, these statements hold as soon as every object in $\mathbf{C}$ has at least one global element.

Definition 44.3.13 (Differential cohesion). Consider a cohesive $(\infty, 1)$-topos $\mathbf{H}$. An **infinitesimal cohesive neighbourhood** of $\mathbf{H}$ is a cohesive $(\infty, 1)$-topos $\mathbf{H}_{\text{th}}$ equipped with a strongly connected adjoint quadruple $(\iota_{\text{inf}} \dashv \Pi_{\text{inf}} \dashv \mathsf{Disc}_{\text{inf}} \dashv \Gamma_{\text{inf}}) : \mathbf{H} \to \mathbf{H}_{\text{th}}$ such that ι_{inf} is also fully faithful (and, hence, so is $\mathsf{Disc}_{\text{inf}}$). If such a neighbourhood exists, $\mathbf{H}_{\text{th}}$ is said to be differentially cohesive (or **elastic**) over $\mathbf{H}$.[2]

Remark 44.3.14. The interpretation of the infinitesimal adjoint quadruple is slightly different from that in the definition of cohesion. Here, $\Pi_0 \dashv \mathsf{Disc} \dashv \Gamma$ corresponds to the triple $\Pi_{\text{inf}} \dashv \mathsf{Disc}_{\text{inf}} \dashv \Gamma_{\text{inf}}$. Moreover, the adjoint triple $\Pi_{\text{th}} \dashv \mathsf{Disc}_{\text{th}} \dashv \Gamma_{\text{th}}$ characterizing $\mathbf{H}_{\text{th}}$ as a cohesive topos factors exactly through these two triples by uniqueness of adjoint and global sections (see Eq. (44.16) for the full structure):

$$\begin{tikzcd} \mathbf{H}_{\text{th}} \arrow[r, shift left=3, "\Pi_{\text{inf}}"] \arrow[r, shift right=3, "\Gamma_{\text{inf}}"'] & \mathbf{H} \arrow[l, hook', "\mathsf{Disc}_{\text{inf}}"'] \arrow[r, shift left=3, "\Pi_0"] \arrow[r, shift right=3, "\Gamma"'] & \infty\mathbf{Grpd} \arrow[l, hook', "\mathsf{Disc}"'] \end{tikzcd} = \begin{tikzcd} \mathbf{H}_{\text{th}} \arrow[r, shift left=3, "\Pi_{\text{th}}"] \arrow[r, shift right=3, "\Gamma_{\text{th}}"'] & \infty\mathbf{Grpd} \arrow[l, hook', "\mathsf{Disc}_{\text{th}}"'] \end{tikzcd} \tag{44.14}$$

Property 44.3.15 (Differential modalities). Consider a differentially cohesive $(\infty, 1)$-topos $\mathbf{H}$. The adjoint quadruple to its infinitesimal thickening induces an adjoint cylinder of modalities:

$$(\mathfrak{R} \dashv \mathfrak{I} \dashv \&) := (\iota_{\text{inf}} \circ \Pi_{\text{inf}} \dashv \mathsf{Disc}_{\text{inf}} \circ \Pi_{\text{inf}} \dashv \mathsf{Disc}_{\text{inf}} \circ \Gamma_{\text{inf}}).$$

There are called the **reduction**, **infinitesimal shape** and **infinitesimal flat** modalities, respectively. The modal types of $\mathfrak{R}$ and $\mathfrak{I}$ are said to be **reduced** and **coreduced**, respectively.

[2]The subscript 'th' indicates that $\mathbf{H}_{\text{th}}$ can be seen as an 'infinitesimal thickening' of $\mathbf{H}$. (See Definition 44.4.31 for more information.)

The modalities coming from being cohesive and differentially cohesive fit into the following diagram, where $\vee$ indicates inclusion of modal types:

$$\begin{array}{ccccccc} \Re & \dashv & \Im & \dashv & \& & & \\ & & \vee & & \vee & & \\ & & \int & \dashv & \flat & \dashv & \sharp \\ & & & & \vee & & \vee \\ & & & & \emptyset & \dashv & * \end{array} \tag{44.15}$$

Definition 44.3.16 (Jet comonad). Consider a differentially cohesive topos $\mathbf{H}$. The jet bundle $J^\infty E$ of a bundle $E \in \mathsf{ob}(\mathbf{H}/B)$ over some base space $B \in \mathsf{ob}(\mathbf{H})$ is given by the base-change comonad (Definition 12.2.1) along the unit $\eta^\Im_B : B \to \Im B$:[3]

$$J^\infty E := (\eta^\Im_B)^*(\eta^\Im_B)_* E\,.$$

Analogously, the **infinitesimal (or formal) disk bundle** is defined as the adjoint monad induced by the base-change triple:

$$T^\infty E := (\eta^\Im_B)^*(\eta^\Im_B)_! E\,.$$

By definition of the base-change triple and the pasting laws (Property 3.4.73), this is equivalent to $T^\infty B \times_B E$, where the infinitesimal disk bundle $T^\infty B$ of an object is defined as the pullback of the counit $\eta^\Im_B$ along itself.

If $\Im$ is interpreted as identifying all infinitesimally close points, the fibre over each (global) point of the infinitesimal disk bundle can indeed be interpreted as the infinitesimal disk around that point (again by the pasting laws).

The (co)algebras over the jet comonad correspond to common objects in geometry. The Eilenberg–Moore category $\mathbf{H}^{J^\infty}$ corresponds to PDEs with variables in B and the Kleisli category $\mathbf{Kl}(J^\infty)$ corresponds to bundles over B equipped with a differential operator.

In analogy to Definition 6.2.14 and Definition 10.2.31, one can also define local diffeomorphisms/étale morphisms between formal smooth sets.

Definition 44.3.17 (Formally étale morphism[4]). A morphism $f : X \to Y$ in a differentially cohesive topos $\mathbf{H}$ such that the naturality square of the shape modality is a

[3] Sometimes, the object $(\eta^\Im_B)_* E \to \Im B$ is called the jet bundle of $E \to B$.
[4] Also called a **local diffeomorphism**.

pullback square:

$$\begin{tikzcd} X \arrow[r, "\eta_X"] \arrow[d, "f"'] \arrow[dr, phantom, "\mathrm{pb}"] & \mathfrak{J}X \arrow[d, "\mathfrak{J}f"] \\ Y \arrow[r, "\eta_Y"'] & \mathfrak{J}Y\,. \end{tikzcd}$$

Definition 44.3.18 (Super-differential cohesion). Consider a differentially cohesive $(\infty,1)$-topos $\mathbf{H}_{\mathrm{bos}} \to \mathbf{H}_{\mathrm{red}}$. A super-differentially cohesive (or **solid**[5]) topos over $\mathbf{H}_{\mathrm{bos}}$ is a differentially cohesive $(\infty,1)$-topos $\mathbf{H} \to \mathbf{H}_{\mathrm{red}}$ equipped with an adjoint quadruple $(\mathsf{even} \dashv \iota_{\mathrm{sup}} \dashv \Pi_{\mathrm{sup}} \dashv \mathsf{Disc}_{\mathrm{sup}}) : \mathbf{H} \to \mathbf{H}_{\mathrm{bos}}$ such that ι_{sup} (and, hence, also $\mathsf{Disc}_{\mathrm{sup}}$) is fully faithful.

These structures give the following diagram of $(\infty,1)$-functors (again by uniqueness of adjoints and global sections):

$$\begin{tikzcd} \mathbf{H} \arrow[r, shift left=4, "\mathsf{even}"] \arrow[r, shift left=1, "\Pi_{\mathrm{sup}}"] \arrow[r, shift right=1, "\Gamma_{\mathrm{sup}}"'] \arrow[rr, shift right=6, "\mathsf{coDisc}"'] & \mathbf{H}_{\mathrm{bos}} \arrow[l, hook', shift right=4, "\iota_{\mathrm{sup}}"'] \arrow[l, hook', shift left=1, "\mathsf{Disc}_{\mathrm{sup}}"] \arrow[r, shift left=1, "\Pi_{\mathrm{bos}}"] \arrow[r, shift right=1, "\Gamma_{\mathrm{bos}}"'] & \mathbf{H}_{\mathrm{red}} \arrow[l, hook', shift left=1, "\iota_{\mathrm{bos}}"] \arrow[l, hook', shift right=1, "\mathsf{Disc}_{\mathrm{bos}}"'] \arrow[r, shift left=1, "\Pi_{\mathrm{red}}"] \arrow[r, shift right=1, "\Gamma_{\mathrm{red}}"'] & \infty\mathbf{Grpd} \arrow[l, hook', shift left=1, "\mathsf{Disc}_{\mathrm{red}}"] \end{tikzcd} \tag{44.16}$$

Property 44.3.19 (Solid modalities). As for cohesive and differentially cohesive topoi, the adjoint functors induce a set of (adjoint) modalities:

$$(\rightrightarrows \dashv \rightsquigarrow \dashv \mathsf{Rh}) := (\iota_{\mathrm{sup}} \circ \mathsf{even} \dashv \iota_{\mathrm{sup}} \circ \Pi_{\mathrm{sup}} \dashv \mathsf{Disc}_{\mathrm{sup}} \circ \Pi_{\mathrm{sup}})\,. \tag{44.17}$$

These are called the **fermionic**, **bosonic** and **rheonomic** modalities, respectively. Note that, by Diagram 44.16, a super-differentially cohesive topos also inherits the cohesive and differential modalities.

@@ COMPLETE (e.g. work by Schreiber) @@

44.4 Smooth spaces

In this section, some generalizations of spaces that are better behaved when considering their properties as a whole are introduced. Before moving to the smooth setting, a bit

[5]This terminology stems from the fact that supergeometry is necessary for treating fermions, which are in turn responsible for condensed matter physics and, hence, the study of solid matter.

of history will be given, starting from the ordinary topological setting.

The first problem in the study of the global properties of spaces arose in algebraic topology. When consider mapping spaces, it is sometimes useful to use the currying operation

$$C(X \times Y, Z) \to C\big(X, C(Y, Z)\big)\,. \tag{44.18}$$

However, in general, this is not a homeomorphism, i.e. currying does not define an adjunction and, therefore, **Top** is not Cartesian closed (Definition 4.1.20). This problem was treated by *Steenrod* and others, and the solution was simply to restrict to a smaller class of better behaved spaces: the compactly generated Hausdorff spaces.[6]

Whilst studying varieties in algebraic geometry, people experienced similar problems. For this reason, *Grothendieck* invented schemes (see Chapter 10 and Section 10.2 in particular). The main takeaway of this approach was that "it's better to work with a nice category containing some nasty objects, than a nasty category containing only nice objects" as *Baez* phrases it succinctly.

The category **Diff** of finite-dimensional smooth manifolds suffers the same problems, namely the space of smooth functions $C^\infty(X, Y)$ is, in general, some kind of infinite-dimensional manifold and, hence, cannot be defined internally. It becomes even worse if one studies the mapping spaces between those. *Kriegl* and *Michor* have introduced a framework in which one can work safely, but the main problem with their solution is that not all spaces of interest are included. Certain other operations such as quotients and (co)limits are also not guaranteed to exist within that category.

44.4.1 Frölicher spaces

Definition 44.4.1 (Frölicher space). A triple (X, C_X, F_X) consisting of the following data:

- a set X,
- a set of curves $C_X \subseteq \mathbf{Set}(\mathbb{R}, X)$, and
- a set of functionals $F_X \subseteq \mathbf{Set}(X, \mathbb{R})$.

These are required to satisfy the following closure and consistency conditions:

1. For all $c \in C_X$ and $f \in F_X$, the composite is smooth: $f \circ c \in C^\infty(\mathbb{R})$.
2. If, given $c \in \mathbf{Set}(\mathbb{R}, X)$, the composite $f \circ c \in C^\infty(\mathbb{R})$ for all functionals $f \in F_X$, then $c \in C_X$.

[6] In practice, this is not a problem since all interesting spaces, such as CW complexes, belong to this class.

3. If, given $f \in \mathbf{Set}(X, \mathbb{R})$, the composite $f \circ c \in C^\infty(\mathbb{R})$ for all curves $c \in C_X$, then $f \in F_X$.

Morphisms of Frölicher spaces are those functions that preserve curves, functionals and smoothness under composition.

Property 44.4.2. The category of Frölicher spaces is bicomplete and Cartesian closed.

Definition 44.4.3 (Topology). Consider a Frölicher space (X, C_X, F_X). The **curvaceous topology** on X is the final topology with respect to the curves C_X. The **functional topology** on X is the initial topology with respect to the functionals F_X.

@@ COMPLETE (e.g. Isbell enveloppe / duality) @@

44.4.2 Smooth sets

Definition 44.4.4 (Concrete site). Consider a site $(\mathbf{C}, J)$. It is said to be concrete if

1. the functor $\mathbf{C}(*, -) : \mathbf{C} \to \mathbf{Set}$ is faithful, and
2. for every covering family $\{f_i : U_i \to U\}_{i \in I} \in J(U)$, the morphism

$$\bigsqcup_{i \in I} \mathbf{C}(*, f_i) : \bigsqcup_{i \in I} \mathbf{C}(*, U_i) \to \mathbf{C}(*, U) \tag{44.19}$$

is surjective.

Definition 44.4.5 (Concrete presheaf). Consider a concrete site $(\mathbf{C}, J)$. For every presheaf $X \in \mathbf{Psh}(\mathbf{C})$ and object $U \in \mathsf{ob}(\mathbf{C})$, denote by

$$X_U : X(U) \to \mathbf{Set}\big(\mathbf{C}(*, U), X(*)\big) \tag{44.20}$$

the adjunct of the restriction morphism

$$X(U) \times \mathbf{C}(*, U) \to X(*)\,. \tag{44.21}$$

The presheaf X is said to be concrete if, for each $U \in \mathsf{ob}(\mathbf{C})$, the map X_U is injective. This says that concrete presheaves are subobjects of presheaves of the form

$$U \mapsto \mathbf{Set}\big(\mathbf{C}(*, U), X(*)\big)\,. \tag{44.22}$$

This definition also agrees with Definition 44.3.10.

Property 44.4.6. The category of sheaves on a concrete site is a local topos (Definition 44.3.1). The right adjoint $\mathsf{coDisc} : \mathbf{Set} \to \mathbf{H}$ sends a set A to the functor $U \mapsto \mathbf{Set}\big(\mathbf{C}(*, U), A\big)$.

This property allows to define concrete objects in any local topos in an alternative way (by slightly modifying Property 12.2.10 to allow for separation instead of locality).

Definition 44.4.7 (Concrete object). Consider a local geometric morphism $\Gamma : \mathbf{H} \to \mathcal{S}$. Consider the class V of local isomorphisms with respect to Γ. An object in $\mathbf{H}$ is said to be concrete if it is V-separated, i.e. its Yoneda embedding maps morphisms in V to monos.

Remark 44.4.8. Note that concrete sheaves carry two kinds of sheaf-theoretic information. On the one hand they are actual sheaves with respect to the Grothendieck topology of the underlying site and, on the other hand, they are separated presheaves with respect to the topology generated by global elements $* \to U$.

Property 44.4.9 (Concretification). Consider a local 1-topos $\mathbf{H}$.[7] Every topos has an epi/mono factorization giving rise to the image $\mathsf{im}(f)$ of a morphism $f \in \mathsf{Hom}(\mathbf{H})$. The concretification of an object $X \in \mathsf{ob}(\mathbf{H})$ is given by the image $\mathsf{im}(\eta^\sharp_X)$.

In the case of the site of Cartesian spaces, the following notion is obtained.

Definition 44.4.10 (Diffeological space). A diffeology on a set X is a concrete sheaf $\mathcal{D}_X$ on $\mathbf{CartSp}_{\text{diff}}$ such that $\Gamma\mathcal{D}_X = X$.

Alternative Definition 44.4.11 (Diffeological space). Let X be a set. A diffeology $\mathcal{D}_X$ on X is defined as a collection of functions $f : U \subseteq \mathbb{R}^n \to X$, called **plots**, satisfying the following conditions (where U, V and W are open sets):

1. $\mathcal{D}_X$ contains all constant functions.
2. If $\{U_i\}_{i\in I}$ is an open cover of U and if $f|_{U_i} \in \mathcal{D}_X$ for all $i \in I$, then $f \in \mathcal{D}_X$.
3. If $f \in \mathcal{D}_X$ and $g : W \subseteq \mathbb{R}^m \to \mathsf{dom}(f)$ is smooth, then $f \circ g \in \mathcal{D}_X$.

The set X can be turned into a topological space by equipping it with the $\mathcal{D}_X$**-topology**, the final topology with respect to $\mathcal{D}_X$.

Remark 44.4.12. Note that, in contrast to ordinary manifolds, the plots in a diffeology can have domains of different dimensions.

Definition 44.4.13 (Smooth map). Let $\mathcal{D}_X$ and $\mathcal{D}_Y$ be diffeological spaces. A map $g : X \to Y$ is said to be smooth if $g \circ f \in \mathcal{D}_Y$ for all $f \in \mathcal{D}_X$. The diffeological spaces together with their differentiable morphisms form a category $\mathbf{DiffSp}$.

Definition 44.4.14 (Chen space). If the open sets in the definition of a diffeological space are replaced by convex sets, the notion of smooth spaces due to *Chen* is obtained.

Alternative Definition 44.4.15 (Manifold). A diffeological space that is locally diffeomorphic to a Euclidean space. A map between manifolds is smooth in the diffeological sense if and only if it smooth in the sense of Definition 31.1.10.

[7] The restriction to ordinary topoi ensures that ordinary image factorization suffices. For higher topoi, this would not be the case and higher image factorizations would be required.

There exist two trivial smooth structures.

Example 44.4.16 (Discrete structure). The minimal smooth structure, i.e. the one obtained by taking the plots to be the constant functions.

Example 44.4.17 (Indiscrete structure). The maximal smooth structure, i.e. the one obtained by taking all functions to be plots.

The notion of diffeological spaces can be further generalized by passing to the full sheaf topos on Cartesian spaces.

Definition 44.4.18 (Smooth set).

$$\mathbf{SmoothSet} := \mathbf{Sh}(\mathbf{CartSp}_{\mathrm{diff}}) \tag{44.23}$$

The topology on this site is generated by the coverage of differentiably good covers (Definition 31.1.16). In fact, this topology coincides with the usual one consisting of open covers. The category of smooth sets is also denoted by $\mathbf{C}^\infty$.

To phrase this in the sense of Section 44.3, smooth sets form a cohesive topos over **Set**, where Disc and coDisc assign the discrete and codiscrete structures from the preceding examples, respectively. Moreover, this cohesive topos satisfies the *Nullstellensatz* 44.3.12.

Property 44.4.19. There exists an adjunction

$$\mathbf{Top} \underset{\mathrm{diff}}{\overset{\mathrm{top}}{\underset{\longrightarrow}{\overset{\longleftarrow}{\perp}}}} \mathbf{C}^\infty . \tag{44.24}$$

The right adjoint endows a topological space X with the smooth structure for which every continuous map $U \to X$ is a plot. Its left adjoint sends a smooth space to the topological space equipped with the finest topology for which all plots become continuous maps.

Property 44.4.20. The adjunction $\Pi_0 \dashv \mathrm{Disc}$ exhibits the discrete smooth sets, i.e. the constant sheaves, as the reflective subcategory on $\mathbb{R}$-local objects (see also Property 13.2.8).

As noted above, the subcategory of diffeological spaces exactly consists of the concrete smooth sets, i.e. those smooth sets that have an underlying set of points. A common example of smooth sets that do not have such an underlying set is given by de Rham spaces.

Example 44.4.21 (Differential forms). Consider the kth de Rham functor Ω^k on the category **Diff**. This functor assigns to every smooth manifold its space of differential k-forms. Locally defined forms can be glued together if they agree on intersections, i.e. they satisfy the sheaf condition. This shows that Ω^k defines a smooth set, albeit one that is far from an ordinary smooth manifold. It is called the **(universal) moduli**

space of differential k-forms. Note, however, that for a given smooth set $X \in \mathbf{C}^\infty$, the plots of $[X, \Omega^k]$ correspond to differential k-forms on the product $\mathbb{R}^n \times X$ and not to $\mathbb{R}^n$-parametrized forms on X. For the latter, one should consider the concretification $\mathsf{conc}[X, \Omega^k]$.

One can go even further and characterize specific geometric structures in terms of this functor. Consider for example the subfunctor Ω^2_{cl} that assigns closed two-forms to a manifold. This also defines a smooth space and, hence, one can consider the slice category $\mathbf{C}^\infty/\Omega^2_{\mathrm{cl}}$. It is not hard to show that the category **SpMfd** of symplectic manifolds admits an embedding into this slice category.

Property 44.4.22 (Universal differential forms). Consider the differential n-form

$$\omega_{\mathrm{univ}} \in \Omega^n(\Omega^n) \tag{44.25}$$

modulated by the identity morphism on Ω^n, the universal differential n-form. Every differential n-form, on any smooth set, can be obtained as a pullback of ω_{univ}.

Example 44.4.23 (Path space). Since every topos is Cartesian closed, all internal homs exist. The path space of a smooth set X is defined as

$$\mathbf{P}X := [\mathbb{R}, X]\,. \tag{44.26}$$

Its underlying set is given by the set of smooth plots (trajectories): $\Gamma[\mathbb{R}, X] \cong \mathbf{C}^\infty(\mathbb{R}, X)$.

Analogously, the internal hom out of X gives the moduli space of smooth functions on X:

$$\mathbf{C}^\infty(X) := [X, \mathbb{R}]\,. \tag{44.27}$$

Definition 44.4.24 (Transgression). Consider a smooth set $X \in \mathsf{ob}(\mathbf{C}^\infty)$. Transgression of differential forms on X along a compact k-manifold Σ is defined as the composite:

$$\tau_\Sigma \omega := \int_\Sigma [\Sigma, \omega] : [\Sigma, X] \to [\Sigma, \Omega^n] \xrightarrow{\int_\Sigma} \Omega^{n-k}\,. \tag{44.28}$$

An alternative approach uses a pullback along the evaluation morphism $\mathsf{ev}_\Sigma : \Sigma \times [\Sigma, X] \to X$.

Property 44.4.25. If the compact k-manifold Σ has a boundary, transgression of closed differential forms along Σ and $\partial\Sigma$ is related by the de Rham differential:

$$\tau_{\partial\Sigma}(\omega) = (-1)^{k+1} \mathsf{d} \tau_\Sigma \omega \tag{44.29}$$

for all $\omega : X \to \Omega^n_{\mathrm{cl}}$.

44.4.3 Smooth algebras

Definition 44.4.26 (Smooth algebra). For any smooth manifold M, the algebra of smooth functions can be obtained as a hom-object:

$$C^\infty(M) := C^\infty(M, \mathbb{R}) = \mathbf{Diff}(M, \mathbb{R}) . \tag{44.30}$$

Since hom-functors are (finite) product-preserving, the multiplication $C^\infty(M) \times C^\infty(M) \to C^\infty(M)$ can be seen to be induced by the multiplication on $\mathbb{R}$:

$$C^\infty(M, \mathbb{R} \times \mathbb{R}) \cong C^\infty(M) \times C^\infty(M) . \tag{44.31}$$

Furthermore, the hom-functor is covariant in the second argument and, hence, defines a copresheaf on the category $\mathbf{CartSp}_{\mathrm{diff}}$. Generalizing this situation, smooth algebras are defined as finite product-preserving copresheaves on $\mathbf{CartSp}_{\mathrm{diff}}$. This (functor) category is denoted by $\mathbf{C^\infty Alg}$.

Definition 44.4.27 (Underlying algebra). Given a smooth algebra $R \in \mathbf{C^\infty Alg}$, its underlying algebra $U(R)$ is defined as the set $R(\mathbb{R})$ equipped with the canonically induced ring operations.

Definition 44.4.28 (Finitely generated smooth algebra). Since ordinary R-algebras are finitely generated if and only if they are of the form $R[x_1, \dots, x_k]/I$ for some integer $k \in \mathbb{N}$ and some ideal I, a smooth algebra is said to be finitely generated if it is of the form $C^\infty(\mathbb{R}^n)/I$ for some $n \in \mathbb{N}$ and some ideal I in the underlying algebra.

Definition 44.4.29 (Smooth locus). Let $\mathbf{C^\infty Alg^{fin}}$ denote the category of finitely generated smooth algebras. The category of **smooth loci** is defined as $(\mathbf{C^\infty Alg^{fin}})^{\mathrm{op}}$. The smooth locus corresponding to a smooth algebra R is often denoted by ℓR.

44.4.4 Supergeometry

In this section, the definition of smooth spaces (and sets) is generalized to the odd (fermionic) sector, i.e. 'super smooth sets' will be defined. This gives an explicit example of a super-differentially cohesive $(\infty, 1)$-topos.

Definition 44.4.30 (Superscheme). The category of affine superschemes is defined as the opposite of the category of supercommutative superalgebras, the commutative monoids internal to $\mathbf{sVect}$:

$$\mathbf{Aff}(\mathbf{sVect}) := \mathbf{sCAlg}^{\mathrm{op}} := \mathbf{CMon}(\mathbf{sVect}) . \tag{44.32}$$

More generally, one can define affine schemes internal to any symmetric monoidal category.

Definition 44.4.31 (Infinitesimally thickened space). First, consider a point $\mathbb{R}^0$. Its infinitesimal thickening should be a space such that every function that vanishes at

the origin is actually nilpotent (this is essentially a version of the Kock–Lawvere axiom 43.1):

$$\mathbb{D} := \mathsf{Spec}(A)\,, \tag{44.33}$$

where $A := \mathbb{R} \oplus V$ for V a finite-dimensional nilpotent ideal.[8] A Euclidean space can be infinitesimally thickened by taking the product with $\mathbb{D}$ or, at the algebraic level, by taking the tensor product with A. A morphism of such spaces is defined by an $\mathbb{R}$-algebra morphism between their associated algebras. These form the category $\mathbf{FormalCartSp}_{\mathrm{diff}}$.

Definition 44.4.32 (Superpoint). A space of the form $\mathsf{Spec}(\Lambda^\bullet\mathbb{R}^n)$. This is often denoted by $\mathbb{R}^{0|n}$. The **super-Euclidean space** $\mathbb{R}^{m|n}$ is obtained as the product of an ordinary Euclidean space $\mathbb{R}^m$ and the superpoint $\mathbb{R}^{0|n}$, i.e. its algebra of smooth functions is $C^\infty(\mathbb{R}^m) \otimes \Lambda^\bullet\mathbb{R}^n$.[9]

Example 44.4.33 (First-order neighbourhood). For $A = \mathbb{R}[\varepsilon]/\varepsilon^2$, the definition of an infinitesimal thickening recovers the first-order infinitesimal neighbourhood of Definition 43.3.1:

$$\mathbb{D}^1 := \mathsf{Spec}(\mathbb{R}[\varepsilon]/\varepsilon^2)\,. \tag{44.34}$$

The morphism dual to the mapping implied by the Kock–Lawvere axiom 43.1 gives an inclusion map $\mathbb{D}^1 \hookrightarrow \mathbb{R}^1$. (This example can easily be generalized to k^{th}-order neighbourhoods.) This algebra is the algebra of functions on the even part of the superplane:

$$\mathcal{O}\big(\mathsf{even}\,\mathbb{R}^{0|2}\big) \cong \mathbb{R}[\varepsilon]/\varepsilon^2\,. \tag{44.35}$$

Property 44.4.34 (Morphisms). First, consider the morphisms from a Euclidean space into an infinitesimal neighbourhood $\mathbb{D}^k$. Since such morphisms are dual to algebra morphisms, one should consider morphisms of the form $\mathbb{R}[\varepsilon]/\varepsilon^{k+1} \to C^\infty(\mathbb{R}^n)$. However, being an algebra morphism implies that $f(1) = 1$ and that nilpotents are mapped to nilpotents. The algebra of smooth functions on a Euclidean space does not contain nilpotents and, hence, there exists a unique function into an infinitesimal neighbourhood (the one that factorizes through the one-point set).

For morphisms out of (first-order) infinitesimal neighbourhoods, one obtains the property, known from synthetic geometry, that morphisms of the form $\mathbb{R}^n \times \mathbb{D}^1 \to \mathbb{R}^n$ correspond to vector fields on $\mathbb{R}^n$ exactly when the postcomposition with the inclusion $\mathbb{R}^n \hookrightarrow \mathbb{R}^n \times \mathbb{D}^1$ is the identity.

Definition 44.4.35 (Formal smooth set). A sheaf on the site of infinitesimally thickened Euclidean spaces, where covers are of the form $\{U_i \times \mathsf{Spec}(A) \mid U_i \subseteq \mathbb{R}^n\}$, for A a commutative (hence, even) algebra. The category of formal smooth sets or, equivalently, the sheaf topos on $\mathbf{FormalCartSp}_{\mathsf{diff}}$ is also called the **Cahiers topos**. The sets

[8] Algebras of this form are also called **Weil algebras** or **local Artin algebras**.
[9] Compare this to the definition of supermanifolds (Example 44.1.1).

in the image of a formal smooth set X are called the sets of **plots** of X and can be interpreted as sets of functions into X (in analogy with the definition of smooth sets).

Definition 44.4.36 (Super smooth set). By combining infinitesimal thickenings and superpoints, the category $\mathbf{SuperFormalCartSp}_{\mathbf{diff}}$ is obtained. A sheaf on this category is called a super smooth set. Since *Dubuc* introduced the Cahiers topos, the category of super smooth sets is sometimes called the **super-Dubuc topos**.

Property 44.4.37 (Cohesion). The full subcategory inclusion $\mathbf{CAlg} \hookrightarrow \mathbf{sCAlg}$ is part of an adjoint cylinder, where the left and right adjoints are given by projection onto the even part and quotienting out the odd part:

$$\mathsf{even} \dashv \iota \dashv -/(-)_{\mathrm{odd}}\,, \tag{44.36}$$

where $(-)_{\mathrm{odd}}$ indicates that the ideal generated by odd elements is quotiented out, i.e. every term containing at least one odd factor is set to zero. These define canonical functors, even and Π_{sup}, from the category of affine superschemes to affine schemes (and, hence, also between the larger categories of superschemes and schemes). Moreover, the inclusion of affine schemes into infinitesimally thickened spaces also admits a right adjoint given by reduction: $\Pi_{\mathrm{inf}}(\mathbb{R}^n \times \mathbb{D}) := \mathbb{R}^n$.

These, in turn, induce a system of ajunctions between the sheaf topos of smooth sets, the Cahiers topos and the super-Dubuc topos, characterizing them as a cohesive, elastic and solid topoi, respectively:

$$\begin{tikzcd}
\textbf{Super Smooth Set} \arrow[r, shift left=4, "\mathsf{even}"] \arrow[r, shift right=4, "\Pi_{\mathrm{sup}}"'] \arrow[rr, shift right=8, "\Gamma_{\mathrm{sup}}"'] & \textbf{Formal Smooth Set} \arrow[l, hook, shift left=4, "\iota_{\mathrm{sup}}"] \arrow[r, shift left=4, "\Pi_{\mathrm{inf}}"] \arrow[r, shift right=4, "\Gamma_{\mathrm{inf}}"'] & \mathbf{C}^\infty \arrow[l, hook, shift left=4, "\iota_{\mathrm{inf}}"] \arrow[r, shift left=4, "\Pi_{\mathrm{red}}"] \arrow[r, shift right=4, "\Gamma_{\mathrm{red}}"'] & \textbf{Set} \arrow[l, hook, shift left=4, "\mathsf{Disc}_{\mathrm{sup}}"] \arrow[l, hook, shift right=4, "\mathsf{Disc}_{\mathrm{inf}}"'] \arrow[lll, hook, shift right=8, "\mathsf{coDisc}"'] \arrow[l, hook, shift right=4, "\mathsf{Disc}_{\mathrm{red}}"']
\end{tikzcd} \tag{44.37}$$

The top four functors in every column are obtained by Kan extension from the functors between affine spaces (Property 3.4.93). The others follow from uniqueness and composition properties of adjoints and the fact that the coverages on **SuperSmoothSet** and **FormalSmoothSet** are trivial along infinitesimal thickenings and odd dimensions.

By the general theory of cohesive topoi, the diagram of adjunctions induce a another diagram of adjoint modalities. These are listed below.

Definition 44.4.38 (Fermionic modalities). The adjoint cylinder induced by the inclu-

sion **CAlg** $\hookrightarrow$ **sCAlg** in turn gives rise to the adjoint modalities

$$\overrightarrow{\overrightarrow{(-)}} \dashv \overset{\rightsquigarrow}{(-)} . \tag{44.38}$$

The left adjoint sends a super smooth set to its 'bifermionic' space, where only paired fermions occur. The right adjoint sends a super smooth set to its underlying bosonic space and might, therefore, be called the **bosonic modality**.

Moreover, since these modalities come from an adjoint quadruple, a further right adjoint exists. The rheonomy modality corresponds to localization at the superpoint:

$$\mathsf{Rh} \cong L_{\mathbb{R}^{0|1}} , \tag{44.39}$$

akin to Property 44.4.20.

The following definition is dual to Definition 2.4.47.

Definition 44.4.39 (Reduction). By Property 44.3.15, **SuperSmoothSet** admits differential modalities to both **FormalSmoothSet** and **SmoothSet**. The reduction modality $\mathfrak{R}$ simply drops all infinitesimal directions, both even- and odd-graded:

$$\mathfrak{R}(\mathbb{R}^{m|n} \times \mathbb{D}) = \mathbb{R}^m . \tag{44.40}$$

The **infinitesimal neighbourhood** (to arbitrary order) of a formal smooth subset $Y \hookrightarrow X$ is defined by taking its plots to be those plots of X for which the reductions factorize through plots of Y.

Definition 44.4.40 (Shape modality). The **(infinitesimal) shape** or **de Rham shape** $\mathfrak{J}X$ of a super smooth set X is given, throught duality/adjointness, by the super smooth set obtained by reducing the plots of X:

$$\mathfrak{J}X(U) := X(\mathfrak{R}(U)) . \tag{44.41}$$

This construction identifies all infinitesimally close points.

Definition 44.4.41 (Tangent bundle). Consider a super smooth set X. Its tangent bundle is given by the internal hom out of the infinitesimal disk:

$$TX := [\mathbb{D}^1, X] . \tag{44.42}$$

Precomposition by the unique global element of $\mathbb{D}^1$ gives the bundle projection $TX \to X$. Analogously, the internal hom out of the superpoint gives the odd tangent bundle (see also Definition 44.1.3):

$$\Pi TX := [\mathbb{R}^{0|1}, X] . \tag{44.43}$$

Definition 44.4.42 (Local diffeomorphism[10]). Recall Definition 44.3.17. In the setting of super smooth sets, this is equivalent to a morphism $f : X \to Y$ such that the thickened plots of X can be identified with those of Y whose reduction comes from a Euclidean plot of X. Equivalently, by taking internal homs, a morphism of super smooth sets is formally étale if and only if

$$\begin{tikzcd} {[\mathbb{D}, X]} \arrow[r] \arrow[d, "{[\mathbb{D}, f]}"'] \arrow[dr, phantom, "\text{pb}"] & X \arrow[d, "f"] \\ {[\mathbb{D}, Y]} \arrow[r] & Y \end{tikzcd}$$

is a pullback square for all infinitesimally thickened (super)points. By taking $\mathbb{D} = \mathbb{D}^1$, the Inverse Function Theorem 34.1.10 can be recovered.

This can also be related to the definition in commutative algebra (and dually, that in algebraic geometry) as follows. The shape modality is right adjoint to the reduction modality. Sending a ring (extension) to its representable presheaf and using the Yoneda lemma gives the diagram

$$\begin{tikzcd} \mathbf{CRing}(A, B) \arrow[r] \arrow[d] \arrow[dr, phantom, "\text{pb}"] & \mathbf{CRing}(A, B/I) \arrow[d] \\ \mathbf{CRing}(R, B) \arrow[r] & \mathbf{CRing}(R, B/I) \end{tikzcd}$$

This square being a pullback exactly corresponds to the lifting condition in Definition 2.4.48.

Property 44.4.43. Local diffeomorphisms are preserved by the bosonic modality. If f is formally étale, so is $\overset{\rightsquigarrow}{f}$.

Alternative Definition 44.4.44 (Smooth manifold). A diffeological space (in its incarnation as a formal smooth set) equipped with a family of local diffeomorphisms from Euclidean spaces (also regarded as formal smooth sets) such that every point of the space lies in the image of at least one such morphism (and such that the final topology induced by the plots of the smooth set is paracompact Hausdorff).

Definition 44.4.45 (V-manifold). Consider a group object V in **SuperSmoothSet**. A V-manifold is a super smooth set X equipped with a span

$$\begin{tikzcd} & U \arrow[dl, "\text{ét}"'] \arrow[dr, "\text{ét, epi}"] & \\ V & & X \end{tikzcd}$$

Definition 44.4.46 (Supermanifold). A supermanifold of **(super)dimension** $(m \mid n)$ is a super smooth set X equipped with a formally étale epimorphism $\sqcup_{i \in I} \mathbb{R}^{m|n} \to X$ for some index set I.

Using the notion of supermanifold, one can also define super fibre bundles (and other common notions from differential geometry). Given such a bundle $E \to M$, its (super)space of sections is the fibre product

$$\Gamma(E) := [M, E] \times_{[M,M]} \mathcal{Y}\{\mathbb{1}_M\}\,.$$

Property 44.4.47. By Property 44.4.43 above, the bosonic space underlying a V-manifold is a $\overset{\infty}{V}$-manifold.

44.5 Higher Lie theory

In this section, some notions about groups, Lie groups and groupoids (Sections 2.2, 32.1 and 3.5.1) are extended to the setting of higher category theory.

44.6 Gauge theory ♣

Recall the notions of Chapter 12, in particular the notions of stacks and higher topoi. The $(\infty, 1)$-category of smooth ∞-stacks can be described in terms of the (left Bousfield) localization of a suitable presheaf category by Lurie's theorem 14.2.1.

The first possibility is the category of ∞-presheaves on **Diff** with the localization at open covers. The second possibility is the dense subsite $\mathbf{CartSp}_{\mathrm{diff}}$ with localization at good open covers. Both will result in a Čech model structure (Property 14.4.3). However, the specific properties will differ.

Example 44.6.1 (Classifying stacks). Consider the example of a Lie group G and its classifying stack $\mathbf{B}G$. In the first model structure, the mapping space $\mathbf{H}(M, \mathbf{B}G)$, for M a smooth manifold, is simply presented[11] by $\mathrm{Hom}(M, \mathbf{B}G)$, since M is cofibrant as a representable presheaf and $\mathbf{B}G$ is fibrant by gluing over covers. So, mapping spaces $\mathbf{H}(M, \mathbf{B}G)$ are just given by groupoids of G-bundles over M.

On the subsite $\mathbf{CartSp}_{\mathrm{diff}}$, the presheaves represented by manifolds are not cofibrant anymore. However, Čech nerves of open covers give a cofibrant replacement. On the other hand, over Cartesian spaces, the stacks are trivial and can be presented as action groupoids $* /\!\!/ G$ (the ordinary deloopings). A fibrant replacement is given by the

[11] This means the homotopy-invariant hom-object in the underlying presheaf category, where the domain is replaced by a cofibrant object and the codomain by a fibrant object.

presheaf

$$U \mapsto N_\Delta\big(*/\!\!/C^\infty(U,G)\big)\,. \tag{44.44}$$

This presheaf is also equivalent to the groupoid of G-bundles (over U). The derived mapping space in this situation is given by (normalized) G-valued Čech cocycles.

44.6.1 Lie groupoids

Definition 44.6.2 (Lie groupoid[12]). A groupoid internal to **Diff**.

Remark 44.6.3. Note that Definition 3.5.2 requires the existence of pullbacks. In the category **Diff**, this is equivalent to assuming that the source and target morphisms are (surjective) submersions.

In the Ehresmannian approach, the manifold of composable morphisms $D_1 \times_{D_0} D_1$ is given as part of the data. Hence, no further assumptions have to be made about the source and target morphisms.

Definition 44.6.4 (Lie algebroid). A vector bundle $\pi : E \to M$ together with a morphism $\rho : E \to TM$, called the **anchor map**, and a Lie bracket on $\Gamma(E)$ such that the following Leibniz-type property is satisfied:

$$[X, fY] = f[X,Y] + \rho(X)(f)Y\,. \tag{44.45}$$

This property also implies that ρ preserves the Lie bracket:

$$\rho([X,Y]) = [\rho(X), \rho(Y)]\,. \tag{44.46}$$

In local coordinates x^i and for a local basis of sections s_α, the bracket and anchor can be expressed in terms of structure functions:

$$\begin{aligned} \rho(s_\alpha) &= R^i_\alpha \partial_i\,, \\ [s_\alpha, s_\beta] &= C^\gamma_{\alpha\beta}(x) s_\gamma\,. \end{aligned} \tag{44.47}$$

The Lie algebroid properties then imply the following conditions on these structure functions:

$$R^j_\alpha \frac{\partial R^i_\beta}{\partial x^j} - R^j_\beta \frac{\partial R^i_\alpha}{\partial x^j} = R^i_\gamma C^\gamma_{\alpha\beta} \tag{44.48}$$

and

$$R^i_\alpha \frac{\partial C^\kappa_{\beta\gamma}}{\partial x^i} + C^\kappa_{\alpha\mu} C^\mu_{\beta\gamma} + (\alpha \leftrightarrow \beta \leftrightarrow \gamma) = 0\,. \tag{44.49}$$

[12] In a similar way one could define *topological groupoids*, *étalé groupoids*, ...

Example 44.6.5 (Tangent Lie algebroid). The tangent bundle over a smooth manifold is a Lie algebroid with $\rho = \mathbb{1}_{TM}$.

Consider the **pair groupoid** or **codiscrete groupoid** $\mathbf{M} \times \mathbf{M}$:

- **Objects**: M, and
- **Morphisms**: $M \times M$, i.e. between every two points there exists a unique morphism.

Both the fundamental groupoid $\mathbf{\Pi}_1(M)$ of Definition 7.1.17 and the pair groupoid $\mathbf{M}\times\mathbf{M}$ *integrate* the tangent Lie algebroid.

@@ ADD LIE INTEGRATION OF ALGEBROIDS @@

One can generalize the dual construction of L_∞-algebras even further.

Definition 44.6.6 (L_∞-algebroid). Consider Definition 4.9.15 of the Chevalley–Eilenberg algebra for an L_∞-algebra. By replacing the base field by a smooth algebra $C^\infty(M)$ for some smooth manifold M and the (graded) vector space V by a module of sections $\Gamma(E)$ of a (graded) vector bundle $E \to M$, one obtains the notion of a L_∞-algebroid.

Property 44.6.7. L_∞-algebras can be recovered by considering the special case $M = \{*\}$.

Example 44.6.8 (de Rham complex). Consider the tangent algebroid of a smooth manifold M. The associated Chevalley–Eilenberg complex is equivalent to the de Rham complex $\Omega^\bullet(M)$.

44.7 Weak groups

Definition 44.7.1 (Weak 2-group). Let $(\mathbf{C}, \otimes, \mathbf{1})$ be a monoidal category. This category is called a weak 2-group or **categorical group** if it satisfies the following conditions:

1. All morphisms are invertible.
2. Every object is weakly invertible with respect to the monoidal structure.

By Property 14.1.7, one can equivalently define a weak 2-group as a 2-category with a single object, weakly invertible 1-morphisms and invertible 2-morphisms.

The definition of a weak 2-group can be strengthened to that of a **coherent 2-group** (sometimes called a **gr-category**), where the isomorphisms $x \otimes x^{-1} \to \mathbf{1}$ and $x^{-1} \otimes x \to \mathbf{1}$ glue together to form an adjoint equivalence.

Example 44.7.2 (Automorphism 2-group). Consider a 2-category $\mathbf{C}$. For every object $c \in \mathrm{ob}(\mathbf{C})$, the automorphism category $\mathbf{Aut}(c)$ of autoequivalences of c and invertible

2-morphisms between them forms a coherent 2-group.[13]

Definition 44.7.3 (2-groupoid). A 2-groupoid is a 2-category in which all 1-morphisms are invertible and every 2-morphisms has a 'vertical' inverse. (The 'horizontal' inverse can be constructed from the other ones.)

Definition 44.7.4 (Strict 2-group). A (strict) 2-group is defined as a (strict) 2-groupoid with only one object. From this it follows that the set of 1-morphisms forms a group and so does the set of 2-morphisms under horizontal composition. However, the 2-morphisms do not form a group under vertical composition because the sources/targets may not match.

This definition is equivalent to the following internal version. A (strict) 2-group is a group object in **Cat** or an internal category in **Grp**. If **Grp** is replaced by **Lie**, the notion of a (strict) Lie 2-group is obtained.

Definition 44.7.5 (∞-groupoid). A ∞-category in which all morphisms are invertible. This is equivalent to a $(\infty, 0)$-category in the language of (n, r)-categories.

Property 44.7.6 (Lie crossed modules). The 2-category of (strict) 2-groups is biequivalent to the 2-category of (Lie) crossed modules (Definition 2.3.18). Given a 2-group $\mathcal{G}$, a crossed module is obtained as follows:

- $G := \mathsf{ob}(\mathcal{G})$,
- $H := \{h \in \mathsf{hom}(\mathcal{G}) \mid \mathfrak{s}(f) = e\}$,
- $t(h) := \mathfrak{t}(h)$, and
- $\alpha(g)h := \mathbb{1}_g h \mathbb{1}_g^{-1}$,

where $\mathfrak{s}, \mathfrak{t}$ are the source and target morphisms in $\mathcal{G}$.

To every Lie crossed module one can also assign a **differential crossed module**. This consists of the following data:

- two Lie algebras $\mathfrak{g}, \mathfrak{h}$,
- a Lie algebra morphism $\partial : \mathfrak{h} \to \mathfrak{g}$, and
- a Lie algebra morphism $\rho : \mathfrak{g} \to \mathrm{Der}(\mathfrak{h})$.

The equivariance and Peiffer conditions induce similar conditions for the above data:

- $\partial(\rho(h)g) = [h, \partial g]$, and
- $\rho(\partial h)(h') = [h, h']$,

[13] Relaxing the objects to weakly invertible 1-morphisms gives a weak 2-group.

where $g \in \mathfrak{g}$ and $h, h' \in \mathfrak{h}$. The biequivalence of crossed modules and strict 2-groups induces a biequivalence of differential crossed modules and strict Lie 2-algebras.

Example 44.7.7 (Automorphism 2-group). Given a Lie group H, one can construct a crossed module with $G := \mathrm{Aut}(H)$, t assigning inner automorphisms (conjugations) and α the obvious map. The associated 2-group $\mathbf{Aut}(H)$ is the automorphism 2-group of H of the delooping $\mathbf{B}H$.

Definition 44.7.8 (Exponentiable group). A smooth group for which every smooth function $f : [0,1] \to \mathfrak{g}$ corresponds to a smooth function $g : [0,1] \to G$ such that

$$\frac{\mathrm{d}}{\mathrm{d}t} g(t) = f(t)g(t) \tag{44.50}$$

with $g(0) = e$. A smooth 2-group is said to be exponentiable if both of its component groups are exponentiable. Since all Lie groups are exponentiable, all Lie 2-groups are also exponentiable

Remark 44.7.9 (Lie's third theorem). In ordinary Lie theory, Lie's third theorem states that every (finite-dimensional) Lie algebra can be obtained as the infinitesimal version of a Lie group. However, this does not carry over to the 2-group setting. Consider, for example, the Lie 2-algebras $\mathfrak{g}_\lambda$ constructed in Example 4.9.13. As shown by Baez and Lauda (2003), only $\mathfrak{g}_0$ gives rise to a Lie 2-group (or even a topological 2-group). This remark also extends to the setting of Lie algebroids.

44.7.1 Spaces

To overcome the problem encountered in Definition 44.6.2 (see the subsequent remark), one should pass from $\mathbf{Diff}$ to $\mathbf{C}^\infty$. It can be shown that this category admits all pullbacks, quotients, path spaces, etc.

Definition 44.7.10 (Smooth 2-space). A category internal to $\mathbf{C}^\infty$.

In the remainder of this, chapter all spaces will be assumed to be smooth in this generalized sense. The notions of 2-groups as introduced in the previous section are easily generalized to this setting.

Definition 44.7.11 (2-group action). Consider a smooth 2-group G and a smooth 2-space E. A strict action of G on E is a smooth morphism $G \to \mathbf{Aut}(E)$, i.e. a smooth (ana)functor preserving products and inverses.

Definition 44.7.12 (Thin homotopy). Let M be a smooth manifold. A smooth homotopy $H : [0,1]^2 \to M$ is said to be thin if

$$H(s,t) = F(s) \tag{44.51}$$

for some smooth F near $t = 0, 1$ and if it pulls back every two-form to 0:

$$\forall \omega \in \Omega^2(M) : H^*\omega = 0\,. \tag{44.52}$$

Definition 44.7.13 (Lazy path). Let M be a smooth manifold. A path $f : [0,1] \to M$ is said to be lazy or **to have sitting instants** if it is locally constant on some neighbourhoods of 0 and 1.

Definition 44.7.14 (Path groupoid). Let M be a smooth space. The path groupoid $\mathcal{P}_1(M)$ is the smooth groupoid conssting of the following data:

- **Objects**: M, and
- **Morphisms**: thin homotopy classes of lazy paths with fixed endpoints on M.

The laziness combined with the first condition of thin homotopies implies that the morphisms of this groupoid are (locally) constant near the boundary of their domain.

In fact, by suitably generalizing the smoothness properties of the homotopies and paths, one can extend this definition to surfaces, volumes and so on. This results in the n-path n-groupoid $\mathcal{P}_n(M)$.

Remark 44.7.15. The restriction to lazy paths is required to ensure the smoothness of composite paths. The quotient by thin homotopies is required to ensure the validity of the associativity and invertibility properties.

The following definition generalizes Definition 10.3.10.

Definition 44.7.16 (Algebraic stack). A stack $X \in \mathbf{Sh}_{(2,1)}(\mathbf{Sch}_{\text{fppf}})$ on the big fppf-site such that

1. The diagonal $\Delta_X : X \to X \times X$ is representable by an algebraic space (Definition 10.3.10).
2. There exists a scheme $S \in \mathbf{Sch}$ and a morphism $h_S \to X$ that is surjective and smooth.

If the covering morphism is étale and not just smooth, the notion of **Deligne–Mumford stacks** is obtained. To distinguish these cases, algebraic stacks are sometimes called **Artin stack**.

@@ COMPLETE @@

44.8 2-Bundles

A first step is the generalization of the categorical definition of a general bundle (Definition 33.1.1), i.e. as an object of a slice category.

Definition 44.8.1 (Smooth 2-bundle). A triple (E, B, π) where both E and B are smooth 2-spaces and π is a smooth map.

Definition 44.8.2 (Locally trivial 2-bundle). A locally trivial 2-bundle with typical fibre F over a smooth 2-space B is defined as a 2-bundle (E, B, π) with an open cover $\{U_i\}_{i\in I}$ of B such that for every $i \in I$ there exists an equivalence $\varphi_i : E|_{U_i} \cong U_i \times F$ that makes the diagram below commute:

$$\begin{tikzcd} E|_{U_i} \arrow[rr, "\varphi_i"] \arrow[dr, "\pi"'] & & U_i \times F \arrow[dl, "\mathrm{pr}_1"] \\ & U_i & \end{tikzcd}$$

It should be noted that the existence of such a cover is not a trivial matter. The general definition becomes quite involved when allowing for arbitrary smooth 2-spaces B. For convenience, it will always be assumed that B is an ordinary smooth space regarded as a 2-space with only trivial morphisms.

As was the case in Definition 33.1.5, one can also characterize locally trivial 2-bundles by their transition data. Since the trivilizations φ_i are equivalences, they admit an inverse (up to an invertible 2-map) and one can thus construct transition maps $\varphi_i \circ \varphi_j^{-1} = U_{ij} \times F \cong U_{ij} \times F$ as usual. By the commutative diagram above, these transition maps only act on the fibre F. Because $\varphi_i \circ \varphi_j^{-1}$ is itself an (auto)equivalence, the action on F is given by a functor $g_{ij} : U_{ij} \to \mathbf{Aut}(F)$, where the 2-space $\mathbf{Aut}(F)$ is the (coherent) automorphism 2-group of F.

The interesting (and important) part is how the cocycle conditions (Eq. (33.1) and Property 33.1.4) for the maps g_{ij} are modified. Since the equivalences g_{ij} are only invertible up to 2-morphisms, one cannot expect these conditions to hold as equations. Instead, two higher transition maps (i.e. natural isomorphisms) $h_{ijk} : g_{ij} \circ g_{jk} \Rightarrow g_{ik}$ and $k_i : g_{ii} \Rightarrow \mathrm{id}_F$ are obtained. These higher data should in turn satisfy the necessary conditions coming from associativity and unitality constraints (similar to the coherence conditions from Section 4.10.1).

Definition 44.8.3 (G-bundle). A locally trivial 2-bundle with typical fibre F is said to have the 2-group G as its structure (2-)group if the transition data factor through an action $G \to \mathbf{Aut}(F)$. If $F = G$, the 2-bundle is called a **principal G-2-bundle**.

Remark 44.8.4 (Gerbes). If the transition maps k_i are chosen to be trivial and G is chosen to be, respectively, the trivial Lie 2-group associated to an Abelian Lie group G or the automorphism 2-group of a Lie group H, one obtains Abelian and non-Abelian *gerbes*. In fact, it can be shown that the 2-category of principal 2-bundles is equivalent to the 2-category of gerbes for every Lie 2-group of the aforementioned type.

By categorifying Definition 35.3.37 of principal connections, one can define connections for principal n-bundles.

Definition 44.8.5 (n-connection). Let M be a smooth space and let G be a Lie n-groupoid. Given a locally trivial principal n-bundle P over M, an n-connection with n-holonomy is defined by the following data:

- for every coordinate chart $U_i \subseteq M$ a local holonomy n-functor

$$\mathsf{hol}_i : \mathcal{P}_n(U_i) \to G\,, \tag{44.53}$$

- for every double intersection U_{ij} a 1-transfor (Definition 14.1.12)

$$g_{ij} : \mathsf{hol}_i \Rightarrow \mathsf{hol}_j\,, \tag{44.54}$$

- for every triple intersection U_{ijk} a 2-transfor

$$f_{ijk} : g_{ij} \circ g_{jk} \Rrightarrow g_{ik}\,, \tag{44.55}$$

- and so on ...

This is equivalently given by a global n-functor

$$\mathsf{hol} : \mathcal{P}_n(M) \to \mathbf{Trans}_n(P)\,. \tag{44.56}$$

@@ ADD GERBES (e.g. BRYLINSKI) @@

44.9 Space and quantity

In this section, the general notions of spaces and observables are reconsidered. From the start, everything will be formulated in an enriched setting, where $\mathcal{V}$ is a cosmos (Definition 4.2.1). The categories of interest will also be assumed to be small.

In Section 44.4, spaces modelled on a base space or, more generally, on a category of spaces were presented as (concrete) sheaves on a suitable site. Here, this notion is relaxed as much as possible.

Definition 44.9.1 (Space). A (generalized) space modelled on a category $\mathbf{C}$ is a presheaf on $\mathbf{C}$.

As before, the object $X(C)$ can be interpreted as the collection of 'probes' from C to X. The Yoneda lemma assures that ordinary test spaces in $\mathbf{C}$ can be viewed as spaces modelled on $\mathbf{C}$ and that their probes are indeed the ordinary maps in $\mathbf{C}$.

In a similar vein, one can define observables as maps out of a space.

Definition 44.9.2 (Quantity). A (generalized[14]) quantity on a category $\mathbf{C}$ is a copresheaf on $\mathbf{C}$.

Property 44.9.3 (Isbell duality). Given a space X, one can look at the quantities that live on it (in ordinary geometry this would have been its algebra of functions). This defines a functor:

$$\mathcal{O} : \mathbf{Psh}(\mathbf{C}) \to \mathbf{coPsh}^{\mathrm{op}}(\mathbf{C}) : X \mapsto \mathsf{Hom}_{\mathbf{Psh}(\mathbf{C})}(X, \mathcal{Y}-) . \tag{44.57}$$

Similarly, given a quantity Q one can ask on which space it behaves as the algebra of functions. This also defines as functor:

$$\mathsf{Spec} : \mathbf{coPsh}^{\mathrm{op}}(\mathbf{C}) \to \mathbf{Psh}(\mathbf{C}) : Q \mapsto \mathsf{Hom}_{\mathbf{coPsh}(\mathbf{C})}(\mathcal{Y}^{\mathrm{op}}-, Q) , \tag{44.58}$$

where $\mathcal{Y}^{\mathrm{op}}$ denotes the co-Yoneda embedding $\mathbf{C} \to [\mathbf{C}, \mathcal{V}]^{\mathrm{op}} : c \mapsto \mathbf{C}(c, -)$.

The incredible result is now that $(\mathcal{O} \dashv \mathsf{Spec})$ is an adjunction, called the **Isbell adjunction**. Objects that are preserved (up to isomorphism) under the associated (co)monad are said to be **Isbell selfdual**.

Example 44.9.4 (Cartesian spaces). When working over the site $\mathbf{CartSp}$ (with its usual topology) and restricting to coherent sheaves and product-preserving presheaves, the Isbell adjunction maps spaces to smooth algebras.

[14] It is generalized because it 'measures' a category instead of a single object.

Part VII

Probability Theory & Statistics

Table of Contents

45 Probability Theory **838**

45.1 Probability . . . 839
45.2 Conditional probability . . . 840
45.3 Probability distributions . . . 841
45.3.1 Distribution functions . . . 841
45.3.2 Moments . . . 844
45.3.3 Correlation . . . 846
45.3.4 Conditional expectation . . . 847
45.4 Joint distributions . . . 848
45.4.1 Common distributions . . . 850
45.5 Information theory . . . 853
45.6 Extreme value theory . . . 854
45.7 Copulas . . . 855
45.8 Randomness ♣ . . . 856
45.9 Optimal transport . . . 857
45.9.1 Kantorovich duality . . . 857
45.9.2 Convex costs . . . 859
45.9.3 Concave costs . . . 860
45.9.4 Densities . . . 860
45.10 Categorical probability ♣ . . . 862
45.10.1 Probability monads . . . 862
45.10.2 Markov categories . . . 863

46 Stochastic Calculus **866**

46.1 Stochastic processes . . . 867
46.2 Brownian motion . . . 869
46.2.1 Classical Wiener space . . . 869
46.2.2 Wiener process . . . 870
46.2.3 Abstract Wiener space . . . 871
46.3 Martingales . . . 875
46.3.1 Proper martingales . . . 875

46.3.2 Local martingales . . . 879
46.4 Itô calculus . . . 880
46.4.1 Wiener–Paley integral . . . 880
46.4.2 Stochastic integral . . . 882
46.4.3 Properties . . . 885
46.4.4 Lebesgue–Stieltjes integration . . . 889
46.4.5 Representation theorems . . . 890
46.5 Stochastic differential equations . . . 892
46.6 Markov processes . . . 893

47 Information Geometry ♣ 894
47.1 Statistical manifolds . . . 894
47.1.1 Divergences . . . 896
47.1.2 Exponential families . . . 899
47.1.3 Mixtures . . . 900
47.1.4 Compatible divergences . . . 901
47.1.5 Flat structures . . . 903
47.2 Projections . . . 903
47.2.1 Functional descent . . . 904

48 Statistics 906
48.1 Data samples . . . 907
48.1.1 Moment estimators . . . 907
48.1.2 Dispersion . . . 909
48.2 Empirical distribution . . . 910
48.3 Errors . . . 913
48.4 Parameter estimation . . . 914
48.4.1 General properties . . . 914
48.4.2 Common estimators . . . 915
48.4.3 Estimation error . . . 915
48.4.4 Likelihood function . . . 916
48.4.5 Maximum likelihood estimation . . . 916
48.4.6 Least squares estimation . . . 917
48.4.7 Geometric approach . . . 919
48.5 Bayesian modelling . . . 919
48.6 Confidence intervals . . . 920
48.6.1 Interval types . . . 920
48.6.2 General construction . . . 921
48.6.3 Interval for a sample mean . . . 921
48.7 Hypothesis testing . . . 922
48.7.1 Testing . . . 922

48.7.2 Comparison tests . . . 923
48.7.3 Post-hoc tests . . . 925
48.8 Goodness of fit . . . 926
48.8.1 χ^2-test . . . 926
48.8.2 Runs test . . . 927

49 Causal Inference **928**
49.1 Causal DAGs . . . 928
49.2 Do-calculus . . . 929

50 Data Analysis **930**
50.1 Data sampling . . . 931
50.1.1 Inverse CDF sampling . . . 931
50.1.2 Uniform rejection sampling . . . 932
50.1.3 Monte Carlo sampling . . . 932
50.2 Optimization Theory . . . 933
50.2.1 Linear problems . . . 933
50.2.2 Gradient descent . . . 935
50.2.3 Conjugate gradient . . . 936
50.2.4 Nonlinear conjugate gradients . . . 938
50.2.5 Krylov methods . . . 940
50.3 Constrained optimization . . . 940
50.3.1 Lagrange multipliers . . . 940
50.3.2 Riemannian gradient descent . . . 942
50.4 Approximation theory . . . 943
50.4.1 Bayes optimality . . . 943
50.4.2 PAC theory and empirical risk minimization . . . 943
50.5 Classification problems . . . 945
50.5.1 Clustering . . . 945
50.5.2 Nearest-neighbour search . . . 946
50.6 Support-vector machines . . . 946
50.6.1 Kernel methods . . . 946
50.6.2 Decision boundaries . . . 948
50.7 Complexity theory . . . 950
50.7.1 VC dimension . . . 950
50.7.2 Rademacher complexity . . . 952
50.7.3 Relation to Glivenko–Cantelli classes . . . 953
50.8 Time series analysis . . . 953
50.8.1 Stationarity . . . 953
50.8.2 Correlation . . . 954
50.8.3 Autoregressive models . . . 955

50.8.4 Granger causality 957
50.9 Uncertainty quantification 958
50.9.1 Prediction regions 958
50.9.2 Classifier calibration 959
50.9.3 Conditionality 960
50.9.4 Distribution shift 961
50.10 Generative models 963
50.10.1 Normalizing flows 963

51 Fuzzy Sets & Imprecise Probabilities ♣ 965
51.1 Fuzzy sets 965
51.2 Imprecise probabilities 966
51.2.1 Gambles 966
51.2.2 Credal sets 967
51.2.3 Fuzzy measure theory 967

Chapter 45

Probability Theory

The majority of this chapter uses the language of measure theory. See Chapter 18 for an introduction. The section on optimal transport is partially based on Beiglboeck, Leonard, and Schachermayer (2009); Nenna (2020).

Chapter Contents

45.1 Probability **839**
45.2 Conditional probability **840**
45.3 Probability distributions **841**
45.3.1 Distribution functions 841
45.3.2 Moments 844
45.3.3 Correlation 846
45.3.4 Conditional expectation 847
45.4 Joint distributions **848**
45.4.1 Common distributions 850
45.5 Information theory **853**
45.6 Extreme value theory **854**
45.7 Copulas **855**
45.8 Randomness ♣ **856**
45.9 Optimal transport **857**
45.9.1 Kantorovich duality 857
45.9.2 Convex costs 859
45.9.3 Concave costs 860
45.9.4 Densities 860
45.10 Categorical probability ♣ **862**

45.10.1 Probability monads 862
45.10.2 Markov categories 863

45.1 Probability

The Kolmogorov axioms of probability state when a set admits the definition of a probability theory.

Axiom 45.1 (Kolmogorov). A **probability space** (Ω, Σ, P) is a measure space (Definition 18.1.3) with normalized measure $P(X) = 1$. The set Ω is called the **sample space**.

Definition 45.1.1 (Random variable). A measurable function $X : \Omega \to \mathbb{R}$ on a probability space (Ω, Σ, P), i.e.

$$\forall a \in \mathbb{R} : X^{-1}\big([a, +\infty[\big) = \{\omega \in \Omega \mid X(\omega) \geq a\} \in \Sigma\,. \tag{45.1}$$

Definition 45.1.2 (σ-algebra of a random variable). Let X be a random variable defined on (Ω, Σ, P) and denote the Borel σ-algebra of $\mathbb{R}$ by $\mathcal{B}$. The following collection of sets is a σ-algebra:

$$X^{-1}(\mathcal{B}) := \{S \in \Sigma \mid \exists B \in \mathcal{B} : S = X^{-1}(B)\}\,. \tag{45.2}$$

Notation 45.1.3. The σ-algebra generated by the random variable X is often denoted by $\mathcal{F}_X$ instead of Notation 1.6.16.

Definition 45.1.4 (Event). Let (Ω, Σ, P) be a probability space. An element S of the σ-algebra Σ is called an event.

From this definition, it is clear that a single possible outcome of a measurement can be part of multiple events. So, although only one outcome can occur at the same time, multiple events can occur simultaneously.

Remark. The Kolmogorov axioms use the σ-algebra (Definition 1.6.12) of events instead of the power set (Definition 1.6.1) of all events. Intuitively this seems to mean that some possible outcomes are not treated as events. However, one can make sure that the σ-algebra still contains all 'useful' events by using a 'nice' definition of probability spaces.

Definition 45.1.5 (Disjoint events). Two events A and B are said to be disjoint if they cannot happen at the same time:

$$P(A \cap B) = 0\,. \tag{45.3}$$

The finite additivity of measures implies the following rules.

Formula 45.1.6 (Union). Let A, B be two events. The probability that at least one of them occurs is given by the following formula:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)\,. \tag{45.4}$$

Formula 45.1.7 (Complement). Let A be an event. The probability of A being false is denoted as $P(\overline{A})$ and is given by

$$P(\overline{A}) = 1 - P(A)\,. \tag{45.5}$$

Corollary 45.1.8. From the previous formula and de Morgan's laws (1.12) and (1.13), one can derive the following formula:

$$P(\overline{A} \cap \overline{B}) = 1 - P(A \cup B)\,. \tag{45.6}$$

45.2 Conditional probability

Definition 45.2.1 (Conditional probability). Let A, B be two events. The probability of A given that B is true, denoted by $P(A \mid B)$, is defined as follows:

$$P(A \mid B) := \frac{P(A \cap B)}{P(B)}\,. \tag{45.7}$$

By interchanging A and B in the previous equation and by observing that this has no effect on the quantity $P(A \cap B)$, the following important result can be derived.

Theorem 45.2.2 (Bayes). *Let A, B be two events.*

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \tag{45.8}$$

Formula 45.2.3. Let $(B_n)_{n\in\mathbb{N}}$ be a pairwise disjoint cover of Ω. The total probability of an event A can be calculated as follows:

$$P(A) = \sum_{n=0}^{+\infty} P(A \mid B_n)P(B_n)\,. \tag{45.9}$$

Definition 45.2.4 (Independent events). Let A, B be two events. A and B are said to be independent if they satisfy the following relation:

$$P(A \cap B) = P(A)P(B)\,. \tag{45.10}$$

Corollary 45.2.5. If A and B are two independent events, Bayes' theorem simplifies to

$$P(A \mid B) = P(A)\,. \tag{45.11}$$

The above definition can be generalized to multiple events.

Definition 45.2.6 (Independent events). The events $\{A_1, \dots, A_n\}$ are said to be independent if, for each choice of $k \leq n$ events, the probability of their intersection is equal to the product of their individual probabilities.

This definition can be stated in terms of σ-algebras.

Definition 45.2.7 (Independence). The σ-algebras $\{\mathcal{F}_1, \dots, \mathcal{F}_n\}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ are said to be independent if, for all choices of distinct indices $i_1, \dots, i_k$ and for all choices of sets $F_{i_n} \in \mathcal{F}_{i_n}$, the following equation holds:

$$P(F_{i_1} \cap \cdots \cap F_{i_k}) = P(F_{i_1}) \cdots P(F_{i_k}) . \tag{45.12}$$

Corollary 45.2.8. Two random variables are independent if the σ-algebras generated by them are independent.

45.3 Probability distributions

45.3.1 Distribution functions

Definition 45.3.1 (Probability distribution). Let X be a random variable on (Ω, Σ, P). The following function is a probability measure on the Borel σ-algebra of $\mathbb{R}$:

$$P_X(B) = P\big(X^{-1}(B)\big) . \tag{45.13}$$

This measure is called the probability distribution of X.

Example 45.3.2 (Rademacher variable). A random variable on $\Omega = \{-1, 1\}$ with probability distribution $P_X(-1) = P_X(1) = \frac{1}{2}$.

Definition 45.3.3 (Density). Consider an integrable function $f \geq 0$ and recall Property 18.2.18. The function f is called the density of the measure $P(A) := \int_A f \, d\lambda$ (with respect to the Lebesgue measure λ). If the measure is a probability measure, i.e. it is normalized to 1, f is called a **probability density function**.

More generally, by the Radon–Nikodym Theorem 18.5.6, every absolutely continuous probability distribution P on $\mathbb{R}^n$ is of the form

$$P(A) = \int_A f \, d\lambda \tag{45.14}$$

for some integrable function f.

When P is discrete, i.e. when one works with respect to the counting measure, the Radon–Nikodym derivative is called the **probability mass function**. (In this compendium, this function will also often be called the density function.)

Definition 45.3.4 (Cumulative distribution function). Consider a random variable X and its associated distribution P_X. The cumulative distribution function $F_X : \mathbb{R} \to [0,1]$ is defined as follows:

$$F_X(a) := P_X\big(\{x \in \mathbb{R} \mid x \leq a\}\big)\,. \tag{45.15}$$

The following is a consequence of the Kolmogorov axioms.

Corollary 45.3.5. Let $F : \mathbb{R} \to [0,1]$ be a function that satisfies the following three properties:

- F is increasing.
- $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.
- F is right-continuous, i.e. $\lim_{y\nearrow y_0} F(y) = F(y_0)$ for all $y_0 \in \mathbb{R}$.

There exists a random variable $X : [0,1] \to \mathbb{R}$ on the probability space $([0,1], \mathcal{B}_{[0,1]}, \lambda_{[0,1]})$, such that $F = F_X$, where $\mathcal{B}_{[0,1]}$ is the Borel σ-algebra of $[0,1]$ with its Euclidean topology.

Definition 45.3.6 (Parametric family). An indexed family of probability distributions.

Example 45.3.7 (Mixture family). Consider a collection of distributions $\mathcal{P} \equiv \{P_i\}_{i\leq n}$. The mixture family generated by $\mathcal{P}$ consist of all convex combinations of elements in $\mathcal{P}$:

$$\left\{ \sum_{i=1}^{n} w_i P_i \,\middle|\, w_i \geq 0, \sum_{i=1}^{n} w_i = 1 \right\}. \tag{45.16}$$

Every element of this family is called a **mixture distribution**.

The following theorem is a specific instance of the more general change-of-variables formula.

Theorem 45.3.8 (Theorem of the unconscious statistician). *Consider a random variable X on a probability space (Ω, Σ, P). The following equality holds for every integrable function $g \in L^1(\mathbb{R})$:*

$$\int_\Omega g \circ X \, dP = \int_{\mathbb{R}} g \, dP_X\,. \tag{45.17}$$

Remark 45.3.9. The name of this theorem stems from the fact that many scientists take this equality to be a definition of the expectation value $\mathsf{E}\big[g(X)\big]$. However, this equality should be proven since the measure on the right-hand side is the one belonging to the random variable X and not $g(X)$.

Formula 45.3.10. Consider an absolutely continuous probability function P on $\mathbb{R}^n$ and let f be the associated density. Let $g : \mathbb{R}^n \to \mathbb{R}$ be integrable with respect to P.

$$\int_{\mathbb{R}^n} g \, dP = \int_{\mathbb{R}^n} f(x) g(x) \, dx \tag{45.18}$$

Corollary 45.3.11. The previous formula, together with Theorem 45.3.8, gives rise to

$$\int_\Omega g \circ X \, dP = \int_{\mathbb{R}^n} f_X(x) g(x) \, dx \,. \tag{45.19}$$

Formula 45.3.12. Let X be a random variable with density function f_X and let $g : \mathbb{R} \to \mathbb{R}$ be smooth and strictly monotone. The random variable $g \circ X$ has an associated density f_g given by

$$f_g(y) = f\big(g^{-1}(y)\big) \left| \frac{\mathrm{d}g^{-1}}{\mathrm{d}y}(y) \right| . \tag{45.20}$$

Weak convergence of measures (Definition 18.1.53) induces a notion of convergence of random variables.

Definition 45.3.13 (Convergence in distribution). A sequence $(X_n)_{n\in\mathbb{N}}$ of random variables is said to converge in distribution to a random variable X if the associated cumulative distribution functions F_{X_n} converge pointwise to F_X, i.e. $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ for all $x \in \mathbb{R}$.

More generally, a sequence $(X_n)_{n\in\mathbb{N}}$ of random variables converges in distribution to a random variable X if the associated probability measures P_{X_n} converge weakly to P_X (Definition 18.1.53).

Notation 45.3.14. If a sequence $(X_n)_{n\in\mathbb{N}}$ converges in distribution to a random variable X, this is often denoted by $X_n \xrightarrow{d} X$. Sometimes, the d (for 'distribution') is replaced by the $\mathcal{L}$ (for 'law').

Theorem 45.3.15 (Slutsky). *Let $(X_n)_{n\in\mathbb{N}}, (Y_n)_{n\in\mathbb{N}}$ be two sequences of random variables converging in probability to a random variable X and a constant c, respectively. The following statements hold:*

- $X_n + Y_n \xrightarrow{d} X + c$,
- $X_n Y_n \xrightarrow{d} cX$, *and*
- $X_n / Y_n \xrightarrow{d} X/c$.

Definition 45.3.16 (Convergence in probability). A sequence $(X_n)_{n\in\mathbb{N}}$ of random variables taking values in a metric space (Ω, d) is said to converge in probability to a random variable X if, for all $\varepsilon > 0$, the following statement holds:

$$\lim_{n\to\infty} \mathrm{Prob}\big(d(X_n, X) > \varepsilon\big) = 0 \,. \tag{45.21}$$

Convergence in probability implies convergence in distribution.

45.3.2 Moments

Definition 45.3.17 (Expectation value). Let X be random variable defined on (Ω, Σ, P).

$$\mathsf{E}[X] := \int_\Omega X \, dP \tag{45.22}$$

Notation 45.3.18. Other common notations are $\langle X \rangle$ and μ_X. However, the latter might be confused with a general measure on the space X and will, therefore, not be used here.

Property 45.3.19 (Markov's inequality). Let X be a random variable. For every constant $a > 0$, the following inequality holds:

$$\mathsf{Prob}(X \geq a) \leq \frac{\mathsf{E}[X]}{a} \,. \tag{45.23}$$

Definition 45.3.20 (Moment of order r). The moment of order r is defined as the expectation value of the rth power of X. By Corollary 45.3.11, this becomes

$$\mathsf{E}[X^r] = \int_\mathbb{R} x^r f_X(x) \, dx \,. \tag{45.24}$$

Definition 45.3.21 (Central moment of order r).

$$\mathsf{E}\Big[\big(X - \mathsf{E}[X]\big)^r\Big] = \int_\mathbb{R} \big(x - \mathsf{E}[X]\big)^r f_X(x) \, dx \tag{45.25}$$

Remark 45.3.22. Moments of order n are determined by central moments of order $k \leq n$ and, conversely, central moments of order n are determined by moments of order $k \leq n$.

Definition 45.3.23 (Variance). The central moment of order 2 is called the variance:

$$\mathsf{Var}[X] := \mathsf{E}\Big[\big(X - \mathsf{E}[X]\big)^2\Big] \,. \tag{45.26}$$

Definition 45.3.24 (Standard deviation).

$$\sigma_X := \sqrt{\mathsf{Var}[X]} \tag{45.27}$$

Property 45.3.25. If $\mathsf{E}[|X|^n]$ is finite for some $n > 0$, then $\mathsf{E}\big[X^k\big]$ exists and is finite for all $k \leq n$.

Property 45.3.26 (Chebyshev's inequality). Let X be a nonnegative random variable. For every constant $a > 0$, the following inequality holds:

$$\mathsf{Prob}(|X - \mathsf{E}[X]| \geq a) \leq \frac{\mathsf{Var}[X]}{a^2} \,. \tag{45.28}$$

Definition 45.3.27 (Moment generating function).

$$M_X(t) := \mathsf{E}\big[e^{tX}\big] = \int_\mathbb{R} e^{tx} f_X(x) \, dx \tag{45.29}$$

Note that this is essentially the Laplace–Stieltjes transform (Definition 19.3.26) of the probability distribution:

$$M_X(t) = \mathcal{L}\{P_X\}(-t)\,. \tag{45.30}$$

Property 45.3.28. If the moment generating function exists, the moments $\mathsf{E}[X^n]$ can be expressed in terms of M_X:

$$\mathsf{E}[X^n] = \left.\frac{\mathsf{d}^n M_X(t)}{\mathsf{d}t^n}\right|_{t=0}\,. \tag{45.31}$$

Method 45.3.29 (Chernoff bound). This bound gives a bound on the tail probabilities of a random variable. For all constants $\lambda > 0$, the Markov inequality 45.3.19 implies the following statement:

$$\mathsf{Prob}(X \geq a) = \mathsf{Prob}\left(e^{\lambda X} \geq e^{\lambda a}\right) \leq \frac{\mathsf{E}\left[e^{\lambda X}\right]}{e^{\lambda a}}\,. \tag{45.32}$$

If one has more information about the moment generating function, the Chernoff bound can be used to obtain improved concentration inequalities by optimizing over λ.

Property 45.3.30 (Hoeffding's inequalities). Consider a collection of bounded, independent random variables $\{X_1, \dots, X_n\}$. Without loss of generality, one can assume that they are constrained to the unit interval, i.e. $0 \leq X_i \leq 1$. For every constant $\lambda \in \mathbb{R}^+$, the following inequality holds:

$$\mathsf{Prob}\left(\overline{X} - \mathsf{E}[\overline{X}] \geq \lambda\right) \leq \exp\left(-2n\lambda^2\right)\,. \tag{45.33}$$

If one can sharpen the bounds for the variables such that $X_i \in [a_i, b_i]$, then

$$\mathsf{Prob}\left(\overline{X} - \mathsf{E}[\overline{X}] \geq \lambda\right) \leq \exp\left(-\frac{2n^2\lambda^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)\,. \tag{45.34}$$

Definition 45.3.31 (Characteristic function).

$$\varphi_X(t) := \mathsf{E}\left[e^{itX}\right] \tag{45.35}$$

Note that this is essentially the Fourier–Stieltjes transform (Definition 19.3.18) of P_X.

Property 45.3.32. The characteristic function has the following properties:

- $\varphi_X(0) = 1$,
- $|\varphi_X(t)| \leq 1$, and
- $\varphi_{aX+b}(t) = e^{itb}\varphi_X(at)$ for all $a, b \in \mathbb{R}$.

Formula 45.3.33. A random variable X has a finite k^{th} moment if and only if $\varphi_X \in C^k(\mathbb{R})$ with

$$\mathsf{E}\left[X^k\right] = \frac{1}{i^k}\frac{\mathsf{d}^k}{\mathsf{d}t^k}\varphi_X(0)\,. \tag{45.36}$$

Formula 45.3.34 (Inversion formula). Let X be a random variable. If the CDF of X is continuous at $a, b \in \mathbb{R}$, then

$$F_X(b) - F_X(a) = \lim_{c\to\infty} \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t)\, dt\,. \tag{45.37}$$

The following statement simply follows from the interpretation of the characteristic function as a Fourier transform.

Formula 45.3.35. If $\varphi_X(t)$ is integrable, the CDF is given by

$$f_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \varphi_X(t)\, dt\,. \tag{45.38}$$

45.3.3 Correlation

Property 45.3.36. Two random variables X, Y are independent if and only if

$$\mathsf{E}\big[f(X)g(Y)\big] = \mathsf{E}\big[f(X)\big]\mathsf{E}\big[g(Y)\big] \tag{45.39}$$

holds for all measurable, bounded functions f, g.

The expectation value $\mathsf{E}\big[XY\big]$ is equal to the inner product $\langle X \mid Y\rangle$ as defined in (18.40). It follows that independence of random variables implies orthogonality. To generalize this concept, the following notions are introduced.

Definition 45.3.37 (Centred random variable). Let X be a random variable with finite expectation value $\mathsf{E}[X]$. The centred random variable X_c is defined as $X_c = X - \mathsf{E}[X]$.

Definition 45.3.38 (Covariance). The covariance of two random variables X, Y is defined as follows:

$$\mathsf{Cov}[X, Y] := \langle X_c \mid Y_c\rangle = \mathsf{E}\big[\big(X - \mathsf{E}[X]\big)\big(Y - \mathsf{E}[Y]\big)\big]\,. \tag{45.40}$$

Some basic math gives

$$\mathsf{Cov}[X, Y] = \mathsf{E}\big[XY\big] - \mathsf{E}[X]\mathsf{E}[Y]\,. \tag{45.41}$$

Definition 45.3.39 (Correlation). The correlation of two random variables X, Y is defined as the cosine of the angle between X_c and Y_c:

$$\rho_{XY} := \frac{\mathsf{Cov}[X, Y]}{\sigma_X \sigma_Y}\,. \tag{45.42}$$

The correlation coefficient is bounded to the interval $[-1, 1]$. It should be noted that its magnitude is only an indicator for the linear dependence.

Corollary 45.3.40. From Property 45.3.36, it follows that independent random variables are uncorrelated.

Corollary 45.3.41. If the random variables X and Y are uncorrelated, they satisfy

$$\mathsf{E}\big[XY\big] = \mathsf{E}[X]\mathsf{E}[Y]\,. \tag{45.43}$$

Formula 45.3.42 (Bienaymé formula). Let $\{X_1, \dots, X_n\}$ be a collection of independent (or uncorrelated) random variables. Their variances satisfy the following equation:

$$\mathsf{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathsf{Var}\big[X_i\big]\,. \tag{45.44}$$

For general random variables, the following formula is valid:

$$\mathsf{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i,j=1}^{n} \mathsf{Cov}[X_i, X_j]\,. \tag{45.45}$$

Remark 45.3.43. For multivariate distributions, the above definitions can be generalized using matrices:

$$\mathsf{Cov}_{ij} := \mathsf{Cov}[X_{(i)}, X_{(j)}]\,, \tag{45.46}$$

$$\rho_{ij} := \rho_{X_{(i)}X_{(j)}}\,. \tag{45.47}$$

45.3.4 Conditional expectation

Let (Ω, Σ, P) be a probability space. Consider a random variable $X \in L^2(\Omega, \Sigma, P)$ and a sub-σ-algebra $\mathcal{G} \subset \Sigma$. Property 18.3.3 implies that the spaces $L^2(\Sigma)$ and $L^2(\mathcal{G})$ are complete and, hence, the Projection Theorem 25.2.20 can be applied. For every $X \in L^2(\Sigma)$, there exists a random variable $Y \in L^2(\mathcal{G})$ such that $X - Y$ is orthogonal to $L^2(\mathcal{G})$. This has the following result:

$$\forall Z \in L^2(\mathcal{G}) : \langle X - Y \mid Z\rangle \equiv \int_\Omega (X - Y)Z\,dP = 0\,. \tag{45.48}$$

Since $\mathbb{1}_B \in L^2(\mathcal{G})$ for every $B \in \mathcal{G}$, Formula 18.2.7 can be rewritten as

$$\int_B X\,dP = \int_B Y\,dP \tag{45.49}$$

for all $B \in \mathcal{G}$. This observation leads to the following definition.

Definition 45.3.44 (Conditional expectation). Let (Ω, Σ, P) be a probability space and let $\mathcal{G}$ be a sub-σ-algebra of Σ. For every Σ-measurable random variable $X \in L^2(\Sigma)$, there exists a unique (up to a null set) random variable $Y \in L^2(\mathcal{G})$ that satisfies Eq. (45.49) for every $B \in \mathcal{G}$. This variable Y is called the conditional expectation of X given $\mathcal{G}$ and it is denoted by $\mathsf{E}\big[X \mid \mathcal{G}\big]$:

$$\int_B \mathsf{E}\big[X \mid \mathcal{G}\big]\,dP = \int_B X\,dP \tag{45.50}$$

for all $B \in \mathcal{G}$.

Remark 45.3.45. Although this construction was based on orthogonal projections, one could also have used the (signed) Radon–Nikodym Theorem 18.7.4 since $B \mapsto \int_B X\,dP$ is absolutely continuous with respect to $P|_{\mathcal{G}}$.

Property 45.3.46. Let (Ω, Σ, P) be a probability space and consider a sub-σ-algebra $\mathcal{G} \subset \Sigma$. If the random variable X is $\mathcal{G}$-measurable, then

$$\mathsf{E}\big[X \,|\, \mathcal{G}\big] = X \text{ a.s.} \tag{45.51}$$

On the other hand, if X is independent of $\mathcal{G}$, then

$$\mathsf{E}\big[X \,|\, \mathcal{G}\big] = \mathsf{E}[X] \text{ a.s.} \tag{45.52}$$

45.4 Joint distributions

Definition 45.4.1 (Joint distribution). Let X, Y be two random variables defined on the same probability space (Ω, Σ, P) and consider the vector random variable $(X, Y) : \Omega \to \mathbb{R}^2$. The distribution of (X, Y) is a probability measure on the Borel algebra of $\mathbb{R}^2$ defined by

$$P_{X,Y}(B) = P\big((X, Y)^{-1}(B)\big)\,. \tag{45.53}$$

Definition 45.4.2 (Joint density). If the probability measure from the previous definition can be written as

$$P_{X,Y}(B) = \int_B f_{X,Y}(x, y)\,dx\,dy \tag{45.54}$$

for some integrable $f_{X,Y}$, X and Y are said to have a joint density.

Definition 45.4.3 (Marginal distribution). The distributions of the one-dimensional random variables is determined by the joint distribution:

$$P_X(A) = P_{X,Y}(A \times \mathbb{R})\,, \tag{45.55}$$

$$P_Y(A) = P_{X,Y}(\mathbb{R} \times A)\,. \tag{45.56}$$

Corollary 45.4.4. If the joint density exists, the marginal distributions are absolutely continuous and the associated density functions are given by

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y)\,dy\,, \tag{45.57}$$

$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y)\,dx\,. \tag{45.58}$$

The converse, however, is not always true. The one-dimensional distributions can be absolutely continuous without the existence of a joint density.

Property 45.4.5 (Independence). Let X, Y be two random variables with joint distribution $P_{X,Y}$. X and Y are independent if and only if the joint distribution coincides with the product measure:

$$P_{X,Y} = P_X \otimes P_Y \,. \tag{45.59}$$

If X and Y are absolutely continuous, the previous properties also applies to the densities.

Formula 45.4.6 (Sum of random variables). Consider two independent random variables X, Y with associated densities f_X, f_Y. The density f_Z of their sum $Z = X + Y$ is given by the following convolution (Formula 19.2.4):

$$f_Z(z) := f_X * f_Y(z) = \int_{\mathbb{R}} f_X(x) f_Y(z-x)\, dx = \int_{\mathbb{R}} f_X(z-y) f_Y(y)\, dy \,. \tag{45.60}$$

Formula 45.4.7 (Product of random variables). Consider two independent random variables X, Y with associated densities f_X, f_Y. The density f_Z of their product $Z = XY$ is given by

$$f_Z(z) = \int_{\mathbb{R}} f_X(x) f_Y(z/x) \frac{dx}{|x|} = \int_{\mathbb{R}} f_X(z/y) f_Y(y) \frac{dy}{|y|} \,. \tag{45.61}$$

Corollary 45.4.8. Taking the Mellin transform 19.3.27 of both the positive and negative part of the above integrand (to be able to handle the absolute value) gives the following relation:

$$\mathcal{M}\{f\} = \mathcal{M}\{g\}\mathcal{M}\{h\} \,. \tag{45.62}$$

Formula 45.4.9 (Conditional density). Let X, Y be two random variables with joint density $f_{X,Y}$. The conditional density of Y given $X \in A$ is

$$h(y \mid X \in A) = \frac{\int_A f_{X,Y}(x,y)\, dx}{\int_A f_X(x)\, dx} \,. \tag{45.63}$$

For $X = \{a\}$, this equation is ill defined since the denominator would become 0. However, it is possible to avoid this problem by formally setting

$$h(y \mid A = a) := \frac{f_{X,Y}(a,y)}{f_X(a)} \,, \tag{45.64}$$

where $f_X(a) \neq 0$. This last condition is nonrestrictive because the probability of having a measurement $(X, Y) \in \{(x,y) \mid f_X(x) = 0\}$ is 0 (for nonsingular measures). One can thus define the conditional probability of Y given $X = a$ as follows:

$$P(Y \in B \mid X = a) := \int_B h(y \mid X = a)\, dy \,. \tag{45.65}$$

Formula 45.4.10 (Conditional expectation).

$$\mathsf{E}\big[Y \mid X\big](\omega) = \int_{\mathbb{R}} y h\big(y \mid X(\omega)\big)\, dy \tag{45.66}$$

Let $\mathcal{F}_X$ denote the σ-algebra generated by the random variable X as before. Using Fubini's theorem, one can prove that, for all sets $A \in \mathcal{F}_X$, the following equality holds:

$$\int_A \mathsf{E}\big[Y \mid X\big]\, dP = \int_A Y\, dP\,. \tag{45.67}$$

This implies that the conditional expectation $\mathsf{E}\big[Y \mid X\big]$ on $\mathcal{F}_X$ coincides with Definition 45.3.44.

Applying Property 45.3.46 to the case $\mathcal{G} = \mathcal{F}_X$ gives the law of total expectation.

Property 45.4.11 (Law of total expectation[1]).

$$\mathsf{E}\big[\mathsf{E}\big[Y \mid X\big]\big] = \mathsf{E}\big[Y\big] \tag{45.68}$$

Theorem 45.4.12 (Bayes' theorem). *The conditional density can be computed without prior knowledge of the joint density:*

$$g(x \mid y) = \frac{h(y \mid x) f_X(x)}{f_Y(y)}\,. \tag{45.69}$$

45.4.1 Common distributions

In this section, some common distribution functions will be covered. In most cases, only a Radon–Nikodym derivative (or density function in accordance with Definition 45.3.3) with respect to the Lebesgue measure on $\mathbb{R}$ will be given.

Definition 45.4.13 (Uniform distribution).

$$f(x; a, b) := \begin{cases} \dfrac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{elsewhere} \end{cases} \tag{45.70}$$

$$\mathsf{E}[x] = \frac{a+b}{2} \tag{45.71}$$

$$\mathsf{Var}[x] = \frac{(b-a)^2}{12} \tag{45.72}$$

Definition 45.4.14 (Gaussian distribution). Let σ be a positive number.

$$\mathcal{N}(x; \mu, \sigma) := \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{45.73}$$

[1] Also called the **tower property**.

This distribution is also called a (univariate) **normal distribution** with **mean** μ and **standard deviation** σ (or **variance** σ^2).

This formula can also be generalized to multivariate distributions. Let Σ be a positive-definite matrix, called the **covariance** matrix. The associated multivariate normal distribution is given by

$$\mathcal{N}(\vec{\mathbf{x}}; \vec{\boldsymbol{\mu}}, \Sigma) = \frac{1}{\sqrt{2\pi}\det(\Sigma)} \exp\left(-\frac{(\vec{\mathbf{x}} - \vec{\boldsymbol{\mu}})^T \Sigma^{-1} (\vec{\mathbf{x}} - \vec{\boldsymbol{\mu}})}{2}\right). \tag{45.74}$$

Definition 45.4.15 (Standard normal distribution).

$$\mathcal{N}(z) := \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \tag{45.75}$$

The cumulative distribution of $\mathcal{N}$ is called the **error function** $\mathrm{erf}(x)$.

Remark 45.4.16. Every Gaussian distribution can be transformed into a standard normal distribution by passing to the random variable $Z = \frac{X-\mu}{\sigma}$. This transformation is often called **standardization**.

Theorem 45.4.17 (Central limit theorem). *A sum of $n \in \mathbb{N}$ i.i.d. random variables X_i distributed according to a distribution with mean μ and variance σ^2 satisfies the following property:*

$$\sqrt{n}\left(\sum_{i=1}^{n} X_i - \mu\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2). \tag{45.76}$$

Remark 45.4.18. If the random variables are not independent, the CLT will not hold. However, a generalization to distributions that are not identical exists. These are the *Lyapunov* and *Lindeberg* CLTs. (This generalization does require additional conditions on the higher moments.)

Formula 45.4.19. The sum of any number of (independent) Gaussian random variables is again Gaussian with the sum of the means and variances as parameters:

$$\forall i \in I : X_i \sim \mathcal{N}(\mu_i, \sigma_i^2) \implies \sum_{i \in I} X_i \sim \mathcal{N}\left(\sum_{i \in I} \mu_i, \sum_{i \in I} \sigma_i^2\right). \tag{45.77}$$

Definition 45.4.20 (Exponential distribution).

$$f(x; \tau) := \frac{1}{\tau} e^{-x/\tau} \tag{45.78}$$

$$\mathrm{E}[x] = \tau \tag{45.79}$$

$$\mathrm{Var}[x] = \tau^2 \tag{45.80}$$

Property 45.4.21. The exponential distribution is **memoryless**:

$$\mathsf{Prob}(X > x_1 + x_2 \mid X > x_2) = \mathsf{Prob}(X > x_1)\,. \tag{45.81}$$

Definition 45.4.22 (Bernoulli distribution). Any binary random variable is described by a Bernoulli distribution. When the possible values are 0 and 1 (without loss of generality), with respective probabilities ρ and $1-\rho$, the distribution is given by

$$p(k;\rho) := \rho^k(1-\rho)^{1-k}\,, \tag{45.82}$$

$$\mathsf{E}\big[k\big] = \rho\,, \tag{45.83}$$

$$\mathsf{Var}\big[k\big] = \rho(1-\rho)\,. \tag{45.84}$$

Definition 45.4.23 (Binomial distribution). A process with $n \in \mathbb{N}$ i.i.d. Bernoulli trials with probability ρ, is described by a binomial distribution:

$$\mathsf{Binom}(k;\rho,n) := \binom{n}{k}\rho^k(1-\rho)^{n-k}\,, \tag{45.85}$$

$$\mathsf{E}\big[k\big] = n\rho\,, \tag{45.86}$$

$$\mathsf{Var}\big[k\big] = n\rho(1-\rho)\,. \tag{45.87}$$

Definition 45.4.24 (Poisson distribution). A process with known possible outcomes but an unknown number of events is described by a Poisson distribution with average expected number of events $\lambda \in \mathbb{N}$:

$$\mathsf{Poisson}(r;\lambda) := \frac{e^{-\lambda}\lambda^r}{r!} \tag{45.88}$$

$$\mathsf{E}[r] = \mathsf{Var}[r] = \lambda \tag{45.89}$$

Property 45.4.25. If multiple independent Poisson processes occur simultaneously, the probability of $r \in \mathbb{N}$ events is also described by a Poisson distribution:

$$\forall i \in I : X_i \sim \mathsf{Poisson}(\lambda_i) \implies \sum_{i\in I} X_i \sim \mathsf{Poisson}\left(\sum_{i\in I}\lambda_i\right). \tag{45.90}$$

The number of events coming from the conditional process described by $\lambda_i \mid r = \sum_{i\in I}\lambda_i$ is given by a binomial distribution $\mathsf{Binom}(r_i;\Lambda_i,r)$ with $\Lambda_i = \frac{\lambda_i}{\sum_{i\in I}\lambda_i}$.

Remark 45.4.26. For $\lambda \longrightarrow +\infty$, the Poisson distribution $\mathsf{Poisson}(r;\lambda)$ can be approximated by a Gaussian distribution $\mathcal{N}(x;\lambda,\sqrt{\lambda})$.

Theorem 45.4.27 (Raikov). *If the sum of two independent random variables is Poisson, the individual random variables are also Poisson.*

Definition 45.4.28 (χ^2-distribution). The sum of $k \in \mathbb{N}$ squared independent (standard) normally distributed random variables Y_i defines the random variable

$$\chi_k^2 := \sum_{i=1}^{k} Y_i^2 \,, \tag{45.91}$$

where k is said to be the number of **degrees of freedom**. The associated density is

$$f(\chi^2; n) := \frac{\chi^{n-2} e^{-\chi^2/2}}{2^{n/2}\Gamma(n/2)} \,. \tag{45.92}$$

Property 45.4.29. Due to the CLT 45.4.17, the χ^2-distribution approximates a Gaussian distribution for large $n \in \mathbb{N}$:

$$f(\chi^2; n) \overset{n>30}{\longrightarrow} \mathcal{N}\left(\sqrt{2\chi^2}; \sqrt{2n-1}, 1\right). \tag{45.93}$$

Definition 45.4.30 (Student-t distribution). The Student-t distribution describes the difference between the true mean and a sample average with estimated standard deviation $\widehat{\sigma}$ (see Section 48.4):

$$f(t; n) := \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma(n/2)\left(1 + \frac{t^2}{n}\right)^{\frac{n+1}{2}}} \,, \tag{45.94}$$

where

$$t := \frac{(x-\mu)/\sigma}{\widehat{\sigma}/\sigma} = \frac{z}{\sqrt{\chi^2/n}} \,. \tag{45.95}$$

Definition 45.4.31 (Cauchy distribution[2]). The density is given by

$$f(x; x_0, \gamma) := \frac{1}{\pi} \frac{\gamma}{(x-x_0)^2 + \gamma^2} \,. \tag{45.96}$$

The associated characteristic function is given by

$$\mathsf{E}\left[e^{itx}\right] = e^{ix_0 t - \gamma|t|} \,. \tag{45.97}$$

Remark 45.4.32. Both the mean and variance of the Cauchy distribution are undefined.

45.5 Information theory

Definition 45.5.1 (Self-information). The self-information of an event x described by a distribution P is defined as follows:

$$I(x) := -\ln P(x) \,. \tag{45.98}$$

This definition is modeled on the following (reasonable) requirements:

[2] Also known, especially in particle physics, as the **Breit–Wigner distribution**.

- Events that are almost surely going to happen, i.e. events x such that $P(x) = 1$, contain only little information: $I(x) = 0$.[3]
- Events that are very rare contain a lot of information.
- Independent events contribute additively to the information.

Definition 45.5.2 (Shannon entropy). The amount of uncertainty in a discrete distribution P is characterized by its (Shannon) entropy

$$H(P) := \mathsf{E}[I(X)] = -\sum_i P_i \ln(P_i)\,. \tag{45.99}$$

Definition 45.5.3 (Kullback–Leibler divergence). Let P, Q be two probability distributions with $P \ll Q$. The Kullback–Leibler divergence (or **relative entropy**) of P with respect to Q is defined as follows:[4]

$$D_{\mathrm{KL}}(P \,\|\, Q) := \int_\Omega \ln\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right) dP\,. \tag{45.100}$$

This quantity can be interpreted as the information gained when using the distribution P instead of Q. Instead of a base-e logarithm, any other logarithm can be used since this simply changes the result by a (positive) scaling constant.

Property 45.5.4 (Gibbs's inequality). By noting that the logarithm is a concave function and applying Jensen's equality (Theorem 16.7.6), one can prove that the Kullback–Leibler divergence is nonnegative:

$$D_{\mathrm{KL}}(P \,\|\, Q) \geq 0\,. \tag{45.101}$$

Furthermore, the Kullback–Leibler divergence is zero if and only if $\frac{\mathrm{d}P}{\mathrm{d}Q} = 1$ holds Q-almost everywhere.

45.6 Extreme value theory

Definition 45.6.1 (Conditional excess). Consider a random variable X with distribution P. The conditional probability that X is larger than a given threshold is given by the conditional excess distribution:

$$F_u(y) = \mathsf{Prob}(X - u \leq y \mid X > u) = \frac{P(u+y) - P(u)}{1 - P(u)}\,. \tag{45.102}$$

Definition 45.6.2 (Extreme value distribution). The extreme value distribution is given by the following formula:

$$F(x;\xi) = \exp\left(-(1 + x\xi)^{-1/\xi}\right). \tag{45.103}$$

[3] And, by extension, $P(x) \approx 1 \implies I(x) \approx 0$.
[4] See Definition 47.1.7 for a generalization.

For $\xi = 0$ one can use the definition of the Euler number to rewrite the definition as

$$F(x;0) = \exp(-e^{-x}) \,. \tag{45.104}$$

The number ξ is called the **extreme value index**.

Definition 45.6.3 (Maximum domain of attraction). The (maximum) domain of attraction of a distribution function H consist of all distribution functions F for which there exist sequences $(a_n > 0)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ such that $F^n(a_n x + b_n) \longrightarrow H(x)$.

Theorem 45.6.4 (Fischer, Tippett & Gnedenko). *Consider a sequence of i.i.d. random variables with distribution F. If F lies in the domain of attraction of G, then G has the form of an extreme value distribution.*

Theorem 45.6.5 (Pickands, Balkema & de Haan). *Consider a sequence of i.i.d. random variables with conditional excess distribution F_u. If the distribution F lies in the domain of attraction of the extreme value distribution, the conditional excess distribution F_u converges to the generalised Pareto distribution when $u \longrightarrow +\infty$.*

45.7 Copulas

Property 45.7.1 (Uniformization). Consider a continuous random variable X and let U be the result of the probability integral transformation, i.e. $U := F_X(X)$. This transformed random variable has a uniform cumulative distribution, i.e. $F_U(u) = u$.

Definition 45.7.2 (Copula). A (joint) cumulative distribution function with uniform marginal distributions.

The following alternative definition is more analytic in nature.

Alternative Definition 45.7.3 (Copula). A function $C : [0,1]^d \to [0,1]$ satisfying the following properties:

1. **Normalization** $C(x_1, \dots, x_d) = 0$ if any of the x_i is zero.
2. **Uniformity:** $C(1, 1, \dots, x_i, 1, \dots) = x_i$ for all $1 \le i \le d$.
3. d**-nondecreasing:** For every box $B = \prod_{1\le i\le d}[a_i, b_i] \subseteq [0,1]^d$, the C-volume is nonnegative:

$$\int_B dC := \sum_{\mathbf{z}\in\prod_i\{a_i,b_i\}} (-1)^{N_b(\mathbf{z})} C(\mathbf{z}) \ge 0 \,, \tag{45.105}$$

   where $N_B(\mathbf{z}) = \mathsf{Card}(\{i \mid a_i = z_i\})$.

Theorem 45.7.4 (Sklar). *For every joint distribution function H with marginals F_i, there exists a unique copula C such that*

$$H(x_1, \dots, x_d) = C\big(F_1(x_1), \dots, F_d(x_d)\big) \,. \tag{45.106}$$

Property 45.7.5 (Fréchet–Hoeffding bounds). Every copula $C : [0,1]^d \to [0,1]$ is bounded in the following way:

$$\max\left(\sum_{i=1}^{d} u_i - d + 1, 0\right) \leq C(u_1, \dots, u_d) \leq \min_i u_i \tag{45.107}$$

for all $(u_1, \dots, u_d) \in [0,1]^d$. Furthermore, the upper bound is sharp, i.e. $\min_i u_i$ is itself a copula.[5]

Definition 45.7.6 (Extreme value copula). A copula C for which there exists a copula $\widetilde{C}$ such that

$$\widetilde{C}\left(u_1^{1/n}, \dots, u_d^{1/n}\right)^n \longrightarrow C(u_1, \dots, u_d) \tag{45.108}$$

for all $(u_1, \dots, u_d) \in [0,1]^d$.

Property 45.7.7. A copula C is an extreme value copula if and only if it is stable in the following sense:

$$C(u_1, \dots, u_d) = C\left(u_1^{1/n}, \dots, u_d^{1/n}\right)^n \tag{45.109}$$

for all $n \geq 1$.

45.8 Randomness ♣

This section is strongly related to Section 76.1 on computability theory.

Definition 45.8.1 (Kolmogorov randomness). Consider a universal Turing machine U. The **Kolmogorov complexity** $C(\kappa)$ of a finite bit string κ (with respect to U) is defined as

$$C(\kappa) := \min\{|\sigma| \mid \sigma \text{ is finite} \wedge U(\sigma) = \kappa\}\,. \tag{45.110}$$

A finite bit string is said to be Kolmogorov random (with respect to U) if there exists an integer $n \in \mathbb{N}$ such that $C(\kappa) \geq |\sigma| - n$.

Property 45.8.2. For every universal Turing machine, there exists at least one Kolmogorov random string. This easily follows from the pigeonhole principle since, for every $n \in \mathbb{N}$, there are 2^n strings of length n, but only $2^n - 1$ programs of length less than n.

Remark 45.8.3. Note that, although universal Turing machines can emulate each other, the randomness of a string is not absolute. Its randomness depends on the chosen machine.

It would be satisfying if this notion of randomness could easily be extended to infinite bit strings by, for example, giving such a string the label 'random' if there exists

[5] The lower bound is only a copula for $d = 2$. In general, this bound is only pointwise sharp.

a uniform choice of constant $k \in \mathbb{N}$ such that all initial segments of the string are k-random. However, by a result of *Martin-Löf*, there does not exist any string satisfying this condition.

45.9 Optimal transport

In this section, a new notion of atomicity of measures will be used.

Definition 45.9.1. A measure on $\mathbb{R}^n$ is said to **give mass to small sets** if there exists a subset of *Hausdorff dimension* $n-1$ (or smaller) that has nonzero measure.

45.9.1 Kantorovich duality

The problem of optimal transport consists of the search for the most cost-efficient transportation scheme that connects a set of producers to a set of consumers. Assume that these are described by the probability spaces (X, Σ_X, μ_X) and (Y, Σ_Y, μ_Y), respectively.

Definition 45.9.2 (Cost function). A measurable function $X \times Y \to \overline{\mathbb{R}}$.

Definition 45.9.3 (Transportation scheme[6]). A joint distribution $\pi \in \mathbb{P}(X \times Y)$ whose marginals coincide with μ_X and μ_Y.

Definition 45.9.4 (Monge–Kantorovich problem). The optimal transportation scheme for a given cost function, according to *Kantorovich*, is the solution of the following optimization problem:

$$c_{\text{opt}} = \inf_{\pi \in \mathbb{P}(X \times Y)} \mathsf{E}_\pi[c]\,. \tag{45.111}$$

The original problem of optimal transportation was considered by *Monge*. However, he studied a restricted problem, where every producer only delivers to a unique consumer. In this case, the joint distributions have a specific form:

$$\int_{X \times Y} c(x,y)\, d\pi(x,y) = \int_X c(x, T(x))\, d\mu_X(x) \tag{45.112}$$

for some measurable function $T : X \to Y$ with $T_* \mu_X = \mu_Y$.

Example 45.9.5 (Finite state spaces). Consider the case where both X and Y are finite with the same cardinality and are both equipped with the uniform distribution. In this case, the joint distributions π can be represented by **bistochastic matrices**, i.e. matrices with nonnegative entries such that every column and every row sum to one. This also implies that the optimization problem reduces to a linear problem on a convex, compact subset. This allows one to use Choquet's theorem 25.3.20 to restrict the attention to the extremal points, which, in this case, are given by permutation matrices.

[6] Also called a **transference plan**.

Hence, the optimal solution is given by the optimal one-to-one pairing of producers and consumers.

Definition 45.9.6 (c-concave function). Consider a cost function $c : X \times Y \to \overline{\mathbb{R}}$. A function $f : \mathbb{R}^n \to [-\infty, +\infty[$, not identically $-\infty$, is said to be c-concave if there exists a set $A \subset \mathbb{R}^n \times \mathbb{R}$ such that

$$f(x) = \inf_{(x',\lambda)\in A} c(x, x') + \lambda \,. \tag{45.113}$$

More generally, two cost functions $c_X : X \to [-\infty, +\infty[$, $c_Y : Y \to [-\infty, +\infty[$ are said to be **competitive** (with respect to c) if

$$c_Y(y) - c_X(x) \leq c(x, y) \tag{45.114}$$

for μ_X-almost all $x \in X$ and μ_Y-almost all $y \in Y$. Their c-transforms are given by

$$\begin{aligned} c_Y^c(x) &= \sup_{y\in Y}\big(c_Y(y) - c(x, y)\big) \,, \\ c_X^c(y) &= \inf_{x\in X}\big(c_X(x) + c(x, y)\big) \,. \end{aligned} \tag{45.115}$$

Functions of the form on the first line are called **c-convex**.

Example 45.9.7 (Legendre transformation). By Definition 16.7.7, it is clear that $c(x, y) = -xy$ leads to the classic notion of convexity (Definition 16.7.3).

Example 45.9.8 (Lipschitz continuity). Let (M, d) be a metric space. The d-convex functions are exactly the 1-Lipschitz functions (Definition 9.2.12). More generally, it can be shown that c-convex functions, for general c, have the same modulus of continuity (Definition 9.2.11) as c, where c is defined on the product space with the taxicab metric (Example 9.3.4).

Theorem 45.9.9 (Kantorovich duality). *Let X, Y be Polish spaces (Definition 18.1.55) with probability measures μ and ν, respectively, and let the cost function $c : X \times Y \to \overline{\mathbb{R}^+}$ satisfy*[7]

1. *lower semicontinuity (Definition 6.2.11), or*
2. *$\mu \otimes \nu$-a.s. boundedness.*

Denote by $\mathbb{P}_{Borel}(\mu, \nu)$ the space of Borel measures on $X \times Y$ whose marginals are given by μ and ν. Moreover, denote by $\Phi_c \subseteq \mathcal{L}^1(X) \times \mathcal{L}^1(Y)$ the competitive integrable functions. Then,

$$\inf_{\pi\in\mathbb{P}_{Borel}(\mu,\nu)} \int_{X\times Y} c \, d\pi = \sup_{(c_X,c_Y)\in\Phi_c} \int_Y c_Y \, d\mu + \int_X c_X \, d\nu \tag{45.116}$$

and the problem admits a solution. Moreover, one can restrict the space of would-be solutions on the right-hand side to those that are

1. *bounded and continuous,*

[7] These are merely sufficient conditions for strong duality.

2. *c-transforms* (c_X, c_X^c),

without changing the solution.

Definition 45.9.10 (Kantorovich distance). Let X be a Polish space (or even Radon space) and consider a metric d on X. The Kantorovich(–Rubinstein) distance $\mathcal{T}_d$ between two Borel probability measures μ, ν on X is defined as the optimal transport cost between them:

$$\mathcal{T}_d(\mu, \nu) := \inf_{\pi \in \mathbb{P}_{\text{Borel}}(X \times X)} \int_{X \times X} d(x, x')\, d\pi(x, x')\,. \tag{45.117}$$

If the metric d is the one inducing the Polish topology, one obtains the definition of the **Wasserstein 1-metric**. This metric can be generalized to complete metric spaces by restricting to the subset of Radon measures.

Kantorovich duality together with Example 45.9.8 imply the following.

Theorem 45.9.11 (Kantorovich–Rubinstein). *Kantorovich duality for $X = Y$ and $c = d$ implies that the Kantorovich distance is equivalently given by*

$$\mathcal{T}_d(\mu, \nu) = \sup\left\{ \mathsf{E}_\mu[\phi] - \mathsf{E}_\nu[\phi] \mid \varphi \in \mathsf{Lip}_1(X, d) \cap \mathcal{L}^1(\mu) \cap \mathcal{L}^1(\nu) \right\}. \tag{45.118}$$

Property 45.9.12 (Translation invariance). The Kantorovich distance is invariant under translations by finite measures.

Property 45.9.13. When $X = Y = \mathbb{R}^n$ with d the Euclidean metric, the Kantorovich distance admits yet another description. In this case, the Lipschitz norm is equal to the supremum norm of the gradient. This gives

$$\mathcal{T}_d(\mu, \nu) = \inf\left\{ \|\sigma\|_1 \,|\, \nabla \cdot \sigma = \mu - \nu \right\}, \tag{45.119}$$

where the condition on σ makes sense by the Riesz–Markov theorem 19.1.6.

45.9.2 Convex costs

This section treats cost functions of the form

$$c(x, y) = h(x - y) \tag{45.120}$$

for some convex function $h : \mathbb{R}^n \to \mathbb{R}$. Moreover, the function h will be assumed to be at least differentiable with locally Lipschitz gradient.

Theorem 45.9.14 (Gangbo–McCann). *If c is strictly convex and μ does not give mass to small sets, the Monge–Kantorovich problem has a a.s. unique minimizer $\pi = (\mathbb{1} \times T)_* \mu$ with*

$$T(x) = x - (\nabla h)^{-1} \nabla \psi(x) \tag{45.121}$$

for some h-concave function $\psi : \mathbb{R}^n \to \overline{\mathbb{R}}$.

Remark 45.9.15. If h is a strictly convex function of the distance $\|x - y\|$, the theorem has to be modified:

- If $\mu \perp \nu$, the theorem still holds.
- If the measures are not singular, one has to restrict to transportation schemes that fix the shared mass. In effect, one removes the shared mass from the problem to recover the previous case.

Note that if h is sufficiently differentiable, the inverse ∇h^{-1} is equal to the gradient of the Legendre transform by Property 16.7.8.

45.9.3 Concave costs

This section treats cost functions of the form

$$c(x, y) = g(\|x - y\|) \tag{45.122}$$

for some concave function $g : \mathbb{R} \to \mathbb{R}$.

Property 45.9.16. Let c be strictly concave. If the transportation cost is not everywhere infinite and if μ does not give mass to small sets, then:

- If $\mu \perp \nu$, there exists a unique optimal transport scheme such that $\nu = T_* \mu$ with
$$T(x) = x - (\nabla g)^{-1} \nabla \varphi(x) \tag{45.123}$$
for some c-concave function φ.
- If the measures are not singular, there still exists a unique optimum by restricting to those schemes that fix shared mass.

45.9.4 Densities

Property 45.9.17 (Continuity equation). Let X be a complete smooth manifold and consider a family $(T_t)_{0 \le t \le 1}$ of locally Lipschitz diffeomorphism on X such that $T_0 = \mathbb{1}_X$ with associated vector fields v_t. If μ is a probability measure on X, the family $(\mu_t := T_{t,*}\mu)_{0 \le t \le 1}$ uniquely satisfies the **continuity equation**[8]:

$$\frac{\partial \mu_t}{\partial t} + \nabla \cdot (\mu_t v_t) = 0\,, \tag{45.124}$$

where the divergence of a measure is defined by duality.

Let $v : \mathbb{R}^n \to \mathbb{R}^n$ be an almost everywhere smooth vector field. This induces a linear, constant-velocity flow as follows:

$$T_t(x) := x - tv(x)\,. \tag{45.125}$$

[8] For a physical incarnation, see Eq. (52.64).

If all T_t are diffeomorphisms, the Eulerian velocity field $v_t(x) := T_t^{-1}\big(v(x)\big)$ satisfies the Eulerian continuity equation:

$$\frac{\partial v_t}{\partial t} + (v_t \cdot \nabla) v_t = 0\,. \tag{45.126}$$

Formula 45.9.18 (Hamilton–Jacobi equation). Given a solution of the continuity equation, the associated flow determines an optimal transport scheme for a cost function c if and only if

$$v_0 = -(\nabla c)^{-1} \nabla \psi \tag{45.127}$$

for some c-concave function ψ. Moreover, if $v_t = (\nabla c)^{-1}\nabla u$ for some function $u(t,x)$, then u satisfies the **Hamilton–Jacobi equation** with Hamiltonian c^*:

$$\frac{\partial u}{\partial t} + c^* \nabla u = 0\,. \tag{45.128}$$

In this section, one considers absolutely continuous measures with respect to the Lebesgue measure on $\mathbb{R}^n$:

$$d\mu_X = \rho_0 dx \qquad\qquad d\mu_Y = \rho_1\, dx\,. \tag{45.129}$$

The transport cost in the Monge problem can then be rewritten as

$$\int_{\mathbb{R}^n} c(x, T(x)) \rho_0(x)\, dx \tag{45.130}$$

with

$$\int_{T^{-1}(A)} \rho_0(x)\, dx = \int_A \rho_1(x)\, dx \tag{45.131}$$

for all measurable $A \subset \mathbb{R}^n$. By the change-of-variables formula, this (weak) integral equation is equivalent to the Jacobian equation for

$$\det\big(DT(x)\big)\rho_1\big(T(x)\big) = \rho_0(x)\,. \tag{45.132}$$

Example 45.9.19 (Euclidean metric). If the cost function c is the square of the Euclidean distance, the optimal transport mapping T, called the **Brenier map**, is given by the gradient of a convex potential:

$$T(x) = \nabla \varphi(x)\,, \tag{45.133}$$

and the optimal cost is equal to the square of the **Wasserstein 2-metric**:

$$\mathcal{T}_{\|\cdot\|_2^2}(\rho_0, \rho_1) = \inf_{\pi \in \mathbb{P}_{\text{Borel}}(\rho_0,\rho_1)} \int_{\mathbb{R}^n} \|x - x'\|\, d\pi(x,x') = W_2^2(\rho_0, \rho_1)\,. \tag{45.134}$$

Moreover, this minimum is unique a.e. It can also be shown that the flow acts affinely:

$$\sigma_t(x) = t\nabla\Phi(x) + (1-t)x\,. \tag{45.135}$$

In fact, the affinity of the flow can be shown more generally.

Property 45.9.20. Consider the time-dependent Monge–Kantorovich problem. If the differential cost $c : \mathbb{R}^n \to \mathbb{R}$ is strictly convex, the flows are given by straight lines:

$$x_t = x + t(x' - x)\,. \tag{45.136}$$

This situation can be generalized to (complete) smooth manifolds, where the minimizers of ℓ^p-costs are geodesics with arc length parametrization.

It is possible to relate optimal transport to mechanics (Section 52.4.1) in the following way.

Method 45.9.21 (Benamou–Brenier formulation). Let ρ_0 and ρ_1 describe the density of particles in a system at time steps $t = 0$ and $t = 1$. Assume that there exists a time-dependent velocity field $v : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$. These are related by the continuity equation (Property 45.9.17):

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0\,. \tag{45.137}$$

The optimization problem now becomes minimizing the *action* or *kinetic energy*:

$$K(\rho, v) := \frac{1}{2}\int_{\mathbb{R}^n}\int_0^T \rho(t,x)\|v(t,x)\|^2\,dt\,dx. \tag{45.138}$$

By making the change of variables $(\rho, v) \longrightarrow (\rho, m := \rho v)$, one obtains a convex problem with a linear constraint (the continuity equation).

Property 45.9.22. The infimum of the Benamou–Brenier action is equal (up to constant factors) to the square of the Wasserstein 2-metric and, hence, gives an equivalent characterization of the Monge–Kantorovich problem for the Euclidean distance.

45.10 Categorical probability ♣

45.10.1 Probability monads

Definition 45.10.1 (Giry monad). Consider the category **Meas** of measurable spaces. On this category, one can define a monad (Definition 3.3.17) that sends a measurable space X to its collection of probability distributions equipped with the σ-algebra generated by all evaluation maps ev_U, where U runs over the measurable subsets of X. Measurable functions are sent to pushforwards.

The unit of the Giry monad $\mathbb{P}$ is defined by assigning Dirac measures:

$$\eta_X(x) := \delta_x\,. \tag{45.139}$$

The multiplication map is given by marginalization:

$$\mu_X(Q)(U) := \int_{P \in \mathbb{P}(X)} \mathsf{ev}_U(P)\, dQ\,. \tag{45.140}$$

Now, consider the Kleisli category (Definition 3.3.21) of the Giry monad. This category has the same objects as **Meas**, but the morphisms $X \to Y$ are given by measurable functions $X \to \mathbb{P}(Y)$. Functions of this type are called **Markov kernels** or **stochastic maps**. This Kleisly category is often denoted by **Stoch**. Composition in **Stoch** boils down to the **Chapman–Kolmogorov equation**:

$$(f \circ_{\mathsf{Stoch}} g)(U \mid x) = \int_Y f(U \mid y)\, dg(y \mid x)\,. \tag{45.141}$$

Global elements in this category are simply probability distributions.

A finitary (or countable) version exists as well. In this case, one considers the category **FinStoch** of finite sets with **stochastic matrices** as morphisms, i.e. matrices $f \in \mathsf{M}_{m,n}(\mathbb{R})$ such that $\sum_{j \le n} f_{ij} = 1$. In this setting, the Chapman–Kolmogorov equations becomes

$$gf(k \mid i) = \sum_{j \le n} g(k \mid j) f(j \mid i)\,, \tag{45.142}$$

where the suggestive notation $f(j \mid i) := f_{ij}$ was again used.

By restricting to Radon measures, an important submonad is found.

Definition 45.10.2 (Kantorovich monad). Consider the category of complete metric spaces and (1-)Lipschitz functions. The Kantorovich monad assigns to every space its set of Radon measures with finite first moment equipped with the Kantorovich–Wasserstein distance.

45.10.2 Markov categories

The structure of the Giry monad and its Kleisli category can be generalized as follows.

Definition 45.10.3 (Markov category). A symmetric semicartesian category (Example 4.1.2) in which every object is an internal commutative comonoid. The comultiplication and counit morphisms $\mathsf{copy}_X : X \to X \otimes X$ and $\mathsf{delete}_X : X \to \mathbf{1}$ are required to be compatible with the monoidal structure in the following way:

$$\begin{aligned} \mathsf{copy}_{X \otimes Y} &= (\mathbb{1}_X \otimes \sigma_{X,Y} \otimes \mathbb{1}_Y) \circ (\mathsf{copy}_X \otimes \mathsf{copy}_Y) \\ \mathsf{copy}_{\mathbf{1}} &= \lambda_{\mathbf{1}}^{-1} \end{aligned} \tag{45.143}$$

for all $X, Y \in \mathsf{ob}(\mathbf{M})$, where σ denotes the braiding. It can be shown that the deletion morphism is natural in X (the copy morphism is not).

Alternative Definition 45.10.4. Instead of requiring $\mathbf{M}$ to be semicartesian, one can also add the following coherence conditions:

$$\mathsf{delete}_X \circ f = \mathsf{delete}_X \tag{45.144}$$

for any endomorphism $f : X \to X$ and

$$\mathsf{delete}_{X \otimes Y} = \mathsf{delete}_X \otimes \mathsf{delete}_Y \tag{45.145}$$

$$\mathsf{delete}_{\mathbf{1}} = \mathbb{1}_{\mathbf{1}} \tag{45.146}$$

for all objects $X, Y \in \mathsf{ob}(\mathbf{M})$, where λ denotes the left unitor.

Notation 45.10.5 (Graphical calculus). As is common for monoidal categories, one can also devise a graphical calculus for Markov categories. The copy and delete morphisms are denoted as follows:

From here on, graphical calculus will be adopted.

Remark 45.10.6 (Strictness). Depending on the strictness of the monoidal structure, the above coherence conditions can be weakened. If $\mathbf{M}$ is strict monoidal, the conditions on $\mathbf{1}$ hold automatically. Moreover, in this case, Eq. (45.144) follows from Eq. (45.143).

Example 45.10.7. Every Cartesian category is Markov in a unique way.

Definition 45.10.8 (Deterministic morphism). A morphism $f : X \to Y$ in a Markov category that satisfies

(45.147)

The name stems from the following intuition: If the morphism would be a random transformation of the input, the two applications on the right-hand side might act differently and, hence, the result would not coincide with the two (identical) copies on the left-hand side.

Definition 45.10.9 (Causal morphism). If the monoidal unit is not required to be terminal, and only coherence conditions (45.145) and (45.146) are required, the resulting structure is sometimes called a **CD**-category (CD stands for copy and discard). A morphism f in a CD-category is said to be causal or **terminal** if

(45.148)

It follows that a Markov category is a CD-category in which all morphisms are causal.

By generalizing the situation of the Giry monad, distributions in a Markov category can be defined.

Definition 45.10.10 (Distribution). Consider a Markov category $(\mathbf{M}, \otimes, \mathbf{1})$. A distribution ψ on an object $X \in \mathsf{ob}(\mathbf{M})$ is a global element $\psi : \mathbf{1} \to X$. Graphically this is depicted as follows:

(45.149)

The category **Stoch**, the motivating example of Markov categories, is clearly not Cartesian. If it were, there would be a bijection between joint distributions $\mathbf{1} \to X \otimes Y$ and their marginals $\mathbf{1} \to X$ and $\mathbf{1} \to Y$. However, it is well known that, in general, probability distributions are not uniquely determined by their marginals (cf. Section 45.7).

To take care of countable samples and asymptotic behaviour, it is necessary to extend the monoidal structure to sequences.

Definition 45.10.11 (Infinite tensor product). Consider a Markov category $\mathbf{M}$. For every two finite subsets $A \subseteq B \subset \mathbb{N}$, one obtains a morphism

$$\bigotimes_{i \in B} X_i \to \bigotimes_{j \in A} X_j \tag{45.150}$$

by applying the deletion morphism to all objects X_k for $k \in B \backslash A$. The infinite tensor product $\bigotimes_{n \in \mathbb{N}} X_n$ is defined as the inverse/filtered limit of this diagram if

1. it exists, and
2. is preserved by all $- \otimes Y$ for $Y \in \mathsf{ob}(\mathbf{M})$.

If all projection morphisms are deterministic, it is called a **Kolmogorov product**.

Chapter 46

Stochastic Calculus

References for this chapter are Elliott and Kopp (2005); Jeanblanc, Yor, and Chesney (2009); Karatzas and Shreve (1991); Salez (2024). The section on abstract Wiener spaces is mainly based on Berger (2002); Dudley, Feldman, and Le Cam (1971); Hairer (2023); Stroock (2008).

Chapter Contents

46.1 Stochastic processes **867**

46.2 Brownian motion **869**

46.2.1 Classical Wiener space 869

46.2.2 Wiener process 870

46.2.3 Abstract Wiener space 871

46.3 Martingales **875**

46.3.1 Proper martingales 875

46.3.2 Local martingales 879

46.4 Itô calculus **880**

46.4.1 Wiener–Paley integral 880

46.4.2 Stochastic integral 882

46.4.3 Properties 885

46.4.4 Lebesgue–Stieltjes integration 889

46.4.5 Representation theorems 890

46.5 Stochastic differential equations **892**

46.6 Markov processes **893**

46.1 Stochastic processes

Definition 46.1.1 (Stochastic process). A sequence of random variables $(X_t)_{t\in T}$ for some index set T. In practice, T will often be a totally ordered set, e.g. $(\mathbb{R}, \leq)$ in the case of a continuous time series.

Definition 46.1.2 (Jump). The jump of a stochastic process $(X_t)_{t\in T}$ at $t \in T$ is defined as

$$\Delta X_t := X_t - X_{t^-}, \tag{46.1}$$

where $X_{t^-} := \lim_{s<t} X_s$.

Definition 46.1.3 (Equivalence). Stochastic process can be be considered equivalent in different ways. The two most common cases are considered here:

- Two stochastic processes $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$ are said to be **stochastically equivalent** if $X_t = Y_t$ a.s. for all $t \in T$. Such processes are also called **modifications** or **versions**.
- Two stochastic processes $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$ are said to be **indistinguishable** or **equivalent up to evanescence** if $X_t = Y_t$ for all $t \in T$ almost surely. Equivalently, they are indistinguishable if their sample paths coincide almost surely.

Definition 46.1.4 (Continuity). A stochastic process $(X_t)_{t\in T}$ on a measurable space (Ω, Σ), where T is a topological space (Chapter 6), is said to be **(sample path) continuous** if the functions $t \mapsto X_t(\omega)$ are continuous for almost all $\omega \in \Omega$.

Property 46.1.5. All continuous stochastic processes are **jointly measurable** when regarded as functions on the product space $T \times \Omega$, where T is equipped with the Borel σ-algebra.

Definition 46.1.6 (Filtered probability space). Consider a probability space (Ω, Σ, P) together with a filtration (Definition 1.6.9) of Σ, i.e. a collection of σ-algebras $\mathbb{F} \equiv (\mathbb{F}_t)_{t\in T}$, such that $i \leq j \implies \mathbb{F}_i \subseteq \mathbb{F}_j$. The quadruple $(\Omega, \Sigma, \mathbb{F}, P)$ is called a filtered probability space. Often, only the filtration and the probability measure are indicated $(\mathbb{F}, P)$, e.g. when the underlying measurable space is of lesser importance.

Often, the filtration is required to be exhaustive and separated (where $\emptyset$ is replaced by $\mathbb{F}_0 = \{\emptyset, \Omega\}$ since any σ-algebra has to contain the total space). Moreover, the filtration will often be required to '**satisfy the usual conditions**' (also called **standard**): $\mathbb{F}_0$ turns the space into a complete measurable space and $\mathbb{F}$ is right-continuous, i.e. $\mathbb{F}_t = \bigcap_{s>t} \mathbb{F}_s$.

Definition 46.1.7 (Adapted process). A stochastic process $(X_t)_{t\in T}$ on a filtered probability space $(\Omega, \Sigma, \mathbb{F}, P)$ is said to be adapted to the filtration $\mathbb{F}$ if X_t is $\mathbb{F}_t$-measurable for all $t \in T$. The smallest filtration with respect to which a stochastic process is adapted, is called its **natural filtration** $\mathbb{F}^X$. For a linearly ordered index set, this is simply the

filtration generated as follows:

$$\mathbb{F}_t^X = \sigma\big(\{X_s\}_{s\leq t}\big)\,. \tag{46.2}$$

Definition 46.1.8 (Progressively measurable). An $\mathcal{X}$-valued stochastic process $(X_t)_{t\in T}$ on a filtered probability space $(\Omega, \Sigma, \mathbb{F}, P)$ is said to be progressively measurable (or simply **progressive**) if for all $t \in T$, the map $T_{\leq t} \times \Omega \to \mathcal{X}$ is measurable with respect to the product σ-algebra $\Sigma_{T_{\leq t}} \otimes \mathbb{F}_t$, where the first factor denotes a suitable σ-algebra on the index set.

Property 46.1.9. Every adapted process with left- (or right-)continuous paths is progressively measurable. Moreover, every adapted process adapts a progressively measurable modification.

Definition 46.1.10 (Predictable process). First, consider $T = \mathbb{N}$. A discrete-time stochastic process $(X_n)_{n\in\mathbb{N}}$ on a filtered probability space $(\Omega, \Sigma, \mathbb{F}, P)$ is said to be predictable if X_{n+1} is $\mathbb{F}_n$-measurable for all $n \in \mathbb{N}$.

More generally, for arbitrary T, consider the **predictable σ-algebra**, i.e. the σ-algebra on $T \times \Omega$ generated by all (left-)continuous, adaptive processes. A process $(X_t)_{t\in T}$ is said to be predictable if it is measurable with respect to the predictable σ-algebra.

Example 46.1.11 (Elementary predictable process).

$$X_t = f_0 \mathbb{1}_{\{0\}}(t) + \sum_{i=1}^{n} f_i \mathbb{1}_{]\tau_{i-1},\tau_i]}(t) \tag{46.3}$$

for a finite sequence of bounded stopping times $\{\tau_i\}_{i\leq n}$, with $\tau_0 \equiv 0$, and bounded $\mathbb{F}_{\tau_i}$-measurable functions f_i. The vector space of elementary predictable processes will be denoted by $\mathcal{E}$.

Property 46.1.12. The predictable σ-algebra is generated by the elementary predictable processes.

Definition 46.1.13 (Increasing). An adapted process $(X_t)_{t\in T}$, where T is linearly ordered and has a minimal element 0, is called increasing if:

1. $X_0 = 0$ a.s.
2. $t \mapsto X_t(\omega)$ is almost surely right-continuous and increasing.

Often, integrability of all X_t is also required. If $\lim_t X_t$ is integrable, the increasing process itself is called **integrable**.

Property 46.1.14 (Naturality). An increasing process $(A_t)_{t\in\mathbb{R}^+}$ is predictable if and only if it is **natural**, i.e. when

$$\mathsf{E}\left[\int_0^t X_s\, dA_s\right] = \mathsf{E}\left[\int_0^t X_{s^-}\, dA_s\right] \tag{46.4}$$

for all $t \in T$ and every bounded, right-continuous martingale $(X_t)_{t\in\mathbb{R}^+}$.

Property 46.1.15 (Measurability hierarchy). The following hierarchy shows how the different notions of measurability of stochastic processes are related (from strong to weak):

1. Continuous and adapted,
2. predictable,
3. *optional*,
4. progressively measurable, and
5. adapted (and jointly measurable).

Definition 46.1.16 (Stopping time). Consider a random variable τ on a filtered probability space $(\Omega, \Sigma, \mathbb{F}, P)$ where the codomain of τ coincides with the index set of $\mathbb{F}$. This variable is called a stopping time for $\mathbb{F}$ if

$$\{\tau \leq t\} \in \mathbb{F}_t \tag{46.5}$$

for all $t \in T$. The stopping time is a 'time indicator' that only depends on the knowledge of the process up to time $t \in T$.

Definition 46.1.17 (Stopped process). Consider a stochastic process $(X_t)_{t\in T}$ and a stopping time τ. The stopped process $(X_t^\tau)_{t\in T}$ is defined as follows:

$$X_t^\tau := X_{t\wedge\tau}\,. \tag{46.6}$$

The notions of convergence from Section 45.3 can be generalized to stochastic processes in different ways. The following is the most common one.

Definition 46.1.18 (UCP convergence). A sequence of jointly measurable stochastic processes $(X_{t,n})_{t\in T, n\in\mathbb{N}}$ is said to converge **uniformly on compacta** (ucp) to a stochastic process $(X_t)_{t\in T}$ if

$$\mathsf{Prob}\left(\sup_{s\leq t} |X_{s,n} - X_s| > K\right) \longrightarrow 0 \tag{46.7}$$

when $n \longrightarrow +\infty$ for all $t \in T$ and $K > 0$.

46.2 Brownian motion

46.2.1 Classical Wiener space

Consider the space $C([0,1])$ of continuous functions on the unit interval. A σ-algebra Σ on the function space $C([0,1])$ can be generated by the following sets, called **cylinder sets**:

$$\mathcal{C}_{t_1,\dots,t_n}(B) := \left\{ f \in X^Y \,\middle|\, \big(f(t_1), \dots, f(t_n)\big) \in B \right\}, \tag{46.8}$$

where B is a Borel set, $n \in \mathbb{N}$ and $\{t_1, \dots, t_n\} \subset [0,1]$. Moreover, for every choice of $\{t_1, \dots, t_n\}$, one obtain a σ-algebra $\mathcal{C}_{t_1,\dots,t_n}$ generated by evaluation at these points. On this measurable space $\big(C([0,1]), \mathcal{C}_{t_1,\dots,t_n}\big)$ one can define the Gaussian measure (see also Definition 45.4.14)

$$\mu_{t_1,\dots,t_n}(A) := \frac{1}{(2\pi)^{n/2}\sqrt{t_1(t_2-t_1)\cdots(t_n-t_{n-1})}} \int_{f^n(A)} \exp\left(-\frac{x_1^2}{2t_1} - \cdots \frac{(x_n - x_{n-1})^2}{2(t_n - t_{n-1})}\right) d^n x \tag{46.9}$$

if $t_1 > 0$ and

$$\mu_{t_1,\dots,t_n}(A) := \frac{1}{(2\pi)^{\frac{n-1}{2}}\sqrt{t_2\cdots(t_n-t_{n-1})}} \int_{A_0} \exp\left(-\frac{x_2^2}{2t_2} - \cdots \frac{(x_n - x_{n-1})^2}{2(t_n - t_{n-1})}\right) d^{n-1} x \tag{46.10}$$

if $t_1 = 0$, where $A_0 := \big\{x \in \mathbb{R}^{n-1} \,|\, (0,x) \in f^n(A)\big\}$. In the case $n = 0$, the measure fully degenerates to a δ-measure at 0.

Given these constructions, it can be shown that there exists a unique measure P^W on $\big(C([0,1]), \Sigma\big)$, called the **Wiener measure**, such that

$$P^W(A) = \mu t_1, \dots, t_n(A) \tag{46.11}$$

for all $A \in \mathcal{C}_{t_1,\dots,t_n}$.

This is an application of a more general theorem.

Theorem 46.2.1 (Kolmogorov extension theorem[1]). *@@ ADD KOLMOGOROV EXTENSION THEOREM (in general) @@*

Theorem 46.2.2 (Kolmogorov continuity theorem). *Consider a stochastic process $(X_t)_{t\in\mathbb{R}^+}$ with*

$$\mathsf{E}\big[|X_t - X_s|^\beta\big] \le C\|t-s\|^{1+\alpha} \tag{46.12}$$

for some $\alpha, \beta, C \in \mathbb{R}_0^+$. There exists a version $(\widetilde{X}_t)_{t\in\mathbb{R}^+}$ with a.s. Hölder-continuous paths for Hölder-exponent $\gamma < \frac{\alpha}{\beta}$ (Definition 9.2.13).

46.2.2 Wiener process

Definition 46.2.3 (Lévy process). A stochastic process $(X_t)_{t\in\mathbb{R}^+}$ satisfying:

1. **Initial condition**: $X_0 = 0$ a.s.
2. **Independent increments**: For all $t_1 \le \cdots \le t_n$, $\{X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}\}$ are independent.
3. **Stationary increments**: For all $s \le t$, $X_t - X_s \overset{d}{=} X_{t-s}$.

[1] Also called the **Kolmogorov–Daniell theorem**.

4. **Continuity in probability**: $(X_t)_{t\in\mathbb{R}^+}$ is almost surely càdlàg.

A famous example of a Lévy process is Brownian motion.

Definition 46.2.4 (Brownian motion). A continuous-time stochastic process $(W_t)_{t\in\mathbb{R}^+}$ satisfying:

1. W_0 is almost surely 0.
2. W has independent Gaussian increments.
3. W is almost surely continuous.

Remark 46.2.5 (Wiener process). A Brownian motion is also called a Wiener process (especially in the mathematics literature). The reason for this terminology is that an explicit construction of Brownian motion can be given through the Wiener measure P^W defined in the previous section. The process

$$B_t : \mathbb{R}^{[0,1]} \to \mathbb{R} : f \mapsto f(t) \tag{46.13}$$

is a Brownian motion on $[0,1]$, i.e. P^W is the law of the $\mathbb{R}^{[0,1]}$-valued random variable B.

Definition 46.2.6 (Geometric Brownian motion).

$$M_t := \exp\left(W_t - \frac{t}{2}\right), \tag{46.14}$$

where $(W_t)_{t\in\mathbb{R}^+}$ is a standard Brownian motion.

Definition 46.2.7 (Wiener measure). Theorem 46.2.2 implies that there exists a continuous version $(\widetilde{B}_t)_{t\in\mathbb{R}^+}$ of Brownian motion. The induced map

$$\widetilde{B} : \mathbb{R}^{[0,1]} \to C([0,1]) : f \mapsto \big(t \mapsto \widetilde{B}_t(f)\big) \tag{46.15}$$

is measurable. As such, one obtains the pushforward measure $\gamma^W := \widetilde{B}_* P^W$ on $C([0,1])$, also called the Wiener measure. The measure space $\big(C([0,1]), \mathcal{B}\big(C([0,1])\big), \gamma^W\big)$ is also called the **classical Wiener space**.

46.2.3 Abstract Wiener space

Theorem 46.2.8 (Cameron–Martin). *Consider the Gaussian measure*

$$\gamma_n(A) := \frac{1}{(2\pi)^{n/2}} \int_A \exp\left(-\frac{\|x\|^2}{2}\right) d^n x\,. \tag{46.16}$$

By Haar's theorem 19.4.3, the Lebesgue measure is the only measure that is translation invariant. The pushforward of γ_n under translation over $v \in \mathbb{R}^n$ is given by the following Radon–Nikodym derivative:

$$\frac{\mathrm{d}v_*\gamma_n}{\mathrm{d}\gamma_n}(x) = \exp\left(\langle v, x\rangle - \frac{\|v\|^2}{2}\right). \tag{46.17}$$

This theorem can be generalized to locally convex vector spaces X (separable Hilbert spaces, in particular). To this end, a measure γ on an LCTVS X is said to be **Gaussian** if, for every $f \in X'$, $f_*\gamma$ is a normal distribution on $\mathbb{R}$. However, on a general infinite-dimensional separable Hilbert space, the definition of the Gaussian measure does not make sense. Since, by the (*strong*) *law of large numbers* (see Theorem 48.1.4), the norm

$$\|x\| = \sum_{n=1}^{+\infty} \langle x \,|\, e_i \rangle^2 \tag{46.18}$$

would diverge almost surely because the individual coefficients, where e_i are orthonormal, would be independent under the Gaussian measure.

The first step in solving this issue is to generalize the notion of cylinder sets (Section 46.2.1) has to be generalized by replacing pointwise evaluations with linear functionals[2] $\phi \in X'$. Two equivalent approaches exist, through finite collections of anchor points or through finite-dimensional projections.

Definition 46.2.9 (Cylinder set). Consider a locally convex vector space X (Definition 25.3.8) and its topological dual X'. A cylinder set in X is a subset of the form

$$C_{\phi_1,\dots,\phi_n}(B) := \left\{ x \in X \,\middle|\, \big(\phi_1(x), \dots, \phi_n(x)\big) \in B \right\}, \tag{46.19}$$

where $\{\phi_1, \dots, \phi_n\} \subset X'$ and $B \in \mathcal{B}(\mathbb{R}^n)$. The σ-algebra generated by all cylinder sets will be denoted by $\mathcal{C}(X)$, whereas the σ-algebra consisting of cylinder sets for a fixed choice of $\{\phi_1, \dots, \phi_n\} \subset X'$ will be denoted by $\mathcal{C}_{\phi_1,\dots,\phi_n}(X)$.[3] $\mathcal{C}(X)$ is, equivalently, the minimal σ-algebra that turns all linear functionals measurable.

The cylinder sets can be defined equivalently as follows. Consider a finite-dimensional subspace $K \leq X'$ and the associated projection $\pi_K : X \to X/K^\perp$, where $K^\perp$ is the orthogonal complement as in Definition 22.2.23. Cylinder sets are equivalently of the form $\pi_K^{-1}(B) \cong B + K^\perp$ for $B \in \mathcal{B}(K')$.

A **cylinder set measure** is a finitely additive function $\mathcal{C}(X) \to \mathbb{R}^+$ that restricts to a proper measure on any sub-σ-algebra $\mathcal{C}_{\phi_1,\dots,\phi_n}(X)$.

Property 46.2.10. On a separable Banach space X, the Borel σ-algebra $\mathcal{B}(X)$ coincides with the cylindrical σ-algebra $\mathcal{C}(X)$ with respect to its topological dual X'.

Now, given a separable Hilbert space $\mathcal{H}$ and a Gaussian cylinder set measure μ, one can try to find a separable Banach space $X \supseteq \mathcal{H}$ such that μ extends to a proper measure γ satisfying

$$\gamma(B) = \mu(B \cap \mathcal{H}) \tag{46.20}$$

[2] In fact, one could consider any dual pair (M, N).

[3] Note the terminology. For fixed indexing functionals, one readily obtains a σ-algebra, whereas the collection of all cylinder sets is only an algebra of sets (Definition 1.6.11). It is sometimes called a **cylindrical algebra**.

for all $B \in \mathcal{B}(X)$. To construct X, one takes the completion of $\mathcal{H}$ with respect to a norm that interacts nicely with μ.[4]

Definition 46.2.11 (Measurable norm). A (continuous) norm $\|\cdot\|$ on a topological vector space X is said to be (Gross-)measurable (with respect to a cylinder set measure μ) if, for every $\varepsilon > 0$, there exists a finite-dimensional subspace $V_\varepsilon \leq X$ such that for all finite-dimensional subspaces $V \leq X$ the following holds:

$$V \perp V_\varepsilon \implies \mu\big(\{v \in V \mid \|v\| > \varepsilon\}\big)\varepsilon\,. \tag{46.21}$$

Example 46.2.12 (Hilbert–Schmidt operators). Consider a separable Hilbert space $\mathcal{H}$ with norm $\|\cdot\|$. For every Hilbert–Schmidt operator $T : \mathcal{H} \to \mathcal{H}$ (Example 26.1.16), the induced norm

$$\|\cdot\|_T := \|T(\cdot)\| \tag{46.22}$$

is measurable with respect to the canonical Gaussian measure on $\mathcal{H}$. Moreover, any other measurable Hilbert norm on $\mathcal{H}$ is exactly of this form.

Property 46.2.13. Every separable Banach space is the Gross-completion of a separable Hilbert space with respect to some measurable norm.

Now, given a Gaussian measure γ on an LCTVS X, one can define the moments of functionals $\phi \in X'$ in analogy with Section 45.3.2 and Theorem 45.3.8:

$$\begin{aligned} \mathsf{E}\big[\phi\big] &:= \int_X \phi\, d\gamma\,, \\ \mathsf{Cov}[\phi,\psi] &:= \int_X \big(\phi - \mathsf{E}\big[\phi\big]\big)\big(\psi - \mathsf{E}\big[\psi\big]\big)\, d\gamma\,. \end{aligned} \tag{46.23}$$

These are, respectively linear and symmetric, bilinear and nondegenerate on X'. Moreover, if X is a separable Banach space, they are continuous as well. The centered functionals (or **measurable linear functionals**), i.e. those with $\mathsf{E}\big[\phi\big] = 0$, form a designated subspace:

$$X'_\gamma := \overline{\{\phi - \mathsf{E}\big[\phi\big] \mid \phi \in X'\}}^{L^2}\,. \tag{46.24}$$

The L^2-closure makes sense since the inner product induced by the covariance operator reduces to the standard $L^2(\gamma)$-inner product when restricting to centered elements.

Theorem 46.2.14. *A measure γ on a locally convex measurable space $(X, \mathcal{C}(X))$ is Gaussian if and only if its Fourier transformation (i.e. its characteristic function 45.3.31) has the form*

$$\widetilde{\gamma}(f) = \exp\left(i\mathsf{E}\big[f\big] - \frac{\mathsf{Cov}[f,f]}{2}\right). \tag{46.25}$$

[4]The following property of norms is merely a sufficient condition for the extension to be possible.

Remark 46.2.15. In the literature, one often focuses on centered Gaussian measures, i.e. where the characteristic function only contains the covariance operator. In this situation, $X'_\gamma = X'$.

Definition 46.2.16 (Cameron–Martin space). Consider a Gaussian measure γ on a locally convex vector space X. The covariance operator Cov induces a norm on X:

$$\|x\|_\gamma := \sup\{\phi(x) \mid \phi \in X', \mathsf{Cov}[\phi, \phi] \leq 1\}\,. \tag{46.26}$$

The Cameron–Martin space of γ is defined as follows:

$$H(\gamma) := \{x \in X \mid \|x\|_\gamma < +\infty\}\,. \tag{46.27}$$

Remark 46.2.17. If X was obtained as the Gross-completion of a separable Hilbert space $\mathcal{H}$, then $\mathcal{H} \cong H(\gamma)$ and $\|\cdot\|_\gamma = \|\cdot\|_{\mathcal{H}}$.

Property 46.2.18. Let X be a separable Banach space. For every centered element $h \in X'_\gamma$, there exists an $x \in X$ such that

$$\mathsf{Cov}[h, \cdot] = \langle \cdot, x \rangle\,. \tag{46.28}$$

More specifically, $\mathsf{Cov} : X'_\gamma \to X$ characterizes $H(\gamma)$:

$$H(\gamma) = \mathsf{Im}(\mathsf{Cov})\,. \tag{46.29}$$

An explicit map is given by the Bochner integral (Section 18.8). The image of $\phi \in X'_\gamma$ is given by

$$\int_X (x - \mathsf{E}[\gamma])\phi(x)\, d\gamma(x)\,. \tag{46.30}$$

Remark 46.2.19. This property follows from the general fact that, given a dense embedding $\mathcal{H} \hookrightarrow \mathcal{B}$ of a Hilbert space in a Banach space, the dual $\mathcal{B}'$ can be densely embedded into $\mathcal{H}$. Every linear form on $\mathcal{B}$ gives a form on $\mathcal{H}$ by restriction. By Riesz's representation theorem 25.2.7, one obtains for every such form $\phi|_{\mathcal{H}}$ an element x_ϕ such that

$$\phi|_{\mathcal{H}}(x) = \left\langle x \,\middle|\, x_\phi \right\rangle.$$

Now, since $\mathcal{H}$ is dense in $\mathcal{B}$, this extends uniquely to $\mathcal{B}$.

Property 46.2.20. $H(\gamma)$ is a separable Hilbert subspace of X and, in particular, a *reproducing kernel Hilbert space* (see Section 50.6.1). However, $\|\cdot\|_X$ and $\|\cdot\|_\gamma$ are not equivalent (the former giving an RKHS, the latter giving it the structure of the Cameron–Martin space).

Theorem 46.2.8 can now be generalized to separable Banach spaces.

Theorem 46.2.21 (Cameron–Martin). *Let X be a separable Banach space. For $x \in X$ and a Gaussian measure γ on X, the translated measure*

$$x_*\gamma(B) := \gamma(B - x) \tag{46.31}$$

is equivalent to γ if and only if $x \in H(\gamma)$ and, in this case, the associated Radon–Nikodym derivative is given by

$$\frac{\mathrm{d}x_*\gamma}{\mathrm{d}\gamma} = \exp\left(h(x) - \frac{\|x\|_\gamma^2}{2}\right), \tag{46.32}$$

where $h(x) \in X'_\gamma$ is the centered element corresponding to x by Property 46.2.18.

The ideas above can be abstracted and formalized as follows.

Definition 46.2.22 (Abstract Wiener space). A triple $(\mathcal{H}, \mathcal{B}, \gamma)$ such that:

1. $\mathcal{H}$ is a separable Hilbert space, called the Cameron–Martin space.
2. $\mathcal{B}$ is a separable Banach space containing $\mathcal{H}$ as a dense subspace.
3. γ is a Gaussian measure on $\mathcal{B}$ with characteristic functional

$$\widetilde{\gamma}(\phi) \equiv \int_{\mathcal{B}} \exp(i\phi)\, d\gamma := \exp\left(-\frac{\|\iota^*(\phi)\|_{\mathcal{H}}^2}{2}\right), \tag{46.33}$$

   where $\iota^* : \mathcal{B}' \to \mathcal{H}' \cong \mathcal{H}$ is the dual of the canonical injection.

Note that it is the Cameron–Martin space that determines the structure.

Although the Cameron–Martin space determines the structure of an abstract Wiener space, it is small in the following sense.

Property 46.2.23. Consider an abstract Wiener space $(\mathcal{H}, \mathcal{B}, \gamma)$. If $\mathcal{H}$ is infinite dimensional, then $\gamma(\mathcal{H}) = 0$.

46.3 Martingales

From here on, the index set T will be $\mathbb{N}$ or $\mathbb{R}^+ \equiv [0, +\infty[$, so that the index t can be interpreted as a time parameter. The explicit choice will be made clear if necessary.

46.3.1 Proper martingales

Definition 46.3.1 (Martingale). A stochastic process $(X_t)_{t\in T}$ on a filtered probability space $(\Omega, \Sigma, \mathbb{F}, P)$ is called an $(\mathbb{F}, P)$-martingale if it satisfies the following conditions:

1. $(X_t)_{t\in T}$ is $\mathbb{F}$-adapted.

2. For all $t \in T$: $X_t \in L^1(P)$.
3. For all $t > s \geq 0$: $\mathsf{E}\left[X_t \,|\, \mathbb{F}_s\right] = X_s$.

If the equality in the last condition is replaced by the inequality $\leq$ (resp. $\geq$), the stochastic process is called a **supermartingale** (resp. **submartingale**).

Example 46.3.2 (Doob martingale). Consider an integrable random variable X and a filtration $\mathbb{F}$. The associated Doob martingale with respect to $(\mathbb{F}, P)$ is given by

$$Y_t := \mathsf{E}\left[X \,|\, \mathbb{F}_t\right]. \tag{46.34}$$

The fact that this is a martingale follows from the law of iterated expectations.

Remark 46.3.3 (Càdlàg). It can be shown that almost surely right-continuous submartingales (and supermartingales) with respect to a filtration that satisfies the usual conditions, always admit a càdlàg modification. From here one, all submartingales will assumed to be càdlàg unless mentioned otherwise.

Theorem 46.3.4 (Optional stopping). *Let $(X_t)_{t\in T}$ be a martingale. If τ is a stopping time satisfying either of the following conditions:*

1. *τ is almost surely bounded, or*
2. *the stopped process $(X_t^\tau)_{t\in T}$ is almost surely bounded,*

then the stopped process $(X_t^\tau)_{t\in T}$ is again a martingale.

The following result is an example of weak type inequalities (Definition 18.3.14).

Property 46.3.5 (Doob–Ville inequality). Consider a càdlàg submartingale $(X_t)_{t\in T}$.

$$\mathsf{Prob}\left(\sup_{t\leq\tau} X_t \geq \lambda\right) \leq \frac{\mathsf{E}\left[\max(0, X_\tau)\right]}{\lambda} \tag{46.35}$$

for all $\lambda \geq 1$ and $\tau \in T$.

The following property generalizes the Hoeffding inequalities Property 45.3.30.

Property 46.3.6 (Hoeffding–Azuma inequality). Let $(X_n)_{n\in\mathbb{N}}$ be a (super)martingale with bounded differences, i.e. there exist constants $c_k \in \mathbb{R}$ such that

$$|X_k - X_{k-1}| \leq c_k. \tag{46.36}$$

The following inequality holds for all $\lambda \geq 0$:

$$\mathsf{Prob}\left(X_N - X_0 \geq \lambda\right) \leq \exp\left(-\frac{\lambda^2}{2\sum_{i=1}^N c_i^2}\right). \tag{46.37}$$

A symmetric result for the lower tail holds for (sub)martingales. Moreover, if there exist predictable processes $(A_n)_{n\in\mathbb{N}}, (B_n)_{n\in\mathbb{N}}$ such that

$$A_k \leq X_k - X_{k-1} \leq B_k \tag{46.38}$$

and

$$B_k - A_k \leq c_k \tag{46.39}$$

for all $k \in \mathbb{N}$, the inequality can be sharpened:

$$\mathsf{Prob}\big(X_N - X_0 \geq \lambda\big) \leq \exp\left(-\frac{2\lambda^2}{\sum_{i=1}^N c_i^2}\right). \tag{46.40}$$

Now, consider a function $f : \Omega^n \to \mathbb{R}$ such that

$$\sup_{x_1,\dots,x_n,x_k'} |f(x_1,\dots,x_k,\dots,x_n) - f(x_1,\dots,x_k',\dots,x_n)| \leq c_k \tag{46.41}$$

for all $k \in \mathbb{N}$. By applying the above inequalities to the Doob martingale

$$Z_m := \mathsf{E}\big[f(X_1,\dots,X_n) \mid X_1,\dots,X_m\big]\,, \tag{46.42}$$

one obtains the following inequality:

$$\mathsf{Prob}\big(f(X_1,\dots,X_n) - \mathsf{E}\big[f\big] \geq \lambda\big) \leq \exp\left(-\frac{2\lambda^2}{\sum_{i=1}^n c_i^2}\right). \tag{46.43}$$

This inequality is sometimes called the **McDiarmid inequality**.

Theorem 46.3.7 (Doob decomposition). *Any integrable, adapted process $(X_n)_{n\in\mathbb{N}}$ can be decomposed (almost surely uniquely) as*

$$X_n = X_0 + M_n + A_n\,, \tag{46.44}$$

where $(M_n)_{n\in\mathbb{N}}$ is a martingale and $(A_n)_{n\in\mathbb{N}}$ is an integrable, predictable process. These two processes are constructed iteratively as follows:

$$\begin{aligned} A_0 &= 0 & M_0 &= 0 \\ \Delta A_n &= \mathsf{E}\big[\Delta X_n \mid \mathbb{F}_{n-1}\big] & \Delta M_n &= \Delta X_n - \Delta A_n\,. \end{aligned} \tag{46.45}$$

Furthermore, $(X_n)_{n\in\mathbb{N}}$ is a submartingale if and only if $(A_n)_{n\in\mathbb{N}}$ is (almost surely) increasing.[5]

Definition 46.3.8 (Quadratic variation). Consider the special case $X_n = Y_n^2$ for some integrable martingale $(Y_n)_{n\in\mathbb{N}}$. By Jensen's inquality 16.7.6, $(X_n)_{n\in\mathbb{N}}$ is an integrable submartingale. The process $(A_n)_{n\in\mathbb{N}}$ in the Doob decomposition of $(X_n)_{n\in\mathbb{N}}$ is often called the **quadratic variation process** of $(Y_n)_{n\in\mathbb{N}}$ and is denoted by $([Y]_n)_{n\in\mathbb{N}}$.

[5] Replacing increasing by decreasing gives a similar statement for supermartingales.

The Doob decomposition has an analogue for $\mathbb{R}^+$-indexed processes.

Theorem 46.3.9 (Doob–Meyer). *Consider the class $\mathcal{T}_a$ of stopping times which are almost surely bounded by $a \in \mathbb{R}$ and a right-continuous submartingale $(T_t)_{t\in T}$ such that $\{X_\tau\}_{\tau\in\mathcal{T}_a}$ is uniformly integrable*[6]*. Then $(X_t)_{t\in T}$ can be decomposed (almost surely uniquely) as*

$$X_t = X_0 + M_t + A_t\,, \tag{46.46}$$

where $(M_t)_{t\in T}$ is a right-continuous martingale and $(A_t)_{t\in T}$ is increasing and natural.

Remark 46.3.10. The DL condition can be replaced by requiring (almost sure) nonnegativity.

As in the case of discrete-time martingales, the quadratic variation could be defined through the Doob–Meyer decomposition of X^2. However, an alternative definition goes as follows.

Alternative Definition 46.3.11 (Quadratic variation). Let $(X_t)_{t\in\mathbb{R}^+}$ be a martingale. For every stochastic partition $P = \{0 = \tau_0 \leq \tau_1 \leq \cdots \nearrow +\infty\}$ of $\mathbb{R}^+$, i.e. sequence of stopping times starting at zero and going to infinity, define the quadratic variation as

$$[X]_t^P := \sum_{i=1}^{+\infty} (X_{t\wedge\tau_i} - X_{t\wedge\tau_{i-1}})^2\,. \tag{46.47}$$

The quadratic variation $[X]$ is obtained by taking the limit over all stochastic partitions as the mesh size

$$|P| := \max_{n\in\mathbb{N}}(t\wedge\tau_n - t\wedge\tau_{n-1}) \tag{46.48}$$

goes to zero in probability. (The limit converges in probability and even ucp.[7]) Moreover, if X is continuous and square-integrable, so is $[X]$ and this construction coincides with that through the Doob–Meyer decomposition.

Remark 46.3.12 (Quadratic covariation). The formula for the quadratic variation can easily be modified to obtain the quadratic covariation:

$$[X,Y]_t^P := \sum_{i=1}^{+\infty} (X_{t\wedge\tau_i} - X_{t\wedge\tau_{i-1}})(Y_{t\wedge\tau_i} - Y_{t\wedge\tau_{i-1}})\,. \tag{46.49}$$

Property 46.3.13 (Linearity). Consider three processes $(X_t)_{t\in\mathbb{R}^+}$, $(Y_t)_{t\in\mathbb{R}^+}$ and $(Z_t)_{t\in\mathbb{R}^+}$. For any two real numbers $a, b \in \mathbb{R}$, the following relation holds:

$$[aX + bY, Z]_t = a[X,Z]_t + b[Y,Z]_t\,. \tag{46.50}$$

Definition 46.3.14 (Orthogonality). Two stochastic processes are said to be orthogonal if their quadratic covariation vanishes:

$$X \perp Y \iff [X,Y] = 0\,. \tag{46.51}$$

[6] Stochastic processes that satisfy this condition for some $a \in \mathbb{R}$ are said to be of **class DL**

[7] For more general processes, the quadratic variation is well defined exactly if this limit exists.

46.3.2 Local martingales

Definition 46.3.15 (Local martingale). Consider a filtered probability space $(\Omega, \Sigma, \mathbb{F}, P)$. A stochastic process $(X_t)_{t\in T}$ is called a local martingale (relative to $\mathbb{F}$) if there exists stopping times $(\tau_n)_{n\in\mathbb{N}}$ (with respect to $\mathbb{F}$) satisfying the following conditions:

1. $(\tau_n)_{n\in\mathbb{N}}$ is almost surely increasing.
2. $(\tau_n)_{n\in\mathbb{N}}$ is almost surely divergent.
3. The stopped process $\{X^{\tau_n}\}_{t\in T}$ is a martingale (relative to $\mathbb{F}$) for all $n \in \mathbb{N}$.

This idea can be generalized to any kind of property of stochastic processes.

Definition 46.3.16 (Localization). Consider a class $\mathcal{P}$ of stochastic processes. A stochastic process $(X_t)_{t\in T}$ is said to be locally in $\mathcal{P}$ if there exists a sequence $(\tau_n)_{n\in\mathbb{N}}$ of stopping times, with $\tau_n \nearrow +\infty$ almost surely, such that the stopped process $(X_t^{\tau_n})_{t\in T}$ is in $\mathcal{P}$ for all $n \in \mathbb{N}$.

Definition 46.3.17 (Local integrability). A stochastic process is said to be locally integrable if its maximum process $X_t^* := \sup_{s\leq t} X_s$ is locally in the class of integrable stochastic processes.

Property 46.3.18. A càdlàg, adapted stochastic process is of class D(L) if and only if it is locally integrable

Property 46.3.19. Every local martingale is locally integrable and a local martingale is a martingale if and only if it is of class DL. (Note that, by the preceding property, this happens exactly when the local martingale is càdlàg and adapted.)

Property 46.3.20 (Convergence). The limit of a sequence of continuous local martingales that converges ucp is itself a continuous local martingale.

Property 46.3.21 (Covariation). If $(X_t)_{t\in T}, (Y_t)_{t\in T}$ are local martingales, then $([X,Y]_t)_{t\in T}$ is the unique FV process that starts at 0 and such that

1. $\left(X_tY_t - [X,Y]_t\right)_{t\in T}$ is a local martingale, and
2. $\Delta[X,Y] = \Delta X \Delta Y$.

Property 46.3.22 (Square-integrability). A local martingale $(X_t)_{t\in\mathbb{R}^+}$ is an L^2-martingale if and only if $\mathsf{E}\left[X_0^2\right] < +\infty$ and $[X]$ is integrable. More generally, a local martingale is a martingale if it is dominated by an integrable random variable.

Definition 46.2.4 was rephrased by *Lévy* in terms of martingale properties.

Alternative Definition 46.3.23 (Brownian motion). A local martingale $(X_t)_{t\in\mathbb{R}^+}$ such that:

1. $X_0 = 0$,

2. X is continuous, and

3. $(X_t^2 - t)_{t\in\mathbb{R}^+}$ is a local martingale or, equivalently by Property 46.3.21, $[X]_t = t$.

Although (local) martingales are often the most interesting objects in practice, they can be generalized in a Doob-like manner. Moreover, these generalized martingales are of major importance for the stochastic calculus that will introduced in the next section.

Definition 46.3.24 (Semimartingale). A real-valued stochastic process that can be decomposed as the sum of a local martingale and a càdlàg, adapted process of locally bounded variation[8].

Remark 46.3.25. The càdlàg, adapted processes of locally bounded variation are also called **FV processes**.

Property 46.3.26. If $(X_t)_{t\in T}$ is a semimartingale, then $(\langle X\rangle_t)_{t\in T}$ is càdlàg, adapted and increasing. Moreover, if $(Y_t)_{t\in T}$ is also a semimartingale, then $(\langle X, Y\rangle_t)_{t\in T}$ is an FV process.

Definition 46.3.27 (Émery topology). The space of semimartingales on a filtered probability space $(\Omega, \Sigma, \mathbb{F}, P)$ admits a natural topology (Chapter 6). Consider the following metric (Chapter 9):

$$d_{\text{sm}}(S,T) := \sup_{H\in\mathcal{E},\|H\|_\infty\leq 1} \mathsf{E}\left[\sup_{t\in\mathbb{R}^+} |H\cdot(S-T)_t| \wedge 1\right], \tag{46.52}$$

where $\mathcal{E}$ denotes the set of elementary predictable process on $(\Omega, \Sigma, \mathbb{F}, P)$. This metric turn Ω into a complete metric space (Definition 9.4.2).

46.4 Itô calculus

Two approaches exist to extending integration theory to stochastic processes. One is through the *Itô isometry* (see Property 46.4.32 below) and the other is through a more axiomatic approach as for the ordinary Lebesgue integral. Both approaches will be explored in this section, with the latter coming first. For simplicity (and unless stated otherwise), all filtered probability spaces will be assumed to satisfy the usual conditions and all martingales will be assumed to be càdlàg.

46.4.1 Wiener–Paley integral

The first step towards defining a notion of stochastic integral will consists of the subclass of integrators on Brownian motions/Wiener processes (Definition 46.2.4).

[8]This means that the sample paths are almost surely of locally bounded variation.

As for the Lebesgue measure, one starts with the simple functions (Definition 18.2.2) as integrands:

$$f \equiv \sum_{i=1}^{n} a_i \mathbb{1}_{]t_{i-1},t_i]} \,. \tag{46.53}$$

If $(W_t)_{t\in\mathbb{R}^+}$ denotes a Brownian motion on a standard probability space (Ω, Σ, P), the Wiener integral of $f : \mathbb{R}^+ \to \mathbb{R}$ with respect to W is then given by

$$\int_0^{+\infty} f(s)\, dW_s := \sum_{i=1}^{n} a_i \big(W_{t_i} - W_{t_{i-1}}\big) \,. \tag{46.54}$$

It is not hard to show, using the definition of Brownian motion, that the stochastic integral defines a partial isometry on the subspace of $L^2(\mathbb{R}^+)$ of simple functions. By Property 25.2.10, one obtains an isometry $L^2(\mathbb{R}^+) \to L^2(\Omega)$. In fact, by generalizing to locally integrable functions, one can define a new stochastic process

$$f \cdot W_t \equiv \int_0^t f\, dW \equiv \int_0^t f_s\, dW_s := \int_0^{+\infty} f \mathbb{1}_{[0,t]}\, dW_s \,. \tag{46.55}$$

Formula 46.4.1. The process $(f \cdot W_t)_{t\in\mathbb{R}^+}$ is a continuous, square-integrable martingale.

Formula 46.4.2 (Covariance).

$$\mathsf{Cov}\left[\int_0^t f(s)\, dW_s, \int_0^{t'} g(s)\, dW_s\right] = \int_0^{\min(s,t)} f(x)g(x)\, dx \tag{46.56}$$

To extend this construction to stochastic integrands, one has to require that the integrand $(\phi_t)_{t\in\mathbb{R}^+}$ is progressive (Definition 46.1.8). Progressive measurability ensures that, by Fubini's theorem (Section 18.4.2), $\int_0^t \phi_s\, ds$ is adapted if and only if $\int_0^t |\phi_s|\, ds$ is a.s. bounded. Let $M^2(\mathbb{R}^+)$ denote the space of square-integrable progressive processes $(\phi_t)_{t\in\mathbb{R}^+}$:

$$\mathsf{E}\left[\int_0^{+\infty} \phi_s^2\, ds\right] < +\infty \,. \tag{46.57}$$

Note that $M^2(\mathbb{R}^+)$ contains random step functions

$$X \equiv \sum_{i=1}^{n} X_i \mathbb{1}_{]t_{i-1},t_i]} \,, \tag{46.58}$$

where the coefficients $X_i \in L^2(\Omega)$. The Wiener–Paley integral of such a function is defined similarly:

$$\int_0^{+\infty} X\, dW_s(\omega) := \sum_{i=1}^{n} X_i(\omega)\big(W_{t_i} - W_{t_{i-1}}\big) \,. \tag{46.59}$$

Extending this partial isometry, gives the **Itô integral** of ϕ with respect to W. The space of progressively measurable stochastic processes satisfying

$$\mathsf{E}\left[\int_0^{+\infty} \phi_s^2 \, ds\right] < +\infty \tag{46.60}$$

is denoted by $L^2(W)$.

Property 46.4.3 (Centrality). Consider a Brownian motion $(W_t)_{t\in\mathbb{R}^+}$ with natural filtration $\mathbb{F}$. For all $\mathbb{F}$-progressive processes $(\phi_t)_{t\in\mathbb{R}^+}$, the following equation holds for all $t \in \mathbb{R}^+$:

$$\mathsf{E}\left[\int_0^t \phi_s \, dW_s\right] = 0\,. \tag{46.61}$$

46.4.2 Stochastic integral

Recall Example 46.1.11 for the definition of the space of elementary predictable process $\mathcal{E}$. The **stochastic integral** of such a process $(\phi_t)_{t\in\mathbb{R}^+}$ with respect to another stochastic process $(M_t)_{t\in\mathbb{R}^+}$, usually a martingale (see further on) is defined as follows:

$$\phi \cdot M_t \equiv \int_0^t \phi \, dM \equiv {}^{9}\int_0^t \phi_s \, dM_s := \sum_{i=1}^n f_i(M_{t\wedge\tau_i} - M_{t\wedge\tau_{i-1}})\,. \tag{46.62}$$

For discrete-time processes, the stochastic integral attains a simple form.

Example 46.4.4 (Martingale transform). Let $(M_n)_{n\in\mathbb{N}}$ be a discrete-time martingale and let $(\phi_n)_{n\in\mathbb{N}}$ be predictable. The (discrete) stochastic integral of ϕ with respect to M is given by:

$$\phi \cdot M_n := \sum_{i=1}^n \phi_i \, \Delta M_i\,. \tag{46.63}$$

For $n = 0$, the convention $\phi \cdot M_0 = 0$ is used.

Since, the elementary predictable processes are dense in the space of all predictable processes, bounded convergence (in probability) allows to extend the stochastic integral to all (bounded) predictable processes.

Definition 46.4.5 (Stochastic integral). An L^0-valued functional

$$(\phi_t)_{t\in\mathbb{R}^+} \mapsto \phi \cdot M_t \tag{46.64}$$

on the space of bounded predictable processes such that

1. it coincides with Eq. (46.62) for elementary predictable processes, and

[9] This abuse of notation will be explained in Section 46.4.4.

2. it satisfies bounded convergence in probability, i.e. if $(\phi_{n,t})_{t\in T, n\in\mathbb{N}} \subset \mathcal{E}$ is a uniformly bounded sequence that converges pointwise to a stochastic process $(\phi_t)_{t\in\mathbb{R}^+}$, then

$$\phi_n \cdot M_t \longrightarrow \phi \cdot M_t \tag{46.65}$$

in probability.

The following property states necessary conditions on the **integrator** $(M_t)_{t\in\mathbb{R}^+}$ for all integrals to exist. This property (uniquely) defines the stochastic integral as in the integration theory of *Daniell*.

Property 46.4.6. If an adapted stochastic process $(M_t)_{t\in\mathbb{R}^+}$ is a well-behaved integrator, i.e. the stochastic integral with respect to $(M_t)_{t\in\mathbb{R}^+}$ exists, then:

1. $(M_t)_{t\in\mathbb{R}^+}$ is right continuous in probability[10], and
2. the set

$$\{\phi \cdot X_t \mid \phi \in \mathcal{E} \wedge |\phi| \leq 1\} \tag{46.66}$$

is bounded in probability for all $t \in T$.

The first item can actually be strengthened, since good integrators always admit a càdlàg modification. As such, all integrators will be assumed to be càdlàg from here on. It can also be shown that these necessary conditions are also sufficient.

The following theorem gives an alternative characterization of the integrators satisfying the above property.

Theorem 46.4.7 (Bichteler–Dellacherie). *The semimartingales are exactly the well-behaved integrators for the stochastic integral.*

Boundedness is, of course, a rather strong property and it would be better to have a more general class of integrands. To this end, bounded convergence should be weakened to dominated convergence (cf. Theorem 18.2.22).

Definition 46.4.8 (Integrability). Let $(M_t)_{t\in\mathbb{R}^+}$ be a semimartingale. The space $L^1(M)$ of M-integrable stochastic processes consists of those predictable processes $(\phi_t)_{t\in\mathbb{R}^+}$ such that if $(\phi_{t,n})_{t\in T, n\in\mathbb{N}}$ is a sequence of bounded predictable processes converging to 0 with $|\phi_{n,t}| \leq |\phi_t|$ for all $t \in T$ and $n \in \mathbb{N}$, then

$$\phi \cdot M_t \longrightarrow 0 \tag{46.67}$$

in probability for all $t \in T$.

[10] For this notion of continuity, replace the convergence in Property 6.2.5 with convergence in probability.

Remark 46.4.9. An equivalent definition uses an approach that is similar to the definition of well-behaved integrators. $L^1(M)$ consists exactly of those predictable stochastic processes $(X_t)_{t\in\mathbb{R}^+}$ for which the set in Eq. (46.66), where $(\phi_t)_{t\in\mathbb{R}^+}$ is replaced by a bounded and X-dominated process, is bounded in probability for all $t \in T$.

Property 46.4.10. For the subclass of integrators belonging to Brownian motions, the space $L^1(M)$ can, equivalently, be described as the space of progressively measurable processes.

Notation 46.4.11 (Differential). A very common notation in stochastic calculus is the differential notation (also common in applied sciences). An equation of the form

$$\mathrm{d}Y = \alpha\,\mathrm{d}X \tag{46.68}$$

means that

$$Y_t = Y_0 + \alpha \cdot X_t \tag{46.69}$$

for some constant $Y_0 \in \mathbb{R}$. For the quadratic (co)variation, a potentially more confusing notation is sometimes used:

$$\begin{aligned} \mathrm{d}X^2 &:= \mathrm{d}[X]\,, \\ \mathrm{d}X\mathrm{d}Y &:= \mathrm{d}[X,Y]\,. \end{aligned} \tag{46.70}$$

Definition 46.4.12 (Itô process). A stochastic process $(X_t)_{t\in\mathbb{R}^+}$ satisfying the following stochastic differential equation (SDE):

$$\mathrm{d}X = \mu\mathrm{d}t + \sigma\mathrm{d}W\,, \tag{46.71}$$

where $(\sigma_t)_{t\in\mathbb{R}^+}, (\mu_t)_{t\in\mathbb{R}^+}$ are adapted. Such processes are actually semimartingales.

Property 46.4.13. The quadratic variation of the Itô process in Eq. (46.71) is given by

$$[X]_t = \int_0^t \mu_s^2\,ds\,. \tag{46.72}$$

Definition 46.4.14 (σ-martingale). An adapted stochastic process $(X_t)_{t\in\mathbb{R}^+}$ such that

$$X_t = \phi \cdot M_t \tag{46.73}$$

for some predictable stochastic process $(\phi_t)_{t\in\mathbb{R}^+}$ and martingale $(M_t)_{t\in\mathbb{R}^+}$.

Property 46.4.15. All (local) martingales are also σ-martingales.

Definition 46.4.16 (Admissible process). Let $(M_t)_{t\in\mathbb{R}^+}$ be a semimartingale. A predictable process $(\phi_t)_{t\in\mathbb{R}^+}$ is said to be (λ-)admissible, where $\lambda \in \mathbb{R}^+$, with respect to $(M_t)_{t\in\mathbb{R}^+}$ if $\phi \cdot M_t \geq -\lambda$ for all $t \in \mathbb{R}^+$.

Definition 46.4.17 (Doléans-Dade exponential). The **stochastic exponential** of a semimartingale $(M_t)_{t\in\mathbb{R}^+}$ is the unique (strong) solution of the stochastic differential equation

$$\mathrm{d}X_t = X_{t^-}\,\mathrm{d}M_t \tag{46.74}$$

with initial value condition $X_0 = 1$.

If $(M_t)_{t\in\mathbb{R}^+}$ is a.s. continuous, the stochastic exponential is given by

$$\mathcal{E}(M)_t = \exp\left(M_t - M_0 - \frac{1}{2}[M]_t\right). \tag{46.75}$$

More generally, when continuity is not assumed, it is given by

$$\mathcal{E}(M)_t = \exp\left(M_t - M_0 - \frac{1}{2}[M]_t^c\right)\prod_{s\le t}(1+\Delta M_s)\exp(-\Delta M_s), \tag{46.76}$$

where $[\cdot]^c$ denotes the continuous part of the quadratic variation.

Property 46.4.18 (Stochastic logarithm). The stochastic exponential (and its left-continuous version) of a semimartingale $(M_t)_{t\in\mathbb{R}^+}$ are strictly positive if and only if the jumps of $(M_t)_{t\in\mathbb{R}^+}$ are strictly greater than -1.

In such cases, the exponential can be inverted to obtain the **stochastic logarithm**. This is given by

$$\mathcal{L}(X)_t = \frac{1}{X_-}\cdot X_t. \tag{46.77}$$

46.4.3 Properties

Note that L^0 is the set of equivalence classes of random variables that are almost surely equal. The consequence is that the stochastic process given by

$$\phi\cdot M_t \equiv \int_0^t \phi_s\,dM_s \tag{46.78}$$

has sample paths that are not necessarily well defined. However, they admit well-behaved modifications.

Property 46.4.19 (Continuity). If $(M_t)_{t\in T}$ is an a.s. continuous semimartingale, then any stochastic integral with respect to M is also a.s. continuous.

Martingales can also be characterized in terms of the centrality of their stochastic integral. (This generalizes Property 46.4.3.)

Property 46.4.20 (Centrality). An integrable, adapted stochastic process $(X_t)_{t\in\mathbb{R}^+}$ is a martingale if and only if

$$\mathsf{E}\left[\int_0^{+\infty} \phi_s \, dX_s\right] = 0 \tag{46.79}$$

for all $(\phi_t)_{t\in\mathbb{R}^+} \in \mathcal{E}$. If this only holds after replacing $=$ by $\geq$, then $(X_t)_{t\in\mathbb{R}^+}$ is a submartingale.

Property 46.4.21 (Stability). If $(M_t)_{t\in\mathbb{R}^+}$ is a martingale and $(\phi_t)_{t\in\mathbb{R}^+}$ a bounded predictable process, then $(\phi \cdot M_t)_{t\in\mathbb{R}^+}$ is also a martingale.

Property 46.4.22 (Change of variables). Let $(M_t)_{t\in\mathbb{R}^+}$ be a semimartingale and consider a stochastic process $(X_t)_{t\in\mathbb{R}^+} \in L^1(M)$. $Z_t := X \cdot M_t$ is an adapted stochastic process admitting a càdlàg modification and, moreover, is a semimartingale such that

$$\int_0^t \phi \, dZ = \int_0^t \phi X \, dM \tag{46.80}$$

for all $t \in T$ and bounded, predictable processes $(\phi_t)_{t\in\mathbb{R}^+}$. Any càdlàg, adapted process satisfying this property will be called 'the' stochastic integral of $(X_t)_{t\in\mathbb{R}^+}$ with respect to $(M_t)_{t\in\mathbb{R}^+}$.

Moreover, a predictable process $(Y_t)_{t\in\mathbb{R}^+}$ is Z-integrable if and only if

$$\int_0^t Y \, dZ = \int_0^t Y X \, dM \,. \tag{46.81}$$

Property 46.4.23 (Optional stopping). Consider a semimartingale $(M_t)_{t\in\mathbb{R}^+}$ and a stopping time τ. If $(X_t)_{t\in\mathbb{R}^+}$ is an M-integrable process, it is also M^τ-integrable and

$$(X \cdot M)^\tau = X \cdot M^\tau \,. \tag{46.82}$$

Property 46.4.24. A predictable process is M-integrable for some semimartingale $(M_t)_{t\in\mathbb{R}^+}$ if and only if it is locally M-integrable. Moreover, a stochastic process is a semimartingale if and only if it is a semimartingale.

This property also implies that all locally bounded predictable processes are integrable with respect to any semimartingale. In fact, for every càdlàg, adapted stochastic process $(X_t)_{t\in\mathbb{R}^+}$, the left-limit process $(X_{t^-})_{t\in T}$ is integrable with respect to any semimartingale.

Theorem 46.4.25 (Dominated convergence). *Let $(\phi_{t,n})_{t\in T, n\in\mathbb{N}}$ be a convergent sequence of predictable processes with limit $(\phi_t)_{t\in\mathbb{R}^+}$ and let $(M_t)_{t\in\mathbb{R}^+}$ be a semimartingale. If the sequence is dominated by an M-integrable stochastic process, then*

$$\int_0^t \phi_n \, dM \xrightarrow{ucp} \int_0^t \phi \, dM \tag{46.83}$$

for all $t \in T$.

Property 46.4.26 (Linearity). The set of semimartingales is a (real) vector space and, moreover, the stochastic integral is linear with respect to this structure.

The following formula should be compared to Formula 18.6.12 for Lebesgue–Stieltjes integrals (an additional relation is given in the next section).

Formula 46.4.27 (Integration by parts). Let $(X_t)_{t\in\mathbb{R}^+}, (Y_t)_{t\in\mathbb{R}^+}$ be semimartingales.

$$X_tY_t = X_0Y_0 + \int_0^t X_{s^-}\, dY_s + \int_0^t Y_{s^-}\, dX_s + [X,Y]_t \tag{46.84}$$

In differential notation, this becomes:

$$\mathsf{d}(XY) = X_\mathsf{d}Y + Y_\mathsf{d}X + \mathsf{d}X\mathsf{d}Y\,. \tag{46.85}$$

Note that this formula could be used as an alternative definition of the quadratic variation process.

Formula 46.4.28 (Jump process). If $(X_t)_{t\in\mathbb{R}^+}$ is a semimartingale and $(P_t)_{t\in\mathbb{R}^+} \in L^1(X)$, then

$$\mathsf{d}Y = P\,\mathsf{d}X \implies \Delta Y = P\Delta X\,. \tag{46.86}$$

Moreover, the continuity (resp. predictability) of $(X_t)_{t\in\mathbb{R}^+}$ implies the continuity (resp. predictability) of $(Y_t)_{t\in\mathbb{R}^+}$. Moreover,

$$\Delta[X,Y] = \Delta X\Delta Y\,. \tag{46.87}$$

Corollary 46.4.29. If $(M_t)_{t\in\mathbb{R}^+}$ is a semimartingale, then

$$\sum_{s\le t} \Delta M_s^2 \le [M]_t < +\infty\,. \tag{46.88}$$

The following inequality should be compared to the Cauchy–Schwarz inequality (Formula 18.3.4).

Formula 46.4.30 (Kunita–Watanabe inequality). Let $(X_t)_{t\in\mathbb{R}^+}, (Y_t)_{t\in\mathbb{R}^+}$ be semimartingales and consider two stochastic processes $(\alpha_t)_{t\in\mathbb{R}^+}, (\beta_t)_{t\in\mathbb{R}^+}$.

$$\int_0^t |\alpha_s\beta_s|\,|d[X,Y]_s| \le \sqrt{\int_0^t \alpha_s^2\, d[X]_s \int_0^t \beta_s^2\, d[Y]_s}\,, \tag{46.89}$$

where $|d[X,Y]_s|$ denotes the total variation measure (Definition 18.7.3) and all integrals are Lebesgue–Stieltjes integrals (by Section 46.4.4).

Corollary 46.4.31 (Brownian motion). Let $(W_t)_{t\in\mathbb{R}^+}, (\overline{W}_t)_{t\in\mathbb{R}^+}$ be standard Brownian motions. The Kunita–Watanabe inequality implies that

$$\int_0^t \alpha_s^2 |d[W,\overline{W}]_s| \le \int_0^t \alpha_s^2\, ds\,. \tag{46.90}$$

By the Radon–Nikodym theorem 18.5.6, there exists a predictable process $(\rho_t)_{t\in\mathbb{R}^+}$, bounded by 1, such that

$$\mathrm{d}W\mathrm{d}\overline{W} = \rho\mathrm{d}t\,. \tag{46.91}$$

$(\rho_t)_{t\in\mathbb{R}^+}$ is called the **instantaneous correlation** of the Brownian motions.

An equivalent way of defining general stochastic integrals would be by mirroring the previous section on Wiener integrals through the following isometry.

Property 46.4.32 (Itô isometry). Consider a local martingale $(M_t)_{t\in\mathbb{R}^+}$ and a predictable process $(\phi_t)_{t\in\mathbb{R}^+}$ such that

$$\mathsf{E}\left[\int_0^t \phi_s^2\, d[M]_s\right] < +\infty\,. \tag{46.92}$$

Then $\phi \in L^1(M)$ and

$$\mathsf{E}\Big[\big(\phi\cdot M_t\big)^2\Big] = \mathsf{E}\left[\int_0^t \phi_s^2\, d[M]_s\right]. \tag{46.93}$$

Property 46.4.33 (Ansel–Stricker). Let $(M_t)_{t\in\mathbb{R}^+}$ be a local martingale and consider $X \in L^1(M)$. If $(X\cdot M_t)_{t\in\mathbb{R}^+}$ is uniformly bounded from below by an integrable random variable, then $(X\cdot M_t)_{t\in\mathbb{R}^+}$ is a local martingale and even a supermartingale.

Theorem 46.4.34 (Girsanov). *Consider a filtered probability space $(\Omega,\Sigma,\mathbb{F},P)$ and an a.s. continuous semimartingale $(M_t)_{t\in\mathbb{R}^+}$. Moreover, let $(W_t)_{t\in\mathbb{R}^+}$ be a Brownian motion. If*

$$M_t := \mathcal{E}(M)_t = \exp\left(\int_0^t M_s\, dW_s - \frac{1}{2}\int_0^t [M]_s\, ds\right) \tag{46.94}$$

is an $\mathbb{F}$-martingale, then

$$\widetilde{W}_t := W_t - \int_0^t M_s\, ds \tag{46.95}$$

is a Brownian motion with respect to the transformed measure $M_t\, dP$.

Example 46.4.35 (Itô process). A specific case, which is often taken to be Girsanov's theorem, is given by considering the Itô process $\mathrm{d}X_t = \mu_t\mathrm{d}t + \sigma_t\mathrm{d}W_t$, where $(W_t)_{t\in\mathbb{R}^+}$ is a standard Brownian motion with respect to some base measure P. If the ratio $\eta_t := \mu_t/\sigma_t$ satisfies the **Novikov condition**

$$\mathsf{E}\left[\exp\left(\frac{1}{2}\int_0^t |\eta_s|^2\, ds\right)\right] < +\infty \tag{46.96}$$

for all $t \in \mathbb{R}^+$, then $M_t := \mathcal{E}\left(\int_0^t \eta_s \, dW_s\right)$ is a martingale[11] and

$$\widetilde{W}_t := W_t - \int_0^t \eta_s \, ds \tag{46.97}$$

is a standard Brownian motion with respect to the transformed measure $M_t \mathrm{d}P$. A generalization to higher dimensions is straightforward.

A more general statement is the following.

Theorem 46.4.36. *Let P, Q be (locally) equivalent measures on a filtered measurable space such that the associated Radon–Nikodym derivatives $(L_t)_{t\in\mathbb{R}^+}$ are continuous. If $(M_t)_{t\in\mathbb{R}^+}$ is an a.s. continuous P-local martingale, the process*

$$\mathrm{d}\widetilde{M}_t := \mathrm{d}M_t - \frac{1}{L_t}\mathrm{d}[M, L]_t \tag{46.98}$$

is an a.s. continuous Q-local martingale. Moreover, the process $(L_t)_{t\in\mathbb{R}^+}$ can be written as a Doléans-Dade exponential of a local martingale $(\xi_t)_{t\in\mathbb{R}^+}$ and $M_t - [M, \xi]_t$ is also a Q-local martingale.

Corollary 46.4.37. The transformed process $(\widetilde{M}_t)_{t\in\mathbb{R}^+}$ will be a Q-local martingale if and only if M and L are orthogonal.

46.4.4 Lebesgue–Stieltjes integration

Property 46.4.38. Let $(X_t)_{t\in\mathbb{R}^+}$ be a continuous stochastic process. If for each $t > 0$, the p-variation[12] converges in probability to a random variable taking values in $\mathbb{R}^+$ almost surely, then the q-variations for all $q < p$ vanish in probability, and the q-variations for all $q > p$ diverge in probability (on the event where the quadratic variation is nonzero).

By Definition 46.3.11, the continuous, square-integrable martingales have a well-defined quadratic variation process. Then, by the property above, this means that their first variation diverges almost surely and that their higher variations are almost surely zero. This shows that it is impossible to obtain a well-defined integration theory with (square-integrable) martingales as integrators. However, in general, instead of looking at the martingales themselves, one can look at their quadratic variation process. These are continuous and increasing and, hence, are equal to their own first variation process. This allows one to define the Lebesgue–Stieltjes integral

$$I(\omega) := \int_0^t Y_s(\omega) \, d[X]_s(\omega) \tag{46.99}$$

[11] This is ensured by the Novikov condition.

[12] The stochastic process obtained by replacing the square in Eq. (46.48) by the power $p > 0$.

along sample paths. The following property shows that, in good cases, the stochastic integral and pathwise Lebesgue–Stieltjes integral coincide.

Property 46.4.39. Let $(M_t)_{t\in\mathbb{R}^+}$ be a semimartingale and a predictable process $(\phi_t)_{t\in\mathbb{R}^+}$. If

$$\int_0^t |\phi_s|\,|dM_s| < +\infty \tag{46.100}$$

almost surely for all $t \in T$, where $|dM_s|$ denotes the total variation measure (Definition 18.7.3), then $(\phi_t)_{t\in\mathbb{R}^+} \in L^1(M)$ and

$$\int_0^t \phi_s(\omega)\,dM_s(\omega) = \phi \cdot M_t(\omega) \tag{46.101}$$

for all $t \in T$ and $\omega \in \Omega$.

Formula 46.4.40. Let $(M_t)_{t\in\mathbb{R}^+}$ be a semimartingale and let $(X_t)_{t\in\mathbb{R}^+}$ be an FX process.

$$[M, X]_t = \int_0^t \Delta M_s\,dX_s = \sum_{s\le t} \Delta M_s \Delta X_s\,, \tag{46.102}$$

where the integral with respect to X is a Lebesgue–Stieltjes integral. In particular, if either process is continuous, their quadratic covariation vanishes. This property is the reason why quadratic variations are a rare sight in ordinary calculus.

Corollary 46.4.41. Let $(M_t)_{t\in\mathbb{R}^+}$ be a semimartingale and let $(X_t)_{t\in\mathbb{R}^+}$ be an FX process.

$$M_t X_t = M_0 X_0 + \int_0^t M_s\,dX_s + \int_0^t X_{s^-}\,dM_s\,, \tag{46.103}$$

where the integral with respect to X is, again, a Lebesgue–Stieltjes integral.

@@ COMPLETE SECTION @@

46.4.5 Representation theorems

Theorem 46.4.42 (Dudley). *Let $(W_t)_{t\in\mathbb{R}^+}$ be a standard Brownian motion and consider its natural filtration $\mathbb{F}$. For every $\mathbb{F}_\infty$-measurable random variable ξ, there exists an $\mathbb{F}$-progressive process $(\phi_t)_{t\in\mathbb{R}^+} \in L^2(W)$ such that*

$$\xi = \int_0^{+\infty} \phi_s\,dW_s\,. \tag{46.104}$$

Theorem 46.4.43 (Martingale representation theorem). *Let $(W_t)_{t\in\mathbb{R}^+}$ be a standard Brownian motion with its natural filtration $\mathbb{F}$. Every $\mathbb{F}$-local martingale $(X_t)_{t\in\mathbb{R}^+}$ can be written as*

$$X_t = X_0 + \phi \cdot W_t \tag{46.105}$$

for a unique predictable $(\phi_t)_{t\in\mathbb{R}^+} \in L^1_{\text{loc}}(W)$.

This theorem can generalized in the setting of Giransov's theorem 46.4.34 as follows. If $(W_t)_{t\in\mathbb{R}^+}$ is a standard Brownian motion with respect to a measure P and Q is (locally) equivalent to P, then every Q-local martingale $(M_t)_{t\in\mathbb{R}^+}$ can be written as

$$M_t = M_0 + \phi \cdot \widetilde{W}_t\,, \tag{46.106}$$

where $(\phi_t)_{t\in\mathbb{R}^+}$ is $\mathbb{F}^W$-predictable and $(\widetilde{W}_t)_{t\in\mathbb{R}^+}$ is the martingale part of W when regarded as a Q-semimartingale.

The following theorem gives an orthogonal decomposition of local martingales.

Theorem 46.4.44 (Kunita–Watanabe). *Let $(M_t)_{t\in\mathbb{R}^+}$ be a continuous $\mathbb{F}$-local martingale with respect to some filtration $\mathbb{F}$. Every other continuous $\mathbb{F}$-local martingale $(X_t)_{t\in\mathbb{R}^+}$ that vanishes at $t = 0$ can be written as*

$$X_t = N_t + \phi \cdot M_t \tag{46.107}$$

where $(\phi_t)_{t\in\mathbb{R}^+}$ is predictable and $(N_t)_{t\in\mathbb{R}^+}$ is an $\mathbb{F}$-local martingale orthogonal to M.

Comparing the previous two statements, the question becomes when the orthogonal part $(N_t)_{t\in\mathbb{R}^+}$ is a constant.

Theorem 46.4.45 (Predictable representation theorem). *Let $(M_t)_{t\in\mathbb{R}^+}$ be a continuous $\mathbb{F}$-local martingale with respect to a filtered probability space $(\Omega, \Sigma, \mathbb{F}, P)$ and consider its natural filtration $\mathbb{F}^M$. Any $\mathbb{F}^M$-local martingale $(X_t)_{t\in\mathbb{R}^+}$ can be decomposed as*

$$X_t = X_0 + \phi \cdot M_t \tag{46.108}$$

for some constant $X_0 \in \mathbb{R}$ and $\mathbb{F}^M$-predictable process $(\phi_t)_{t\in\mathbb{R}^+}$ if and only if P is extremal (Definition 16.7.2) in the set of probability measures for which M is a continuous, local martingale.

Remark 46.4.46 (Generalization). More generally, a collection of $(\mathbb{F}, P)$-local martingales (or even semimartingales) $\left\{(M^1_t)_{t\in\mathbb{R}^+}, \dots (M^d_t)_{t\in\mathbb{R}^+}\right\}$, for a general filtered probability space $(\Omega, \Sigma, \mathbb{F}, P)$, is said to satisfy the (strong) predictable representation property (PRP) if every $(\mathbb{F}, P)$-local[13] martingale $(X_t)_{t\in\mathbb{R}^+}$ can be written as

$$X_t = X_0 + \boldsymbol{\phi} \cdot \mathbf{M}_t = X_0 + \sum_{i=1}^{d} \phi^i \cdot M^i_t \tag{46.109}$$

for some $\mathbb{F}_0$-measurable function X_0 and $\mathbb{F}$-predictable process $(\boldsymbol{\phi}_t)_{t\in\mathbb{R}^+}$ satisfying

$$\int_0^t \phi^{i,2}_s \, d[M]_s < +\infty \tag{46.110}$$

for all $i \leq d$.

[13] In fact, one can relax this to the existence of measure Q (locally) equivalent to P.

Remark 46.4.47 (Brownian motion). The martingale representation theorem in terms of Brownian motion can be recovered by noting that the Wiener measure is an extremal martingale law on the space $C(\mathbb{R}^+, \mathbb{R})$ by the Lévy characterization of Brownian motion (Definition 46.3.23).

Whereas stability under change of distribution was already treated above, the change of filtrations (refinement, in particular) will be treated here.

Definition 46.4.48 (Immersion property). A refinement of filtered probability spaces $(\mathbb{F}, P) \hookrightarrow (\mathbb{G}, P)$ is said to satisfy the immersion property if every $(\mathbb{F}, P)$-martingale is also a $(\mathbb{G}, P)$-martingale. A corollary of this property is that also all $(\mathbb{F}, P)$-semimartingales are $(\mathbb{G}, P)$-semimartingales.

Property 46.4.49. If $(\mathbf{M}_t)_{t\in\mathbb{R}^+}$ satisfies the PRP with respect to $(\mathbb{F}, P)$, then $(\mathbb{F}, Q) \hookrightarrow (\mathbb{G}, Q)$ satisfies the immersion property for all probability measures Q (locally) equivalent to P, for which M is a $(\mathbb{G}, Q)$-local martingale.

A converse, negative result is the following.

Property 46.4.50. Let $(\mathbf{M}_t)_{t\in\mathbb{R}^+}$ satisfy the PRP with respect to $(\mathbb{F}, P)$. If the refinement $(\mathbb{F}, P) \hookrightarrow (\mathbb{G}, P)$ satisfies $\mathbb{G}_0 = \{\emptyset, \Omega\}$, then $\mathbf{M}$ does not satisfy the PRP with respect to $(\mathbb{G}, P)$.

46.5 Stochastic differential equations

Formula 46.5.1 (Itô lemma). Consider an Itô process $(X_t)_{t\in\mathbb{R}^+}$:

$$\mathrm{d}X = \mu \mathrm{d}t + \sigma \mathrm{d}W\,. \tag{46.111}$$

For $f \in C^2(\mathbb{R}^2, \mathbb{R})$, one has

$$\mathrm{d}f(t, X) = \left(\frac{\partial f}{\partial t} + \mu\frac{\partial f}{\partial x} + \frac{\partial^2 f}{\partial x^2}\right)\mathrm{d}t + \sigma\frac{\partial f}{\partial x}\mathrm{d}W\,. \tag{46.112}$$

This also implies that $f(t, X)$ is again an Itô process.

More generally, consider a d-dimensional continuous semimartingale $(X_t)_{t\in\mathbb{R}^+}$ as a function $g \in C^2(\mathbb{R}^d, \mathbb{R})$.

$$\mathrm{d}f(X) = \frac{\partial f}{\partial x^i}(X)\mathrm{d}X^i + \frac{1}{2}\frac{\partial^2 f}{\partial x^i \partial x^j}g(X)\mathrm{d}[X^i, X^j]\,. \tag{46.113}$$

If X is not continuous, a correction has to be applied:

$$\mathrm{d}f(X) = \frac{\partial f}{\partial x^i}(X_-)\,\mathrm{d}X^i + \frac{1}{2}\frac{\partial^2 f}{\partial x^i \partial x^j}g(X)\,\mathrm{d}[X^i, X^j]^c + \left(\Delta f(X) - \frac{\partial f}{\partial x^i}(X_-)\Delta X^i\right). \tag{46.114}$$

46.6 Markov processes

Definition 46.6.1 (Markov process). A Markov process (or chain) is a stochastic process $(X_t)_{t\in T}$ adapted to a filtration $(\mathbb{F}_t)_{t\in T}$ such that

$$\mathsf{Prob}(X_t \mid \mathbb{F}_s) = \mathsf{Prob}(X_t \mid X_s) \tag{46.115}$$

for all $t, s \in T$. For discrete-time processes, the first-order Markov chains are the most common. These satisfy

$$\mathsf{Prob}(X_t \mid X_{t-1}, \dots, X_{t-r}) = \mathsf{Prob}(X_t \mid X_{t-1}) \tag{46.116}$$

for all $t, r \in \mathbb{N}$.

@@ COMPLETE @@

Chapter 47

Information Geometry ♣

The main reference for this chapter is Amari (2016). For more information on differential geometry, see Chapter 31 and onwards.

Chapter Contents

47.1 Statistical manifolds **894**
47.1.1 Divergences 896
47.1.2 Exponential families 899
47.1.3 Mixtures 900
47.1.4 Compatible divergences 901
47.1.5 Flat structures 903
47.2 Projections **903**
47.2.1 Functional descent 904

47.1 Statistical manifolds

In this section, an important subclass of Riemannian manifolds that admit two related flat connections will be introduced. These manifolds will form the geometric backbone of many statistical concepts and methods.

Definition 47.1.1 (Conjugate connections). Let (M, g) be a Riemannian manifold and consider an affine connection ∇ on M. The conjugate (or dual) connection $\widetilde{\nabla}$ is uniquely defined by the following equation (akin to the Leibniz property):

$$X\big(g(Y,Z)\big) = g(\nabla_X Y, Z) + g(Y, \widetilde{\nabla}_X Z)\,, \tag{47.1}$$

where $X, Y, Z \in TM$. Moreover, this construction is involutive:

$$\widetilde{\widetilde{\nabla}} = \nabla\,. \tag{47.2}$$

Property 47.1.2. Consider a pair of conjugate connections $\nabla, \widetilde{\nabla}$ on a Riemannian manifold (M, g) and denote their parallel transport maps by $\mathcal{P}$ and $\mathcal{P}'$, respectively. Although the metric is, in general, not preserved under either $\mathcal{P}$ or $\mathcal{P}'$, it is preserved under conjugate (or dual) transport:

$$g(v, w) = g\big(\mathcal{P}_\gamma v, \widetilde{\mathcal{P}}_\gamma w\big) \tag{47.3}$$

for every smooth path γ.

Property 47.1.3. Consider two conjugate connections $\nabla, \widetilde{\nabla}$ on a Riemannian manifold (M, g). The connection

$$\overline{\nabla} := \frac{\nabla + \widetilde{\nabla}}{2} \tag{47.4}$$

is metric(-preserving), i.e. $\overline{\nabla} g = 0$. Furthermore, if both ∇ and $\widetilde{\nabla}$ are torsion-free, then $\overline{\nabla}$ necessarily coincides with the Levi-Civita connection of g by Theorem 36.1.19.

The above properties lead to the following definition.

Definition 47.1.4 (Statistical manifold). A Riemannian manifold (M, g) equipped with an affine connection that satisfies the **Codazzi condition**

$$\nabla_X g(Y, Z) = \nabla_Y g(X, Z)\,, \tag{47.5}$$

i.e. ∇g is totally symmetric. The Codazzi condition implies vanishing torsion and vice versa. The rank-3 tensor $T := \nabla g$ is sometimes called the **cubic tensor** or **Amari–Chentsov** tensor. In local coordinates, the cubic tensor gives the difference between the Christoffel symbols of ∇ and $\widetilde{\nabla}$:

$$T_{ijk} = \tilde{\Gamma}_{ijk} - \Gamma_{ijk}\,. \tag{47.6}$$

In the case where ∇ has nonvanishing torsion, one can generalize this definition by also relaxing the Codazzi equation:

$$\nabla_X g(Y, Z) - \nabla_Y g(X, Z) = -g\big(T^\nabla(X, Y), Z\big)\,. \tag{47.7}$$

If this equation is satisfied for all X, Y and $Z \in TM$, the dual connection is torsion-free and the tuple (M, g, ∇) is called a **statistical manifold admitting torsion**.[1]

Remark 47.1.5. One can show that the existence of a torsion-free connection is sufficient to a (pseudo-)Riemannian manifold admitting a totally symmetric rank-3 tensor, i.e. to turning it into a statistical manifold admitting torsion.

Definition 47.1.6 (Dually flat manifold). Consider a statistical manifold (M, g, ∇). If ∇ is flat, its conjugate $\widetilde{\nabla}$ is also flat and the tuple $(M, g, \nabla, \widetilde{\nabla})$ is called a dually flat manifold.

[1] This situation arises in the study of quantum systems and density operators.

Because the affine connections are flat, they endow the manifold with an *affine structure*, i.e. there exist coordinate charts such that the coordinate-induced vector fields satisfy

$$\nabla_{\partial_i}\partial_j = 0 \tag{47.8}$$

for all $i, j \leq n$ and such that the transition functions are affine transformations. It can be shown that the conjugate connection induces a similar $\widetilde{\nabla}$-affine coordinate chart such that the coordinate-induced vector fields satisfy the following orthonormality condition:

$$g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y_j}\right) = \delta_i^j . \tag{47.9}$$

This coordinate system is called the **dual (coordinate) system**.

47.1.1 Divergences

Definition 47.1.7 (Divergence). Let M be a set. A smooth function $D(\cdot\|\cdot) : M \times M \to \mathbb{R}$ with the following properties is called a divergence (measure) on M:

1. **Positivity**: $D(p \,\|\, q) \geq 0$ for all $p, q \in M$, and
2. **Nondegeneracy**: $D(p \,\|\, q) = 0$ if and only if $p = q$.

The **dual divergence** D^* is defined by

$$D^*(p \,\|\, q) := D(q \,\|\, p) \tag{47.10}$$

for all $p, q \in M$.

Property 47.1.8 (Induced metric). An interesting feature of these functions is that one can use their Hessians (with respect to either of the two arguments) to construct a Riemannian metric if M is a smooth manifold:

$$g_{ij}(\theta) := \frac{\partial^2 D}{\partial p^i \partial p^j}(p \,\|\, q)\bigg|_{p=q=\theta} = \frac{\partial^2 D}{\partial q^i \partial q^j}(p \,\|\, q)\bigg|_{p=q=\theta} = -\frac{\partial^2 D}{\partial p^i \partial q^j}(p \,\|\, q)\bigg|_{p=q=\theta} . \tag{47.11}$$

Example 47.1.9 (f-divergences and α-divergences). Let $f : \mathbb{R} \to \mathbb{R}$ be a smooth, convex function such that $f(1) = 0$ and let p, q be two probability distributions such that p is absolutely continuous with respect to q. The f-divergence is defined as follows:

$$D_f(p \,\|\, q) := \int_\Omega f\left(\frac{\mathrm{d}p}{\mathrm{d}q}\right) dq . \tag{47.12}$$

In the case where both p and q are absolutely continuous with respect to some given measure μ (with Radon–Nikodym derivatives g, h), one can rewrite the above formula as

$$D_f(p \,\|\, q) = \int_\Omega h(x) f\left(\frac{g(x)}{h(x)}\right) d\mu(x) . \tag{47.13}$$

It is not hard to see that f-divergences remain invariant under affine transformations of the form

$$f(x) \longrightarrow f(x) + c(x-1)\,, \tag{47.14}$$

where the shift $x-1$ is necessary to preserve the condition $f(1)=0$. This implies that one can always choose $f'(1)=0$ without loss of generality. Moreover, one can also easily see that these divergences transform linearly under rescaling and, hence, one can also always choose $f''(1)=1$. f-divergences with these properties are said to be **standard**.

A particular class of f-divergences are the α-divergences (in the sense of *Csiszár*[2]) where

$$f_\alpha(x) := \frac{1-x^\alpha}{\alpha(1-\alpha)}\,. \tag{47.15}$$

Note that some authors replace $(1-x^a)$ by $(x-x^a)$ since this does not make any difference when calculating the divergence for normalized distributions. For the cases $\alpha=0,1$, one can use a workaround. For $\alpha=0$, one can take the limit of the above definition to obtain $f_0(x) = -\ln x$. This results in $D_0(p\,\|\,q) = D_{\mathrm{KL}}(q\,\|\,p)$. For $\alpha=1$, one can look at the general expression of D_α and notice that it is invariant under the simultaneous exchanges $(\alpha \leftrightarrow 1-\alpha)$ and $(p \leftrightarrow q)$. Using this trick, one can see that $D_1(p\,\|\,q) = D_{\mathrm{KL}}(p\,\|\,q)$.

Definition 47.1.10 (Bregman divergence). Let $F:\mathbb{R}^n \to \mathbb{R}$ be a convex function. Because the function is convex, at every point $q\in\mathbb{R}^n$ the tangent plane to the graph of F is a supporting hyperplane, i.e. it lies underneath the graph everywhere and it touches the graph only at q. Using this hyperplane, one can define the Bregman divergence as follows:

$$D_F(p\,\|\,q) := F(p) - F(q) - \vec{\nabla}F(q)\cdot(p-q)\,, \tag{47.16}$$

where the gradient is denoted by $\vec{\nabla}$ to avoid confusion with any occurrences of the ∇-symbol for affine connections. This function gives the difference in 'height' between the function value at p and the position of the tangent plane (defined by q) at p. Because the gradient of a convex function is monotonic, this difference will always increase the further the points are spread apart. (For convex functions, this will, in general, only be nondecreasing. However, in the remainder of this chapter strict convexity will almost always be assumed.)

Example 47.1.11 (Kullback–Leibler divergence). The Kullback–Leibler divergence (Definition 45.5.3) can be obtained as the Bregman divergence associated to the (negative) Shannon entropy $F(\rho) := \sum_{i=1}^n \rho_i \ln \rho_i$. It is also equal to the f-divergence associated to the choice $f = x\ln(x)$.

[2] *Tsallis* and *Rényi* introduced different divergences/entropies with the same name.

A Bregman divergence can also be used to endow the underlying manifold with further structure. By restricting to strictly convex functions, i.e. requiring that the Hessian is positive definite, one can perform a Legendre transformation (Definition 16.7.7) to obtain a new function:

$$\tilde{F}(x^*) := x^* \cdot y - F(y)\,, \tag{47.17}$$

where $x^* = \vec{\nabla} F(y)$.[3] It can be shown that $\tilde{F}$ is again (strictly) convex and, hence, also defines a Bregman divergence. This second Bregman divergence coincides with the dual divergence D_F^*:

$$D_F(p \,\|\, q) = D_{\tilde{F}}(q^* \,\|\, p^*)\,. \tag{47.18}$$

Using this relation one can also rewrite the original expression for the Bregman divergence:

$$D_F(p \,\|\, q) = F(p) + \tilde{F}(q) - x^i(p) y_i(q)\,. \tag{47.19}$$

Now, the two convex functions $F, \tilde{F}$ define two coordinate systems that are related as follows:

$$y = \vec{\nabla} F \qquad \text{and} \qquad x = \vec{\nabla} \tilde{F}(y)\,. \tag{47.20}$$

However, convexity of functions is not preserved under arbitrary coordinate transformations and, hence, one should restrict the class of allowed coordinate transformations. To preserve convexity, only affine transformations are allowed. In affine coordinates, one can express any geodesic, i.e. any path γ such that $\nabla_{\dot{\gamma}} \dot{\gamma} = 0$, as a straight line:

$$\gamma_{q \to p}(t) = t x(p) + (1 - t) x(q)\,. \tag{47.21}$$

Geodesics for the conjugate connection are called **dual geodesics**. It is important to note that the Legendre transform which maps the primary coordinates to the dual coordinates is, in general, not an affine transformation and, hence, does not preserve the dual structure. Moreover, it can be shown that neither of the parallel transport maps, although completely trivial due to the affine structure, are metric-preserving. However, parallel transporting one vector by ∇ and the other by $\tilde{\nabla}$ does preserve the metric. The metric structures induced by the Hessians are also intertwined:

$$g_{ij} = \frac{\partial^2 F}{\partial x^i \partial x^j} \qquad \text{and} \qquad \tilde{g}^{ij} = \frac{\partial^2 \tilde{F}}{\partial y_i \partial y_j} \tag{47.22}$$

are mutual inverses. It can be concluded that a Bregman divergence endows a set with the structure of a dually flat manifold (Definition 47.1.6).

[3] For general convex functions, this relation is not necessarily invertible.

Example 47.1.12 (Euclidean distance). On $\mathbb{R}^n$, the most common choice of divergence measure is the Euclidean distance:

$$D_{\text{eucl}}(p \,\|\, q) := \frac{1}{2}\|p - q\| \,. \tag{47.23}$$

It is not hard to show that this function is, in fact, selfdual with respect to Legendre transformations. This also implies that the primary and dual structures on $\mathbb{R}^n$ (with respect to the Euclidean distance) coincide. The associated connections are equal to the (trivial) Levi-Civita connection.

Property 47.1.13 (Bregman divergence). The dually flat structure on a dually flat manifold $(M, g, \nabla, \widetilde{\nabla})$ enables one to define two affine coordinate systems through two functions ψ, ϕ (called **potentials**). Because the connection ∇ is torsion-free and the metric is symmetric by definition, a function ψ can (locally) be found such that

$$g_{ij} = \partial_i \partial_j \psi \,. \tag{47.24}$$

The positive-definiteness of g further implies that ψ is convex. This implies that ψ can be used to define a Bregman divergence. The induced dually flat structure is exactly $(M, g, \nabla, \widetilde{\nabla})$.

47.1.2 Exponential families

The primary and dual affine geodesics are often given the names *e-geodesic* and *m-geodesic* in the literature. In this and the following section, this terminology is explained.

Definition 47.1.14 (Exponential family). Let $X : \Omega \to \mathbb{R}^k$ be a random variable. For every integer $n \in \mathbb{N}$, every collection of smooth functions $\{h_i : \mathbb{R}^k \to \mathbb{R}\}_{1 \le i \le n}$ and any smooth function $\lambda : \mathbb{R}^k \to \mathbb{R}$, one can define the following family of distributions indexed by some parameter $\theta \in \mathbb{R}^n$:

$$p(X; \theta) := \exp\big(h_i(X)\theta^i + \lambda(X) - \psi(\theta)\big) \,, \tag{47.25}$$

where the function $\psi(\theta)$ is introduced as a normalization function:

$$\psi(\theta) := \ln \int \exp\big(h_i(X)\theta^i\big) e^{\lambda(X)} \, d\mu(X). \tag{47.26}$$

The function λ can be removed through a redefinition of the measure:

$$d\mu(X) \longrightarrow d\nu(X) := \exp^{\lambda(X)} \, d\mu(X) \,. \tag{47.27}$$

Remark 47.1.15. The function ψ is actually the cumulant-generating function (or free energy in physics terminology) of the sufficient statistics $h_i(X)$.

Such a family of exponential distributions forms a manifold with affine coordinates θ^i (these are also called the **natural parameters**). The dual coordinates $\nabla\psi(\theta)$ are the expectation values $\mathsf{E}[X]$ and the associated dual convex function ϕ is the Shannon entropy. Accordingly, the Bregman divergence associated to ψ is given by the dual Kullback–Leibler divergence:

$$D_\psi(\theta \,\|\, \theta') = D_{\mathrm{KL}}(\theta' \,\|\, \theta)\,. \tag{47.28}$$

The metric induced by this structure is the *Fisher information* (see also Definition 48.4.17):

$$g_{ij} = \mathcal{J}_{ij}[X;\theta] := \mathsf{E}\left[\left(\frac{\partial}{\partial\theta^i}\ln p(X;\theta)\right)\left(\frac{\partial}{\partial\theta^j}\ln p(X;\theta)\right)\right]. \tag{47.29}$$

Now, consider an affine combination of natural parameters, i.e. an affine geodesic on the manifold of an exponential family:

$$\theta(t) := t\theta_2 + (1-t)\theta_1\,. \tag{47.30}$$

The probability distributions along this path can themselves be interpreted as constituting an exponential family with natural parameter t and, therefore, one calls the primary geodesic $\theta(t)$ an **e-geodesic** ('e' for exponential).

47.1.3 Mixtures

Another important class of probability distributions is given by the mixture families.

Definition 47.1.16 (Mixture). Consider a collection $\{p_i(X)\}_{i\leq n}$ of probability distributions. For every point $\eta \equiv (\eta_0, \dots, \eta_n)$ in the probability simplex Δ^n, one can define the distribution

$$p(X;\eta) := \sum_{i=0}^{n} \eta_i p_i(X)\,. \tag{47.31}$$

This mixture family forms a manifold with affine coordinates $\{\eta_i\}_{1\leq i\leq n}$ (note that η_0 can be calculated from the other weights and is, therefore, not an independent coordinate).

The (negative) Shannon entropy of a mixture defines a convex function φ and, as noted before, it induces the Kullback–Leibler divergence:

$$D_\varphi(\eta \,\|\, \eta') = D_{\mathrm{KL}}(\eta \,\|\, \eta')\,. \tag{47.32}$$

Example 47.1.17 (Discrete distribution). An interesting example of mixtures is given by the class of discrete (or categorical) distributions:

$$p(X;\eta) = \sum_{i=0}^{n} \eta_i \delta_i(X)\,, \tag{47.33}$$

where $\eta \in \Delta^n$. At the same time, these models can be considered as distributions in an exponential family with affine coordinates $\theta^i := \ln {}^{p_i}/_{p_0}$. For these models, the primary coordinates $\overline{\theta}$, dual to η, coincide with the natural parameters θ.

Consider two points with dual coordinates η_1, η_2. The dual geodesic connecting these points is of the form

$$\eta(t) = t\eta_2 + (1 - y)\eta_1 \,. \tag{47.34}$$

In the case of discrete distributions, where the dual coordinates are given by elements of the probability simplex Δ^n, one can see that such a geodesic induces a linear interpolation of distributions and, accordingly, defines a mixture family. For this reason, one generally calls a dual geodesic an **m-geodesic** ('m' for mixture).[4]

Remark 47.1.18. The reader should be aware of an important source of confusion. The above sections would point to the following naming convention:

e-geodesic	$\leftrightarrow$	primary geodesic, and
m-geodesic	$\leftrightarrow$	dual geodesic.

However, because the Kullback–Leibler divergence is the most widely used divergence measure, it was possible that the above convention got reversed in the bulk of the literature since the KL-divergence is the dual divergence in Eq. (47.28). This reversal also leads one to interchange 'primary' and 'dual' in most statements such as the Pythagorean and projections theorems.

To prevent confusion, it is important to pay attention to which divergence is used. In this text, a distinction has been made (as much as possible) between the e/m-convention and the primary/dual convention. The second convention is the main choice here since this one is uniquely determined once the divergence is known.

47.1.4 Compatible divergences

The question to be answered in this section is the following one: "Given a dually flat manifold, which divergences are compatible with this structure?". In the previous sections, it was shown that exponential families and mixture families naturally give rise to the Kullback–Leibler divergence. However, not all dually flat manifolds are induced by such families.

A basic requirement, as is generally the case with geometric structures, is the requirement that the divergences are invariant under coordinate transformations. To this end, one needs the *monotonicity criterion of Chentsov*. This criterion states that no transformation should increase the divergence between two points (this corresponds to the idea that transformations can never increase the amount of information). Moreover, there exists a class of (noninvertible) transformations that leave the divergence invariant.

Definition 47.1.19 (Sufficient statistic). Consider a random variable X following the distribution $p(X;\theta)$. A transformation $\xi(X)$ is said to be sufficient (with respect to θ)

[4] For arbitrary families, the dual geodesic does not necessarily induce a mixture of distributions.

if the distribution of X conditioned on $\xi(X)$ is independent of θ. The **Fisher–Neyman factorization theorem**[5] states that this is equivalent to the existence of the following decomposition

$$p(X;\theta) = f(X)g_\theta(\xi(X)) \tag{47.35}$$

for some nonnegative functions f, g_θ.

The invariance criterion states that these transformations are the only transformations that leave the divergence invariant.

Axiom 47.1 (Invariance criterion). Consider a dually flat manifold M. Compatible divergences should satisfy the following inequality for all transformations $\overline{x} := \xi(x)$:

$$\overline{D}(\theta \,\|\, \theta') \leq D(\theta \,\|\, \theta'). \tag{47.36}$$

The equality holds if and only if the transformed variable $\overline{x}$ is a sufficient statistic.

Example 47.1.20 (f-divergences). An important class of invariant divergences on the manifold Δ^n is given by the f-divergences introduced in the beginning of this chapter. These also have the additional property that they are **decomposable**

$$D_f(p \,\|\, q) = \sum_{i=0}^{n} \Delta(p_i, q_i) \tag{47.37}$$

for some nonnegative function Δ.

The following result gives a partial characterization of invariant divergences on the manifold of discrete distributions Δ^n.

Property 47.1.21. A divergence D on Δ^n $(n > 1)$ is invariant and decomposable if and only if it it an f-divergence. If the induced geometric structure is required to be flat, then, necessarily, $D = D_{\mathrm{KL}}$. When extended to $\mathbb{R}^n_+$ (the discrete positive measures), the collection of all α-divergences is recovered.

Corollary 47.1.22. Because every Bregman divergence is flat, one can see that D_{KL} is the only Bregman divergence that is also an f-divergence.

One can also go a step further and ask which metrics can arise on such invariant structures. The answer is quite simple (at least for discrete distributions).

Theorem 47.1.23 (Chentsov). *Up to scaling, the only invariant metric that exists on Δ^n is the Fisher information metric. Extending this result to the manifold $\mathbb{R}^n_+$ of discrete positive measures, the only invariant metric on $\mathbb{R}^n_+$ is the Euclidean metric.*

Remark. Extensions to other families/manifolds of distributions can be found in the literature. However, most of these theorems have to make additional assumptions.

[5] Sometimes also called the **Halmos–Savage factorization theorem**.

47.1.5 Flat structures

In this paragraph, the flat structures on $\mathbb{R}^n_+$ are considered. By one of the invariance results above, these are exactly the α-divergences. Following *Amari*, the transformation $\alpha = 2q - 1$ is performed (this maps the *Csiszár divergences* to the *Tsallis divergences*). Given a discrete positive measure with coefficients m_i, the affine coordinates are defined as follows:

$$\theta^i \equiv h_\alpha(m_i) := m_i^{1-\alpha/2}. \tag{47.38}$$

It is not hard to see that the inverse function $\theta^i \mapsto m_i$ is convex (for $|\alpha| \leq 1$) and, hence, one can define a potential as follows:

$$\psi_\alpha(\theta) := \frac{1-\alpha}{2} \sum_{i=0}^{n} m_i = \frac{1-\alpha}{2} \sum_{i=0}^{n} \left(\theta^i\right)^{2/1-\alpha}. \tag{47.39}$$

The dual coordinates are given by

$$\eta_i \equiv h_{-\alpha(m_i)} = m_i^{1+\alpha/2}. \tag{47.40}$$

It should be noted that the normalization condition $\sum_{i=0}^{n} m_i = 1$ that embeds Δ^n in $\mathbb{R}^n_+$ is a nonlinear constraint on the affine coordinates except for $\alpha = -1$ (for the dual coordinates, this happens for $\alpha = 1$). This recovers the fact that the Kullback–Leibler divergence is the only flat, invariant and decomposable structure on Δ^n.

For any monotonic function h, the so-called **h-representation** of x is defined as $h(x)$. Using this representation, one can define the **h-mean** as follows:

$$m_h(x, y) := h^{-1}\left(\frac{h(x) + h(y)}{2}\right). \tag{47.41}$$

The α-representations are exactly the ones inducing linearly scaling h-means:

$$m_\alpha(\lambda x, \lambda y) = \lambda m_\alpha(x, y). \tag{47.42}$$

It is not hard to see that all well-known means, such as the ordinary, geometric and harmonic means, are examples of α-means. Given an α-representation, one can define an α-mixture of distributions as the α-mean of the distributions (up to normalization). By allowing for weighted sums, the so-called **α-family** or **α-integration of distributions** with coordinate system $\{w_i\}_{1 \leq i \leq N}$ is obtained. The cases $\alpha = -1$ and $\alpha = 1$ can be seen to correspond to mixtures and exponential families, respectively.

47.2 Projections

The following theorem generalizes the classic Pythagorean theorem on $\mathbb{R}^n$ (and reduces to it when one chooses the Euclidean distance as the divergence measure).

Theorem 47.2.1 (Pythagoras). *Consider a triangle PQR on a dually flat manifold M with canonical divergence D. If the geodesic PQ and the dual geodesic QR are orthogonal, the following equation holds:*

$$D(P \,\|\, R) = D(P \,\|\, Q) + D(Q \,\|\, R)\,. \tag{47.43}$$

One obtains a conjugate statement by dualizing all quantities.

In Euclidean space (and in general Hilbert spaces), one of the most powerful theorems is the Projection Theorem 25.2.20, which states that the shortest path from a point to a subspace is given by orthogonal projection. This can be generalized to dually flat manifolds.

Definition 47.2.2 (Orthogonal projection). Consider a point $p \in M$ and a submanifold S of a dually flat manifold M such that $p \notin S$. A geodesic (orthogonal) projection of p on S is a point $p^* \in \partial S$ such that the affine geodesic connecting p and p^* is orthogonal to all of $T_{p^*}S$. One can obtain the notion of a dual geodesic projection in a similar way.

Because multiple projections exist in general, a strict minimality theorem cannot be formulated.

Theorem 47.2.3 (Projection theorem). *The extremal points of the map $s \mapsto D(p \,\|\, s)$ are geodesic projections of p onto S. The dual statement also holds.*

The strict projection theorem for Hilbert spaces only holds for affine subspaces. In the manifold setting, this is reflected by a flatness condition.

Property 47.2.4. If the submanifold S is $\widetilde{\nabla}$-flat, the geodesic projection of $p \notin S$ is unique and it minimizes the map $s \mapsto D(p \,\|\, s)$. The dual statement also holds.

The e/m-terminology also exists for projections.

Definition 47.2.5 (Projections). The e- and m-projections are defined as follows:

- e-projection: $\pi_e(p) := \arg\min_{q \in S} D_{\mathrm{KL}}(q \,\|\, p)$, and
- m-projection: $\pi_m(p) := \arg\min_{q \in S} D_{\mathrm{KL}}(p \,\|\, q)$.

47.2.1 Functional descent

The content of this section will be the derivation of an update scheme for probability measures using the functional analytic framework in the spirit of gradient descent. This is based on the paper Chu, Blanchet, and Glynn (2019).

Definition 47.2.6 (Probability functional). Consider the space of finite Borel measures $M_{\mathrm{Borel}}(X)$ on a compact Polish space X. A probability functional on X is a (linear) fuc-

tional $J : M_{\text{Borel}}(X) \to \mathbb{R}$. By the Riesz–Markov theorem 19.1.6[6], this is equivalent to a functional $J : C(X)^* \to \mathbb{R}$.

Definition 47.2.7 (Instance function). Consider a probability functional $J \in M_{\text{Borel}}(X)^*$. An instance function for J at $\mu \in M_{\text{Borel}}(X)$ is a function $\psi_\mu : X \to \mathbb{R}$ such that

$$\mathrm{d}J(\mu; \nu) = \int_X \psi_\mu \, d\nu \tag{47.44}$$

for all $\nu \in M_{\text{Borel}}(X)$.

Remark. Note that the Gâteaux differential (Definition 25.3.12) only makes sense for functions on vector spaces, which is why $M_{\text{Borel}}(X)$ was used instead of $\mathbb{P}(X)$.

Locally around $\mu_0 \in M_{\text{Borel}}(X)$, one can write down the Taylor expansion

$$J(\mu) \approx J(\mu_0) + \mathrm{d}J(\mu_0; \mu - \mu_0) \,.$$

If J admits an instance function at μ_0, this can be rewritten as

$$\begin{aligned} J(\mu) &\approx J(\mu_0) + \int_X \psi_{\mu_0}(x) \, d\mu - \int_X \psi_{\mu_0} \, d\mu_0 \\ &= J(\mu_0) - \mathsf{E}_{\mu_0}[\psi_{\mu_0}] + \mathsf{E}_\mu[\psi_{\mu_0}] \,. \end{aligned} \tag{47.45}$$

The first two terms are constants, so only the last expectation matters during minimization. An expression of this form is sometimes called a **von Mises representation**. The probability functional descent algorithm tries to improve the current estimate μ_0 of the true distribution $\hat{\mu}$ by iteratively decreasing the expectation $\mathsf{E}_\mu[\psi_{\mu_0}]$.

A different approach could be followed to define the probability functional and its Gâteaux derivative. The set $\mathbb{P}(X)$ has the structure of a manifold inside $M_{\text{Borel}}(X)$ modelled on a locally convex space. Definition 44.2.2 then gives rise to the tangent bundle $T\mathbb{P}(X)$.

@@ FORMALIZE THIS IDEA @@

[6] Compactness was only assumed to freely apply this theorem.

Chapter 48

Statistics

In this chapter, most definitions and formulas will be based on either a standard calculus approach or a data-driven approach. See Chapter 45 for a measure-theoretic approach. For some sections, the language of information geometry will be used (as introduced in the previous chapter).

Chapter Contents

48.1 Data samples **907**
48.1.1 Moment estimators 907
48.1.2 Dispersion 909
48.2 Empirical distribution **910**
48.3 Errors **913**
48.4 Parameter estimation **914**
48.4.1 General properties 914
48.4.2 Common estimators 915
48.4.3 Estimation error 915
48.4.4 Likelihood function 916
48.4.5 Maximum likelihood estimation 916
48.4.6 Least squares estimation 917
48.4.7 Geometric approach 919
48.5 Bayesian modelling **919**
48.6 Confidence intervals **920**
48.6.1 Interval types 920
48.6.2 General construction 921
48.6.3 Interval for a sample mean 921

48.7 Hypothesis testing **922**
48.7.1 Testing 922
48.7.2 Comparison tests 923
48.7.3 Post-hoc tests 925
48.8 Goodness of fit **926**
48.8.1 χ^2-test 926
48.8.2 Runs test 927

48.1 Data samples

In this section, $(x_i)_{i\leq n}$ will always denoted a finite data set in $\mathbb{R}$ (or a suitable vector space V). Note the notation for families instead of subsets (cf. Section 1.6.1). What are called data sets in statistics and data science are actually not necessarily sets but **multisets**, i.e. copies are allowed.

48.1.1 Moment estimators

Formula 48.1.1 (r^{th} sample moment).

$$\mu_r := \overline{x^r} = \frac{1}{n}\sum_{i=1}^{n} x_i^r \tag{48.1}$$

Example 48.1.2 (Arithmetic mean). The arithmetic mean is used to average out differences between measurements. It is defined as the first sample moment:

$$\overline{x} := \frac{1}{n}\sum_{i=1}^{n} x_i\,. \tag{48.2}$$

Theorem 48.1.3 (Weak law of large numbers[1]). *Assume that the sequence $(X_n)_{n\in\mathbb{N}}$ of random variables is i.i.d. with finite expectation μ. The sample average converges in probability (Definition 45.3.16) to the expectation value:*

$$\lim_{n\to\infty} \mathrm{Prob}\Big(|\overline{X}_n - \mu| > \varepsilon\Big) = 0 \tag{48.3}$$

for all $\varepsilon > 0$.

Convergence in probability can also be strengthened to almost sure convergence.

Theorem 48.1.4 (Strong law of large numbers[2]). *Assume that the sequence $(X_n)_{n\in\mathbb{N}}$ of random variables is i.i.d. with finite expectation μ. The sample average converges almost surely*

[1] Also called **Khinchin's law**.
[2] Also called **Kolmogorov's law**.

(Definition 18.1.5) to the expectation value:

$$\mathsf{Prob}\left(\lim_{n\to\infty} \overline{X}_n = \mu\right) = 1\,. \tag{48.4}$$

In fact, the i.i.d. assumption can be weakened. The convergence

$$\overline{X}_n \xrightarrow{a.s.} \mathsf{E}\left[\overline{X}_n\right] \tag{48.5}$$

holds as long as the random variables are independent and satisfy the ***Kolmogorov criterion***

$$\sum_{k=1}^{+\infty} \frac{1}{k^2}\mathsf{Var}\left[X_k\right] < +\infty\,. \tag{48.6}$$

Definition 48.1.5 (Weighted mean). Let $f : \mathbb{R} \to \mathbb{R}^+$ be a weight function. The f-weighted mean is given by

$$\overline{x} := \frac{\sum_i f(x_i)x_i}{\sum_i f(x_i)}\,. \tag{48.7}$$

Example 48.1.6 (Binned mean). If the data has been grouped in $b \in \mathbb{N}$ bins, the weight function is often given by the number of elements in each bin.

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{b} n_i x_i\,, \tag{48.8}$$

where $n_i \in \mathbb{N}_0$ is the number of observations in the bin b_i.

Remark 48.1.7. In the above definitions, the measurements x_i can be replaced by function values $f(x_i)$ to calculate the mean of the function $f(x)$. This follows from Theorem 45.3.8. However, it is also important to keep in mind that $\overline{f(x)} \neq f(\overline{x})$. The equality only holds for linear functions.

Definition 48.1.8 (Geometric mean). Let $(x_i)_{i\leq n}$ be a data set taking values in $\mathbb{R}^+$. The geometric mean is used to average out *normalized* measurements, i.e. ratios with respect to a reference value.

$$g := \left(\prod_{i=1}^{n} x_i\right)^{1/n} \tag{48.9}$$

The following relation exists between the arithmetic and geometic mean:

$$\ln g = \overline{\ln x}\,. \tag{48.10}$$

Definition 48.1.9 (Harmonic mean).

$$h := \left(\frac{1}{n}\sum_{i=1}^{n} x_i^{-1}\right)^{-1} \tag{48.11}$$

The following relation exists between the arithmetic and harmonic means:

$$h^{-1} = \overline{x^{-1}}\,. \tag{48.12}$$

Property 48.1.10. Let $(x_i)_{i\leq n}$ be a data set taking values in $\mathbb{R}^+$.

$$h \leq g \leq \overline{x} \tag{48.13}$$

The equalities only hold when all x_i are equal.

Definition 48.1.11 (Mode). The most occurring value in a data set.

Definition 48.1.12 (Median). The element $x_{(0.5)}$ in a data set such that half of the values is greater than $x_{(0.5)}$ and half of the values is smaller than $x_{(0.5)}$. (Cf. Definition 48.1.23 further below.)

48.1.2 Dispersion

Definition 48.1.13 (Range). The simplest indicator for statistical dispersion:

$$R := x_{\max} - x_{\min} \,. \tag{48.14}$$

It should be noted that this quantity is very sensitive for outliers.

Definition 48.1.14 (Mean absolute difference).

$$\mathsf{MD} := \frac{1}{n}\sum_{i=1}^{n} |x_i - \overline{x}| \tag{48.15}$$

Formula 48.1.15 (r^{th} central sample moment).

$$m_r := \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^r \tag{48.16}$$

Definition 48.1.16 (Sample variance).

$$\mathsf{Var}(x) := \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^2 \tag{48.17}$$

Formula 48.1.17. The variance can also be rewritten in the following way (cf. Eq. (45.41)):

$$\mathsf{Var}(x) = \overline{x^2} - \overline{x}^2 \,. \tag{48.18}$$

Remark 48.1.18 (Bessel corection). A better estimator for the variance of a sample is given by the following formula:

$$\hat{s}^2 := \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \overline{x})^2 \,. \tag{48.19}$$

See Formula 48.4.9 for more information.

Definition 48.1.19 (Skewness). The skewness γ describes the 'asymmetry' of a distribution. It is defined as the proportionality constant relating the third central moment and the standard deviation:

$$m_3 = \gamma\sigma^3 \,. \tag{48.20}$$

A positive skewness indicates a tail to the right or, alternatively, a median smaller than $\overline{x}$. A negative skewness indicates a median greater than $\overline{x}$.

Definition 48.1.20 (Pearson's mode skewness).

$$\gamma_P := \frac{\overline{x} - \text{mode}}{\sigma} \tag{48.21}$$

Definition 48.1.21 (Kurtosis). The kurtosis c is an indicator for the 'tailedness'. It is defined as the proportionality constant relating the fourth central moment and the standard deviation:

$$m_4 = c\sigma^4 \,. \tag{48.22}$$

Definition 48.1.22 (Excess kurtosis). The excess kurtosis is defined as $c - 3$. This fixes the excess kurtosis of all univariate normal distributions at 0. A positive excess is an indicator for long, 'fat' tails, whereas a negative excess indicates short, 'thin' tails.

Definition 48.1.23 (Percentile). The p-percentile of a cumulative distribution F is defined as follows:

$$x_{(p\%)} := F^{-1}(p\%) \,. \tag{48.23}$$

The p%-percentile is, hence, such that it is greater than p% of the measurements. The median (Definition 48.1.12) is the 50-percentile. The percentiles for $0, 25\%, 50\%, 75\%$ and 100% are also called **quartiles**. If, instead of using percentages, real numbers $p \in [0, 1]$ are used, these quantities are also called **quantiles**.

Definition 48.1.24 (Interquartile range). The difference between the upper and lower quartile (i.e. the 75- and 25-percentiles, respectively).

Definition 48.1.25 (Full Width at Half Maximum). The difference between the two values of the independent variable where the dependent variable is half of its maximum. This quantity is often denoted by the abbreviation **FWHM**. Note that this only makes sense for unimodal distributions, i.e. distributions with a single peak.

Property 48.1.26. For Gaussian distributions $\mathcal{N}(\mu, \sigma^2)$, the following relation exists between the FWHM and the standard deviation:

$$\text{FWHM} = 2.35\sigma \,. \tag{48.24}$$

48.2 Empirical distribution

In this section, the probability distribution induced by a data sample will be considered.

Definition 48.2.1 (Empirical distribution function). The empirical probability distribution function, corresponding to a data sample $(x_i)_{i \leq n}$, is defined as the uniform mixture distribution with Dirac measures at the data points:

$$P_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i} \,. \tag{48.25}$$

Remark 48.2.2. Most of the preceding section can be recovered by applying Section 45.3.2 to the empirical distribution function of a data sample.

Theorem 48.2.3 (Borel's law of large numbers). *If the sample size approaches infinity, the observed frequencies approach the theoretical probabilities.*

Corollary 48.2.4 (Frequentist probability[3]).

$$\mathrm{Prob}(x) := \lim_{n\to\infty} \frac{f_n(x)}{n}\,, \tag{48.26}$$

where f_n denotes the probability mass function of P_n.

The law of large numbers can also be phrased in terms of the empirical distribution function.

Theorem 48.2.5 (Glivenko–Cantelli). *Consider a cumulative distribution function F and denote the empirical cumulative distribution function of $n \in \mathbb{N}$ random variables by F_n. If the random variables are i.i.d. according to F, then*

$$\sup_{x\in\mathbb{R}} |F(x) - F_n(x)| \xrightarrow{a.s.} 0\,. \tag{48.27}$$

Remark 48.2.6. The law of the large numbers implies pointwise convergence of the empirical distribution function, while the Glivenko–Cantelli theorem strengthens this to uniform convergence.

The quantity in the Glivenko–Cantelli theorem is important enough to get its own name.

Definition 48.2.7 (Kolmogorov–Smirnov statistic). Let F be a given cumulative distribution function. The nth Kolmogorov–Smirnov statistic is defined as follows:

$$D_n := \sup_{x\in\mathbb{R}} |F_n(x) - F(x)|\,. \tag{48.28}$$

@@ ADD KS METRIC AND TV METRIC @@

Definition 48.2.8 (Kolmogorov distribution).

$$F_{\mathrm{Kol}}(x) := 1 - 2\sum_{i=1}^{+\infty} (-1)^{i-1} e^{-2i^2x^2} = \frac{\sqrt{2\pi}}{x} \sum_{i=1}^{+\infty} e^{-(2i-1)^2\pi^2/(8x^2)} \tag{48.29}$$

Property 48.2.9 (Kolmogorov–Smirnov test). Let the null hypothesis H_0 state that a given data sample is described by a cumulative distribution function F. The null hypothesis is rejected at significance level α if

$$\sqrt{n} D_n > K_\alpha\,, \tag{48.30}$$

where K_α is defined by the Kolmogorov distribution: $F_{\mathrm{Kol}}(K_\alpha) := 1 - \alpha$.

[3] Also called the **empirical probability**.

Definition 48.2.10 (Glivenko–Cantelli class). Consider a set of measurable functions $\mathcal{F}$ on a measurable space (Ω, Σ). For every probability measure P on Ω, one can define the $\mathcal{F}$-norm as follows:

$$\|P\|_{\mathcal{F}} := \sup\{\mathsf{E}_P[f] \mid f \in \mathcal{F}\}\,. \tag{48.31}$$

A class $\mathcal{F}$ of measurable functions is said to be Glivenko–Cantelli (GC) with respect to a probability measure P if it satisfies

$$\|P_n - P\|_{\mathcal{F}} \xrightarrow{\text{a.s.}} 0\,, \tag{48.32}$$

where P_n is the empirical measure.[4] The Glivenko–Cantelli theorem 48.2.5 says that the indicator functions of the sets $]-\infty, x]$ form a GC class.[5] In fact, they are **universally GC** because this theorem applies to all probability measures on Ω. A class is said to be **uniformally GC** if the convergence holds uniformly over all probability measures.

Remark 48.2.11. Note that, by the law of large numbers, every singleton class is Glivenko–Cantelli (also universally and uniformly). The above definition strengthens the convergence of all elements of $\mathcal{F}$ to uniform convergence.

Property 48.2.12 (Bracketing number). Consider a collection of measurable functions $\mathcal{F}$ on a meaurable space (Ω, Σ) and recall Definition 25.1.17 of the bracketing number. If the bracketing number $N_{[]}(\varepsilon, \mathcal{F}, \|\cdot\|_1)$ is finite for all $\varepsilon > 0$, then $\mathcal{F}$ is Glivenko–Cantelli (with respect to the probability measure that induces the L^1-norm $\|\cdot\|_1$).

Property 48.2.13 (Metric entropy). Consider a collection of measurable functions $\mathcal{F}$ on a meaurable space (Ω, Σ) and recall Definition 9.5.3 of the metric entropy. Moreover, assume that $\mathcal{F}$ admits an integrable envelope F. Let $\mathcal{F}_M$ denote the collection of functions $f\mathbb{1}_{F \le M}$, where $f \in \mathcal{F}$. If

$$\frac{1}{n} \ln N_C(\varepsilon, \mathcal{F}_M, \|\cdot\|_1) \xrightarrow{d} 0\,, \tag{48.33}$$

where $\|\cdot\|_1$ is the L^1-norm associated to the empirical measure P_n, for all $\varepsilon > 0$ and $M > 0$, then $\mathcal{F}$ is Glivenko–Cantelli.

Property 48.2.14 (Symmetrized empirical measure). Consider a set of i.i.d. Rademacher variables $\{\sigma_1, \dots, \sigma_n\}$ (Example 45.3.2). The symmetrized empirical measure is defined as follows:

$$P_n^\sigma : f \mapsto \frac{1}{n} \sum_{i=1}^{n} \sigma_i f(x_i)\,. \tag{48.34}$$

Given a collection $\mathcal{F}$ of measurable functions on (Ω, Σ, P), the following inequality holds:

$$\mathsf{E}\big[\|P_n - P\|_{\mathcal{F}}\big] \le 2\,\mathsf{E}\big[\|P_n^\sigma\|_{\mathcal{F}}\big]\,. \tag{48.35}$$

[4] If the convergence only holds in probability, the class is said to be **weakly GC**.

[5] Because every indicator function is uniquely associated to a set, one can also speak of GC classes of measurable sets.

Theorem 48.2.15 (Donsker). *Consider an empirical distribution function F_n and define its normalized empirical process as*

$$G_n := \sqrt{n}(F_n - F)\,, \tag{48.36}$$

where F is the cumulative distribution function of the random variables X_i. The central limit theorem says that the empirical process converges in distribution to a standard normal distribution for every $x \in \mathbb{R}$. This can be strengthened as follows:

$$G_n \xrightarrow{d} U \tag{48.37}$$

in $D(\mathbb{R}, \|\cdot\|_\infty)$, where U is a standard Brownian bridge and $D(\mathbb{R}, \|\cdot\|_\infty)$ denotes the space of càdlàg functions equipped with the supremum metric (Example 9.3.3).

48.3 Errors

Definition 48.3.1 (Systematic error). Errors that always have the same effect independent of the measurements itself, i.e. they shift all values in the same way and cannot be directly inferred from the measurements. Note that they are not necessarily independent of each other.

Formula 48.3.2 (Inverse-variance averaging). When performing a sequence of measurements x_i with different variances σ_i^2, it is impossible to use the arithmetic mean (Example 48.1.2) in a meaningful way because the measurements are not of the same type. Therefore, it is also impossible to apply the CLT 45.4.17.

These problems can be resolved by the using the weighted mean (Definition 48.1.5):

$$\overline{x} := \frac{\sum_i x_i \sigma_i^{-2}}{\sum_i \sigma_i^{-2}}\,. \tag{48.38}$$

The variance of the weighted mean is given by

$$\mathsf{Var}[\overline{x}] := \frac{1}{\sum_i \sigma_i^{-2}}\,. \tag{48.39}$$

Formula 48.3.3 (Error propagation). Let X be a vector random variable such that functions of X admit a good first-order Taylor approximation around the mean (this usually means that the covariance is small). The variance of a general function of X is given by

$$\mathsf{Var}\big[f(X)\big] \approx \sum_{i=1}^{n} \left(\frac{\partial f}{\partial X_i}\right)^2 \mathsf{Var}\big[X_i\big] + \sum_{i \neq j} \left(\frac{\partial f}{\partial X_i}\right)\left(\frac{\partial f}{\partial X_j}\right) \mathsf{cov}[X_i, X_j]\,. \tag{48.40}$$

Definition 48.3.4 (Fractional error). Let X, Y be two independent random variables. The standard deviation of $f(X, Y) = XY$ is given by the fractional error:

$$\left(\frac{\sigma_f}{f}\right)^2 = \left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2\,. \tag{48.41}$$

The fractional error of a variable is equal to the fractional error of the reciprocal of that variable.

Property 48.3.5 (Logarithm). Let X be a random variable. The error of the logarithm of X is equal to the fractional error of X.

Formula 48.3.6 (Covariance of functions).

$$\mathrm{cov}[f,g] \approx \sum_{i,j} \left(\frac{\partial f}{\partial X_i}\right)\left(\frac{\partial g}{\partial X_j}\right)\mathrm{cov}[X_i, X_j] \tag{48.42}$$

Corollary 48.3.7. Let $f \equiv (f_1, \dots, f_k)$ be a vector-valued function. The covariance matrix is to first order given by

$$\mathrm{Var}\big[f(X)\big] \approx J\mathrm{Var}[X]J^T\,, \tag{48.43}$$

where J is the Jacobian matrix of f.

Corollary 48.3.8. The correlation coefficient (Definition 45.3.39) of a random variable X and a <u>linear</u> function of X is independent of σ_x and is equal to ± 1.

48.4 Parameter estimation

48.4.1 General properties

Definition 48.4.1 (Consistency). An estimator $\hat{a}$ is said to be consistent if it is asymptotically equal to the true parameter:

$$\lim_{n\to\infty} \hat{a} = a\,. \tag{48.44}$$

Definition 48.4.2 (Unbiased estimator). An estimator $\hat{a}$ is said to be unbiased if its expectation value is equal to the true parameter:

$$\langle \hat{a} \rangle = a\,. \tag{48.45}$$

Note that neither consistency, nor unbiasedness implies the other.

Definition 48.4.3 (Bias).

$$B(\hat{a}) := |\langle \hat{a} \rangle - a|\,. \tag{48.46}$$

Definition 48.4.4 (Mean squared error).

$$\mathrm{MSE}(\hat{a}) := B(\hat{a})^2 + \mathrm{Var}(\hat{a})\,. \tag{48.47}$$

Remark 48.4.5. If an estimator is unbiased, the MSE is equal to the variance of the estimator.

48.4.2 Common estimators

Property 48.4.6 (Unbiased mean). The CLT 45.4.17 implies that the sample mean (Example 48.1.2) is a consistent and unbiased estimator of the population mean.

Formula 48.4.7 (Standard error of the mean). Using the Bienaymé formula 45.3.42, one can show that the standard error of the mean, i.e. the standard deviation of the sample mean, is given by the following formula:

$$\mathsf{Var}\left[\overline{X}\right] = \frac{\sigma^2}{n}\,. \tag{48.48}$$

Formula 48.4.8 (Variance estimator for known mean). If the true mean μ is known, a consistent and unbiased estimator for the variance is given by

$$\widehat{\mathsf{Var}[X]} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2\,. \tag{48.49}$$

Formula 48.4.9 (Variance estimator for unknown mean). If the true mean is unknown and the sample mean has been used to estimate it, a consistent and unbiased estimator is given by

$$\hat{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2\,. \tag{48.50}$$

The modified factor $\frac{1}{n-1}$ is called the **Bessel correction**. It corrects the bias of the estimator given by the sample variance (Definition 48.1.16). The consistency is guaranteed by the CLT.

Property 48.4.10 (Characterization of normal distributions). The class of normal distributions is uniquely characterized by those distributions for which the sample mean and sample variance are independent.

48.4.3 Estimation error

Formula 48.4.11 (Variance of the estimator of the variance).

$$\mathsf{Var}\left(\widehat{\mathsf{Var}[X]}\right) = \frac{(n-1)^2}{n^3}\langle(x - \langle x\rangle)^4\rangle - \frac{(n-1)(n-3)}{n^3}\langle(x - \langle x\rangle)^2\rangle^2 \tag{48.51}$$

Formula 48.4.12 (Variance of the estimator of the standard deviation).

$$\mathsf{Var}\left[\widehat{\sigma}\right] = \frac{1}{4\sigma^2}\mathsf{Var}\left[\widehat{\mathsf{Var}[X]}\right] \tag{48.52}$$

Remark 48.4.13. The previous result is a little odd, as one has to know the true standard deviation to compute the variance of the estimator. This problem can be solved in two ways. Either a value (hopefully close to the real one) inferred from the sample is used as an estimator, or a guess is used in the design phase of an experiment to see what the possible outcomes are.

48.4.4 Likelihood function

Definition 48.4.14 (Likelihood). The likelihood $\mathcal{L}(a;\mathbf{x})$ is the joint density of a set of measurements $\mathbf{x} := \{x_1, \dots, x_n\}$:

$$\mathcal{L}(a;\mathbf{x}) = \prod_{i=1}^{n} f(x_i;a)\,. \tag{48.53}$$

Theorem 48.4.15 (Cramer–Rao bound). *The variance of an unbiased estimator has a lower bound called the Cramer–Rao bound or* ***minimum variance bound*** *(**MVB**):*

$$\mathsf{Var}[\hat{a}] \geq \frac{1}{\left\langle \left(\frac{\mathrm{d}\ln\mathcal{L}}{\mathrm{d}a}\right)^2 \right\rangle}\,. \tag{48.54}$$

For a biased estimator with bias b, the MVB takes on the following form:

$$\mathsf{Var}[\hat{a}] \geq \frac{\left(1 + \frac{\mathrm{d}b}{\mathrm{d}a}\right)^2}{\left\langle \left(\frac{\mathrm{d}\ln\mathcal{L}}{\mathrm{d}a}\right)^2 \right\rangle}\,. \tag{48.55}$$

Remark 48.4.16.

$$\left\langle \left(\frac{\mathrm{d}\ln\mathcal{L}}{\mathrm{d}a}\right)^2 \right\rangle = -\left\langle \frac{\mathrm{d}^2\ln\mathcal{L}}{\mathrm{d}a^2} \right\rangle \tag{48.56}$$

Definition 48.4.17 (Fisher information).

$$I_X(a) := \left\langle \left(\frac{\mathrm{d}\ln\mathcal{L}}{\mathrm{d}a}\right)^2 \right\rangle = n\int \left(\frac{\mathrm{d}\ln f}{\mathrm{d}a}\right)^2 f\,d\mu \tag{48.57}$$

Using this definition one can rewrite the Cramer–Rao inequality as follows:

$$\mathsf{Var}[\hat{a}] \geq I_X(a)\,. \tag{48.58}$$

Definition 48.4.18 (Finite-sample efficiency). An unbiased estimator is said to be (finite-sample) efficient if it saturates the Cramer–Rao bound. In general the **efficiency** of (unbiased) estimators is defined through the Cramer–Rao bound as follows:

$$e(\hat{a}) := \frac{I_X(a)^{-1}}{\mathsf{Var}[\hat{a}]}\,. \tag{48.59}$$

48.4.5 Maximum likelihood estimation

From Definition 48.4.14, it follows that the estimator $\hat{a}_{\mathrm{MLE}}$ that makes the given measurements most probable is the value of a for which the likelihood function is maximal. It is, therefore, not the most probable estimator.

Using Bayes' theorem, one finds $f(a \mid x) = f(x \mid a)\frac{f(a)}{f(x)}$. The prior density $f(x)$ is fixed since the values x_i are given by the measurement and, hence, do not vary. The density $f(a)$ is generally assumed to be uniform if there is no prior knowledge about a. It

follows, that $f(a \mid x)$ and $f(x \mid a)$ are proportional and, hence, the logarithms of these functions differ only by an additive constant. This leads to following method for finding an estimator $\hat{a}$:

Method 48.4.19 (Maximum likelihood estimator). The maximum likelihood estimator $\hat{a}$ is obtained by solving the following equation:

$$\left.\frac{\mathrm{d}\ln\mathcal{L}}{\mathrm{d}a}\right|_{a=\hat{a}} = 0\,. \tag{48.60}$$

Remark 48.4.20. MLE estimators are mostly consistent but often biased.

Property 48.4.21. MLE estimators are invariant under parameter transformations.

Corollary 48.4.22. The invariance implies that the two estimators $\hat{a}$ and $\widehat{f(a)}$ cannot both be unbiased at the same time.

Property 48.4.23. Every consistent estimator asymptotically becomes unbiased and efficient.

Property 48.4.24 (Minimizing KL-divergence). It can be shown that maximizing the log-likelihood is equivalent to minimizing the Kullback–Leibler divergence (Definition 45.5.3) between the would-be density $f(x;\theta)$ and the true density $q(x)$:

$$\begin{aligned}
\arg\max_\theta \ln\mathcal{L} &= \arg\max_\theta \sum_{i=1}^n \ln f(x_i;\theta) \\
&= \arg\max_\theta \sum_{i=1}^n \ln f(x_i;\theta) - \ln q(x_i) \\
&= \arg\min_\theta \frac{1}{n}\sum_{i=1}^n \ln\frac{q(x_i)}{f(x_i;\theta)} \\
&\longrightarrow \arg\min_\theta \int q(x;\theta)\ln\frac{q(x)}{f(x;\theta)}\,dx = \arg\min_\theta D_{\mathrm{KL}}(q\,\|\,f_\theta)\,,
\end{aligned}$$

where the law of large numbers was used in the last line.

48.4.6 Least squares estimation

To fit a (parametric) function $y = f(x;a)$ to a set of 2 variables (x,y), where the x values are exact and the y values have an uncertainty σ_i, one can use the following method.

Method 48.4.25 (Least squares).

1. For every event (x_i, y_i), define the residual $d_i := y_i - f(x_i;a)$.
2. Determine the χ^2-statistic (analytically):

$$\chi^2 := \sum_i \frac{d_i^2}{f_i}\,, \tag{48.61}$$

   where $f_i := f(x_i;a)$.

3. Find the most probable value of $\hat{a}$ by solving the equation

$$\frac{\mathrm{d}\chi^2}{\mathrm{d}a} = 0\,. \tag{48.62}$$

Property 48.4.26. The optimal χ^2-value is asymptotically distributed according to a χ^2-distribution with $N \in \mathbb{N}$ degrees of freedom. The parameter N is equal to the number of events $n \in \mathbb{N}$ minus the number of fitted parameters $k \in \mathbb{N}$. (See more in Section 48.8.1.)

Formula 48.4.27 (Linear fit). When all uncertainties σ_i are equal, the slope $\widehat{m}$ and intercept $\hat{c}$ are given by the following formulas:

$$\widehat{m} := \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{\mathrm{cov}(x,y)}{\mathrm{Var}(x)}\,, \tag{48.63}$$

$$\hat{c} := \overline{y} - \widehat{m}\overline{x} = \frac{\overline{x^2} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2}\,. \tag{48.64}$$

Remark 48.4.28. The equation $\overline{y} = \hat{c} + \widehat{m}\overline{x}$ says that the linear fit passes through the center of mass $(\overline{x}, \overline{y})$.

Formula 48.4.29 (Errors of linear fit).

$$\mathrm{Var}\left[\widehat{m}\right] = \frac{1}{n(\overline{x^2} - \overline{x}^2)}\sigma^2 \tag{48.65}$$

$$\mathrm{Var}[\hat{c}] = \frac{\overline{x^2}}{n(\overline{x^2} - \overline{x}^2)}\sigma^2 \tag{48.66}$$

$$\mathrm{cov}(\widehat{m}, \hat{c}) = \frac{-\overline{x}}{n(\overline{x^2} - \overline{x}^2)}\sigma^2 \tag{48.67}$$

The least squares method is very useful to fit data that has been grouped in bins (histograms).

Method 48.4.30 (Binned least squares).

1. $n \in \mathbb{N}$ i.i.d. events with density $f(X;a)$ divided in $N_B \in \mathbb{N}$ bins, where the interval j is centred on the value x_j, has a width W_j and contains $n_j \in \mathbb{N}$ events.
2. The ideally expected number of events in the jth interval: $f_j = NW_j f(x_j;a)$.
3. The real number of events has a Poisson distribution: $\overline{n}_j = \sigma_j^2 = f_j$.
4. Define the binned χ^2 as

$$\chi^2 := \sum_{i=1}^{N_B} \frac{(n_i - f_i)^2}{f_i^2}\,. \tag{48.68}$$

48.4.7 Geometric approach

Consider a sample $\mathbf{x} := \{x_1, \dots, x_n\}$ drawn from a distribution $f(x;\theta)$ in an exponential family. The likelihood (Definition 48.4.14) is given by

$$\mathcal{L}(\theta;\mathbf{x}) := \prod_{i=1}^{n} f(x_i;\theta)\,.$$

The m-coordinates of the observed point are

$$\eta = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}\,. \tag{48.69}$$

The optimal value for θ can be found by maximizing the log-likelihood as before. In Property 48.4.24, it was shown that this is equivalent to minimizing the Kullback–Leibler divergence between the 'true' distribution $p(x;\xi)$ and the variational solution $f(x;\theta)$. However, in practice, the true distribution is not known. Luckily, one can replace the true distribution by the empirical distribution in the proof of Property 48.4.24. Minimization then corresponds to m-projecting the observed point η on the submanifold S of 'admissible' distributions.

Theorem 48.4.31 (Sanov). *Consider a probability distribution q on a finite set S and draw n i.i.d. samples. Let P_n be the empirical distribution function of the samples (Definition 48.2.1). Furthermore, let Γ be a collection of probability distributions such that $P_n \in \Gamma$. The joint distribution q^n satisfies the following inequality:*

$$q^n(P_n \in \Gamma) \leq (n+1)^{|S|} 2^{-nD_{\mathrm{KL}}(p^* \,\|\, q)}\,, \tag{48.70}$$

where p^* is the information projection of q on Γ. If $\Gamma = \overline{\Gamma^\circ}$, this can be restated as

$$\lim_{n\to\infty} \frac{1}{n} q^n(P_n \in \Gamma) = -D_{\mathrm{KL}}(p^* \,\|\, q)\,. \tag{48.71}$$

48.5 Bayesian modelling

Definition 48.5.1 (Conjugate distributions). Consider a prior distribution $F(\theta)$ and a posterior distribution $F(\theta \mid X)$. If these distributions belong to the same family, e.g. they are both Gaussians, they are said to be conjugate. In this case, the prior $F(\theta)$ is said to be a **conjugate prior** for the likelihood $F(X \mid \theta)$.

Example 48.5.2. A common example is the case of binomial distributions, where the conjugate prior is the *β-distribution*. This can be generalized to multi-class situations. The conjugate prior of a categorical (or even *multinomial*) distribution is the *Dirichlet distribution*.

48.6 Confidence intervals

The true value of a parameter ε can never be known exactly. However, it is possible to construct an interval I in which this value should lie with a certain confidence C.

Example 48.6.1 (Prediction interval). Let X be a normally distributed random variable. An observation of X will lie in the interval $[\mu - 1.96\sigma, \mu + 1.96\sigma]$ with 95% probability. The true value μ lies in the interval $[x - 2\sigma, x + 2\sigma]$ with 95% confidence.

Remark. In the previous example, some assumptions were made. All possible values (left or right side of peak) are given the same probability due to the Gaussian distribution. If one removes this symmetry condition, a more careful approach is required. Furthermore, the apparent symmetry between the uncertainty and confidence levels is only valid for Gaussian distributions.

48.6.1 Interval types

Definition 48.6.2 (Two-sided confidence interval).

$$\mathsf{Prob}(x_- \leq X \leq x_+) = \int_{x_-}^{x_+} f(x)\,dx = C \tag{48.72}$$

There are three possible (often used) two-sided intervals:

- **symmetric interval**: $x_+ - \mu = \mu - x_-$,
- **shortest interval**: $|x_+ - x_-|$ is minimal, or
- **central interval**: $\int_{-\infty}^{x_-} f(x)\,dx = \int_{x_+}^{+} \infty f(x)\,dx = \frac{1-C}{2}$.

The central interval is the most widely used confidence interval.

Remark 48.6.3. For Gaussian distributions, these three definitions are equivalent.

Definition 48.6.4 (One-sided confidence interval).

$$\mathsf{Prob}(x \geq x_-) = \int_{x_-}^{+\infty} f(x)\,dx = C \tag{48.73}$$

$$\mathsf{Prob}(x \leq x_+) = \int_{-\infty}^{x_+} f(x)\,dx = C \tag{48.74}$$

Definition 48.6.5 (Discrete central confidence interval). For a discrete distribution, it is often impossible to find integers $x_\pm$ such that the real value lies with exact confidence

C in the interval $[x_-, x_+]$.

$$x_- = \arg\min_\theta \left[\frac{1-C}{2} - \sum_{x=0}^{\theta-1} p(x)\right] \tag{48.75}$$

$$x_+ = \arg\min_\theta \left[\frac{1-C}{2} - \sum_{x=\theta+1}^{+\infty} p(x)\right] \tag{48.76}$$

48.6.2 General construction

For every value of the true parameter X it is possible to construct a confidence interval. This leads to the construction of two functions $x_-(X)$ and $x_+(X)$. The 2D diagram obtained by plotting $x_-(X)$ and $x_+(X)$ with the x-axis horizontally and X-axis vertically is called the **confidence region**.

Method 48.6.6. Let x_0 be a point estimate of the parameter X. From the confidence region it is possible to infer a confidence interval $[X_-(x), X_+(x)]$, where the upper limit X_+ is not the limit such that there is only a $\frac{1-C}{2}$ chance of having a true parameter $X \geq X_+$, but the limit such that if the true parameter $X \geq X_+$ then there is a chance of $\frac{1-C}{2}$ to have a measurement x_0 or smaller.

48.6.3 Interval for a sample mean

Formula 48.6.7 (Interval with known variance). If the sample size is large enough, the true distribution is unimportant, since the CLT ensures a Gaussian distribution of the sample mean $\overline{X}$. The α-level confidence interval such that $\mathrm{Prob}(-z_{\alpha/2} < Z < z_{\alpha/2})$ with $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ is given by

$$\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]. \tag{48.77}$$

Remark 48.6.8. If the sample size is not sufficiently large, the measured quantity must follow a normal distribution.

Formula 48.6.9 (Interval with unknown variance). To account for the uncertainty of the estimated standard deviation $\widehat{\sigma}$, the student-t distribution (Definition 45.4.30) is used instead of a Gaussian distribution to describe the sample mean $\overline{X}$. The α-level confidence interval is given by

$$\left[\overline{X} - t_{\alpha/2;(n-1)}\frac{s}{\sqrt{n}}, \overline{X} + t_{\alpha/2;(n-1)}\frac{s}{\sqrt{n}}\right], \tag{48.78}$$

where s is the estimated standard deviation (Formula 48.4.9).

Formula 48.6.10 (Wilson score interval). For a sufficiently large sample, a sample proportion $\widehat{P}$ is approximately Gaussian distributed with expectation value π and variance

$\frac{\pi(\pi-1)}{n}$. The α-level confidence interval is given by

$$\left[\frac{(2n\widehat{P} + z_{\alpha/2}^2) - z_{\alpha/2}\sqrt{z_{\alpha/2}^2 + 4n\widehat{P}(1-\widehat{P})}}{2(n + z_{\alpha/2}^2)}, \frac{(2n\widehat{P} + z_{\alpha/2}^2) + z_{\alpha/2}\sqrt{z_{\alpha/2}^2 + 4n\widehat{P}(1-\widehat{P})}}{2(n + z_{\alpha/2}^2)}\right]. \tag{48.79}$$

Remark. The expectation value and variance are these of a binomial distribution (Definition 45.4.23) with $r = X/n$.

48.7 Hypothesis testing

Definition 48.7.1 (Simple hypothesis). A hypothesis where the distribution is fully specified.

Definition 48.7.2 (Composite hypothesis). A hypothesis where the distribution is given relative to some parameter values.

48.7.1 Testing

Definition 48.7.3 (Type I error). Rejection of a true null hypothesis.

Definition 48.7.4 (Type II error). Acceptance of a false null hypothesis.

Definition 48.7.5 (Significance). The probability of making a type-I error.

Property 48.7.6. Let $\alpha_1 > \alpha_2$. An α_2-level test is also significant at the α_1-level.

Remark 48.7.7. For discrete distributions, it is not always possible to achieve an exact level of significance.

Remark. Type-I errors occur occasionally. They cannot be prevented, one can only try to control them.

Definition 48.7.8 (Power). The probability of not making a type-II error.

Remark 48.7.9. A good test is a test with a small significance and a large power. The probabilities P_{I} and P_{II} should be as different as possible.

Definition 48.7.10 (Shapiro–Wilk test). After obtaining a data sample, it is often interesting to see if the data is distributed normally, since many other tests and methods assume a normal distribution. The Shapiro–Wilk test considers the following test statistic:

$$W := \frac{\sum_{i=1}^{n}(a_i x_{(i)})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}, \tag{48.80}$$

where

- $x_{(i)}$ are the order statistics,

- $(a_i, \dots, a_n) := \frac{m^T V}{\|V^{-1} m\|}$,
- $m \equiv (m_1, \dots, m_n)$ are the expectation values of the order statistics of i.i.d. standard normal distributions, and
- V is the covariance matrix of m.

The test statistic does not follow a known distribution and all critical values are calculated with Monte Carlo simulations.

Definition 48.7.11 (Likelihood ratio test). The null hypothesis $H_0 : \theta = \theta_0$ is rejected in favour of the alternative hypothesis $H_1 : \theta = \theta_1$ if the likelihood ratio Λ satisfies the following condition:

$$\Lambda(x) = \frac{\mathcal{L}(\theta_0 \mid x)}{\mathcal{L}(\theta_1 \mid x)} \leq \eta\,, \tag{48.81}$$

where $P(\Lambda(x) \leq \eta \mid H_0) = \alpha$.

Remark. In some references the reciprocal of Λ is used as the definition of the likelihood ratio.

Theorem 48.7.12 (Neyman–Pearson lemma). *The likelihood ratio test is the most powerful test at significance level α.*

Definition 48.7.13 (Familywise error). Given a collection of hypothesis tests, the familywise error is defined as the probability of making at least one type-I error.

Construction 48.7.14 (Bonferroni correction). Consider a set of hypotheses $\{H_i\}_{1 \leq i \leq n}$. The higher the number of tests, the higher the chance that by statistical fluctuations at least one of these hypotheses will be rejected. To avoid this problem of multiple comparisons, one can try to control the family-wise error rate, i.e. the probability of falsely rejecting at least one hypothesis. The easiest way to control this error rate is by modifying the individual significance levels:

$$\alpha \longrightarrow \frac{\alpha}{n}\,. \tag{48.82}$$

48.7.2 Comparison tests

Definition 48.7.15 (McNemar test). Consider two models or hypotheses describing a given data set. Construct the contingency table describing the number of true positives and true negatives for both models:

	TP (model 1)	TN (model 1)
TP (model 2)	a	b
TN (model 2)	c	d

(48.83)

The null hypothesis of the McNemar test is that there is no significant difference between the predictive power of the models, i.e. $p_a + p_c = p_a + p_b$ and $p_b + p_d = p_c + p_d$,

where p_i indicates the proportion of class i. In fact it is easy to see that the diagonal values are irrelevant for this hypothesis:

$$H_0 : b = c\,,$$
$$H_1 : b \neq c\,.$$

The test statistic is the McNemar chi-squared statistic:

$$\chi^2 = \frac{(b-c)^2}{b+c}\,. \tag{48.84}$$

When the values of b and c are large enough (> 25), one can approximate this distribution by an ordinary χ^2-distribution with 1 degree of freedom.

Remark 48.7.16 (Edwards correction). It is common to apply a continuity correction (similar to the *Yates correction* for the ordinary chi-squared test):

$$\chi^2 := \frac{(|b-c|-1)^2}{b+c}\,. \tag{48.85}$$

This follows from the fact that for small b, c the exact p-values should be compared with a binomial test which compares b to $b+c$ (note the factor of 2):

$$p = 2\sum_{i=b}^{b+c} \binom{b+c}{i} 0.5^i (1-0.5)^{b+c-i}\,. \tag{48.86}$$

Definition 48.7.17 (Wilcoxon signed-rank test). Consider a paired data sample, i.e. two dependent data samples for which the entries are uniquely paired. This test checks if the population means (more generally, the location parameters) are different.

First, calculate the differences d_i and rank their absolute values (ties are assigned an average rank). Then, calculate the sums of the ranks R_+, R_- for positive and negative differences and take the smallest of these:

$$T := \min(R_+, R_-)\,. \tag{48.87}$$

For small data samples ($n < 25$) one can look up critical values in the literature. For larger data samples one can (approximately) use a standard normal distribution with statistic

$$z := \frac{T - \frac{1}{4}n(n+1)}{\sqrt{\frac{1}{24}n(n+1)(2n+1)}}\,. \tag{48.88}$$

Remark 48.7.18. The main benefit of this test over a signed t-test is that the Wilcoxon test does not require the data samples to be drawn from a normal distribution. However in the case where the assumptions for a paired t-test are met, the t-test is more powerful.

Remark 48.7.19 (Independent samples). There exists a similar rank-based test for unpaired data samples. This is the **Wilcoxon rank-sum test** or **Mann–Whitney U-test**.

Definition 48.7.20 (Friedman test). Consider $k \in \mathbb{N}$ models tested on N data sets. For every data set, one ranks the models according to decreasing performance. For every $i \leq k$, one defines the average rank $R_i = \frac{1}{N}\sum_{j\leq N} r_i^j$, where r_i^j is the rank of the ith model on the jth data set. Under the null hypothesis "all models perform equally well", the average ranks should be the same for all models.

The Friedman statistic

$$\chi_F^2 := \frac{12N}{k(k+1)}\left(\sum_{i\leq k} R_i^2 - \frac{k(k+1)^2}{4}\right) \tag{48.89}$$

follows a χ^2-distribution with $k-1$ degrees of freedom when $N > 10$ and $k > 5$. For smaller values of these parameters, one can look up the exact critical values in the literature.

Remark 48.7.21. It was shown that the original Friedman test is rather conservative and that a better statistic is

$$F := \frac{(N-1)\chi_F^2}{N(k+1) - \chi_F^2}\,. \tag{48.90}$$

This follows an F-distribution with $k-1$ and $(N-1)(k-1)$ degrees of freedom. A further remark is that the (nonparametric) Friedman test is weaker than the (parametric) *repeated-measures ANOVA* whenever the assumptions for the latter hold (similar to the case of the Wilcoxon signed-rank test).

48.7.3 Post-hoc tests

After successfully using one of the multi-model tests from the previous section to reject the null hypothesis of equal performance, one is often interested in exactly which model outperforms the others. For this one can use one of the following pairwise tests.

Definition 48.7.22 (Nemenyi test). Consider the average ranks R_i from the Friedman test. As a test statistic one uses

$$z := \frac{R_i - R_j}{\sqrt{\frac{k(k+1)}{6N}}}\,, \tag{48.91}$$

where k is the number of models and N is the number of data sets. The exact critical values can either be found in the literature or one can approximately use a normal distribution.

Remark 48.7.23 (Bonferroni–Dunn test). If all one wants to do is see if a particular model performs better than a given baseline model, the Nemenyi test is too conservative since it corrects for $k(k-1)/2$ model comparisons instead of $k-1$. Therefore it is better to use a general method to control the family-wise error for multiple measurements. The Bonferroni–Dunn test modifies the Nemenyi test by performing a Bonferroni correction with $n-1$ degrees of freedom.

A more powerful test is given by the following strategy.

Definition 48.7.24 (Holm test). Consider the p-values of the Nemenyi test. Instead of comparing all values to a single Bonferroni-corrected significance, one can use a so-called 'step-down' method. First, one orders the p-values in ascending order and compares the smallest one to $\frac{\alpha}{k-1}$. If this value is significant, i.e. the hypothesis that the associated models perform equally well is rejected, one compares p_2 to $\frac{\alpha}{k-2}$ and so on until one finds a hypothesis that cannot be rejected. All remaining hypotheses are retained as well.

Remark 48.7.25 (Power). It is possible that the post-hoc test fails to report a significant difference even though the Friedman test rejected the null hypothesis. This is a consequence of the lower power of post hoc tests.

48.8 Goodness of fit

Definition 48.8.1 (Akaike information criterion). Consider a model $f(x;\theta)$ with $k \in \mathbb{N}$ parameters fitted to a given data sample and let $\mathcal{L}_0$ be the maximum of the associated likelihood function. The Akaike information criterion is defined a follows:

$$\mathsf{AIC} := 2k - 2\ln(\mathcal{L}_0)\,. \tag{48.92}$$

From this definition it is immediately clear that the AIC rewards goodness-of-fit but penalizes overfitting due to the first term.

This criterion is often useful when trying to select the best model/parameters to describe a certain data set. However, it should be noted that it is not an absolute measure of quality.

48.8.1 χ^2-test

Property 48.8.2. If there are $N - n$ fitted parameters one has:

$$\int_{\chi^2}^{+\infty} f_{\chi^2}(x \mid n)\,dx \approx 1 \implies \begin{cases} \circ \text{ good fit} \\ \circ \text{ errors were overestimated} \\ \circ \text{ selected measurements} \\ \circ \text{ lucky shot} \end{cases} \tag{48.93}$$

Property 48.8.3 (Reduced χ^2). The reduced chi-squared statistic is defined as follows:

$$\chi^2_{\text{red}} := \chi^2/n\,, \tag{48.94}$$

where $n \in \mathbb{N}$ is the number of degrees of freedom. Depending on the value of this statistic, one can draw the following conclusions (under the right assumptions):

- $\chi^2_{\text{red}} \gg 1$: poor modelling,
- $\chi^2_{\text{red}} > 1$: bad modelling or underestimation of the uncertainties,
- $\chi^2_{\text{red}} \approx 1$: good fit, or
- $\chi^2_{\text{red}} < 1$: (improbable) overestimation of the uncertainties.

48.8.2 Runs test

A good χ^2-test does not mean that the fit is good. As mentioned in Property 48.8.2, it is possible that the errors were overestimated. Another condition for a good fit is that the data points vary around the fit, i.e. there are no long sequences of points that lie above/underneath the fit. This condition is tested with a runs test (see Formula 48.8.5 below).

Remark 48.8.4. The χ^2-test and runs test are complementary. The χ^2-test only takes the absolute value of the differences between the fit and data points into account, the runs test only takes the signs of the differences into account.

Formula 48.8.5 (Runs distribution). Let N_+ and N_- denote the number of points above and below the fit. Under the hypothesis that all points were independently drawn from the same distribution, the number of runs is distributed as follows (approximately Gaussian):

$$P(r_{\text{even}}) = 2\frac{\binom{N_+-1}{\frac{r}{2}-1}\binom{N_--1}{\frac{r}{2}-1}}{\binom{N}{N_+}} \qquad P(r_{\text{odd}}) = \frac{\binom{N_+-1}{\frac{r-3}{2}}\binom{N_--1}{\frac{r-1}{2}} + \binom{N_--1}{\frac{r-3}{2}}\binom{N_+-1}{\frac{r-1}{2}}}{\binom{N}{N_+}}, \tag{48.95}$$

where C_k^n is the binomial coefficient $\binom{n}{k}$. The first two moments of this distribution are given by the following formulas:

$$\begin{aligned} \mathsf{E}[r] &= 1 + 2\frac{N_+N_-}{N}, \\ \mathsf{Var}[r] &= 2\frac{N_+N_-}{N}\frac{2N_+N_- - N}{N(N-1)}. \end{aligned} \tag{48.96}$$

Remark 48.8.6. For $r > 15$, the runs distribution approximates a Gaussian distribution.

Chapter 49

Causal Inference

Chapter Contents

49.1 Causal DAGs 928

49.2 Do-calculus 929

49.1 Causal DAGs

Definition 49.1.1 (Bayesian network). Consider a DAG $\mathcal{G} \equiv (V, A)$. A probability distribution P is said to factorize over $\mathcal{G}$ if it can be written as:

$$P(V) = \prod_{x \in V} P\big(x \mid \mathsf{parents}(x)\big) \,. \tag{49.1}$$

Definition 49.1.2 (Mediator). Consider the DAG

$$A \longrightarrow B \longrightarrow C$$

The variable B is called a mediator for A and C.

Definition 49.1.3 (Fork). A DAG of the form

$$A \longleftarrow B \longrightarrow C$$

Definition 49.1.4 (Collider). A DAG of the form

$$A \longrightarrow B \longleftarrow C$$

Definition 49.1.5 (d-separation). Consider a DAG $\mathcal{G} \equiv (V, A)$. Two subsets $V_1, V_2 \subset V$ are said to be d-separated by a set $V_3 \subset V$ whenever all paths from V_1 to V_2 are blocked

by V_3. A path is said to be **blocked** whenever at least one mediator or fork ancestor on the path is contained in V_3 or at least one of the colliders on the path is not contained in V_3.

Property 49.1.6. Consider a Bayesian network represented by the DAG $\mathcal{G} \equiv (V, A)$. If $V_1, V_2 \subset V$ are d-separated by $V_3 \subset V$, then $V_1 \perp\!\!\!\perp V_2 \mid V_3$.

Definition 49.1.7 (Causal path). A path in a DAG such that all arrows point in the same direction.

Combining the notions of d-separation and causality leads to formulas that allow to separate 'spurious' correlations from actual causal relations. Consider for example the collider diagram. As long as the value of B is not fixed, the variables A and C are independent. However, in each stratum of B, the values of A and C might be correlated, although they are not causally related.

49.2 Do-calculus

In this section, $\mathbf{G}$ will always denote a DAG characterizing the underlying causal model.

Axiom 49.1 (Insertion and deletion of variables). If X, Z are d-separated by X, W in $\mathbf{G}_X$, the subgraph obtained from $\mathbf{G}$ by removing all arrows pointing to X, then

$$P\big(Y \mid \mathsf{do}(X), Z, W\big) = P\big(Y \mid \mathsf{do}(X), W\big)\,. \tag{49.2}$$

Axiom 49.2 (Exchange of action and observation). If Y, Z are d-separated by X, W in $\mathbf{G}_X^Z$, the subgraph obtained from $\mathbf{G}$ by removing all arrows pointing to X and all arrows pointing from Z, then

$$P\big(Y \mid \mathsf{do}(X), \mathsf{do}(Z), W\big) = P\big(Y \mid \mathsf{do}(X), Z, W\big)\,. \tag{49.3}$$

Axiom 49.3 (Insertion and deletion of actions). If X, Z are d-separated by X, W in $\mathbf{G}^*$, the subgraph obtained from $\mathbf{G}$ by removing all arrows pointing to X and all arrows pointing to Z that are not ancestors of W in $\mathbf{G}_X$, then

$$P\big(Y \mid \mathsf{do}(X), \mathsf{do}(Z), W\big) = P\big(Y \mid \mathsf{do}(X), W\big)\,. \tag{49.4}$$

@@ COMPLETE @@

Chapter 50

Data Analysis

The main reference for the sections on optimization problems is Shewchuk (1994). For the geometry of clustering methods, see Boissonnat, Nielsen, and Nock (2010). The main references for the section on conformal prediction are Shafer and Vovk (2008); Vovk, Gammerman, and Shafer (2005). Although a part of this chapter is a continuation of Chapter 48, the focus here lies more on the computational aspect of the analysis of large data sets. For this reason, the chapter starts with some sections on applied linear algebra (for a refresher, see Chapter 22).

Chapter Contents

50.1 Data sampling **931**

50.1.1 Inverse CDF sampling 931

50.1.2 Uniform rejection sampling 932

50.1.3 Monte Carlo sampling 932

50.2 Optimization Theory **933**

50.2.1 Linear problems 933

50.2.2 Gradient descent 935

50.2.3 Conjugate gradient 936

50.2.4 Nonlinear conjugate gradients 938

50.2.5 Krylov methods 940

50.3 Constrained optimization **940**

50.3.1 Lagrange multipliers 940

50.3.2 Riemannian gradient descent 942

50.4 Approximation theory **943**

50.4.1 Bayes optimality 943

50.4.2 PAC theory and empirical risk minimization 943

50.5 Classification problems **945**
50.5.1 Clustering 945
50.5.2 Nearest-neighbour search 946
50.6 Support-vector machines **946**
50.6.1 Kernel methods 946
50.6.2 Decision boundaries 948
50.7 Complexity theory **950**
50.7.1 VC dimension 950
50.7.2 Rademacher complexity 952
50.7.3 Relation to Glivenko–Cantelli classes 953
50.8 Time series analysis **953**
50.8.1 Stationarity 953
50.8.2 Correlation 954
50.8.3 Autoregressive models 955
50.8.4 Granger causality 957
50.9 Uncertainty quantification **958**
50.9.1 Prediction regions 958
50.9.2 Classifier calibration 959
50.9.3 Conditionality 960
50.9.4 Distribution shift 961
50.10 Generative models **963**
50.10.1 Normalizing flows 963

50.1 Data sampling

50.1.1 Inverse CDF sampling

Although one of the most straightforward sampling algorithms, this approach makes the strong assumption that the cumulative distribution function (Definition 45.3.4) is invertible.

Method 50.1.1. Sample a point λ uniformly from the unit interval $[0, 1]$. By Property 45.7.1, the CDF $F_X(X)$ is itself uniformly distributed and, hence, the sampled value can serve as the cumulative distribution of the point to be sampled. The new point x' is, hence, simply given by $F_X^{-1}(\lambda)$.

When F_X is discrete, λ might not lie in the image of F_X and the inverse might not admit an algorithmically useful expression, so one should use a different approach. Given a point $x \in \mathbb{R}$ (the sampling algorithm used here is irrelevant and could even be deterministic) and its associated cumulative probability $F_X(x)$, one can sample a new point x' by increasing (or decreasing) x until the unique point x' is found such that

$$F_X(x' - 1) < \lambda \leq F_X(x')\,. \tag{50.1}$$

50.1.2 Uniform rejection sampling

This method again uses the fact that the value of the cumulative distribution function is itself uniformly distributed on the unit interval $[0, 1]$. The CDF does not have to be invertible for this method, but the probability density should be compactly supported.

Method 50.1.2. Consider an interval $[a, b]$ contained in the support of f_X and let q_0 be an upper bound for f_X. Now, uniformly sample a point x from $[a, b]$ and sample a point q uniformly from $[0, q_0]$. If $f_X(x) \geq q$, then x is a good sample. If not, repeat this procedure.

The proof that this algorithm works is quite easy and mainly depends on the fact that

$$\mathsf{Prob}\big(Q \leq f_X(X)\big) = \frac{1}{q_0(b-a)}\,. \tag{50.2}$$

This value is often called the **acceptance probability**. From this expression, it is clear that if one chooses q_0 too large, the acceptance probability becomes very small and the algorithm will take a long time to produce a sample.

50.1.3 Monte Carlo sampling

The general idea of (Markov chain) Monte Carlo methods is to construct a sequence of points such that, starting from a given index, all points represent good samples and such that the sequence forms a Markov chain.

The first Monte Carlo algorithm uses an acceptance threshold.

Method 50.1.3 (Metropolis–Hastings). Assume that a normal distribution $\mathcal{N}(\mu, \sigma^2)$ is given. The first element is given by $x_0 = \mu$. Subsequent points are constructed as follows:

- Sample a point $x' \in \mathbb{R}$ from the normal distribution.
- Calculate the **acceptance ratio**

$$\lambda := \frac{f_X(x')}{f_X(x_{i-1})}\,. \tag{50.3}$$

- If $\lambda \geq 1$, take $x_i = x'$.
- If not, sample a point q uniformly from $[0,1]$. If $\lambda \geq q$, take $x_i = x'$, else take $x_i = x_{i-1}$.[1]

To obtain an efficient algorithm, it is helpful to choose $\mu = \mathsf{E}[X]$ and $\sigma^2 = \mathsf{Var}[X]$. This ensures that the points are sampled in a region that resembles the form of f_X.

This method can also be drastically generalized. First of all, it is possible to replace the normal distribution by any symmetric transition probability:

$$g(x' \mid x) = g(x \mid x')\,. \tag{50.4}$$

The transition probability can even be asymmetric, in which case the acceptance ratio needs to be modified:

$$\lambda := \frac{f_X(x')g(x_{i-1} \mid x')}{f_X(x_{i-1})g(x' \mid x_{i-1})}\,. \tag{50.5}$$

Remark 50.1.4 (Normalization). It is clear from the definition of the acceptance ratio that one does not need f_X to be normalized. This avoids costly calculations of the normalization factor.

50.2 Optimization Theory

50.2.1 Linear problems

Method 50.2.1 (Normal equation). Given the equation

$$Ax = b$$

as in Section 22.4.1, one can try to numerically solve for x by minimizing the ℓ^2-norm $\|Ax - b\|^2$:

$$\hat{x} := \arg\min_x (Ax - b)^T(Ax - b)\,. \tag{50.6}$$

This leads to the so-called normal equation[2]

$$A^TAx = A^Tb\,. \tag{50.7}$$

This can formally be solved by $x = (A^TA)^{-1}A^Tb$, where $(A^TA)^{-1}A^T$ is the pseudoinverse of A.

[1] This step could be merged with the previous one since $\lambda \geq 1$ always implies $\lambda \geq q$.

[2] The name stems from the fact that the equation $A^TAx = A^Tb$ implies that the residual is orthogonal (normal) to the range of A.

Remark 50.2.2. It is easy to see that the above linear problem is obtained when trying to extremize the quadratic form associated to a symmetric matrix.

Method 50.2.3 (Tikhonov regularization). Consider the linear problem

$$Ax = b\,.$$

The most straightforward way to solve for x is the least squares method introduced in Chapter 48, where the solution is (formally) given by the normal equation: $x = (A^TA)^{-1}A^Tb$. However, sometimes it might happen that A is nearly singular (it is said to be **ill-conditioned**). In this case, a regularization term can be added to the minimization problem:

$$\|Ax - b\|^2 + \|\Gamma x\|^2\,, \tag{50.8}$$

where Γ is called the **Tikhonov matrix**. When $\Gamma = \lambda\mathbb{1}$, one speaks of ℓ^2**-regularization**. This regularization technique benefits solutions with smaller norms.

Remark 50.2.4. The ℓ^2-regularization can be generalized by replacing the 2-norm by any p-norm $\|\cdot\|_p$. For $p = 1$ and $p = 2$, the names **lasso** and **ridge** regression are often used, respectively. For general $p \geq 0$, one sometimes speaks of **bridge** regression.

The minimization procedures for $p \leq 1$ have the important property that they not only shrink the coefficients, but even perform feature selection, i.e. some coefficients become identically zero. However, it can be shown that the optimization problem for $p < 1$ is nonconvex and, hence, is harder to solve. In general, it is found that lasso regression gives the best results.

Another benefit of ℓ^p-regularization is that it can be derived from a Bayesian argument. A Gaussian prior $\mathcal{N}(0, \lambda^{-1})$ gives rise to the ℓ^2-regularized cost function as the posterior (log)likelihood. Accordingly, the ℓ^2-regularized linear regressor is equivalent to the maximum-a-posteriori estimator with Gaussian priors. One can obtain ℓ^p-regularization in a similar way by replacing the Gaussian priors with other distributions such as the Laplace distribution for $p = 1$.

Definition 50.2.5 (Multicollinearity). Let $\{X_i\}_{1\leq i\leq n}$ be a finite set of random variables. These random variables are said to be perfectly (multi)collinear if there exists an affine relation between them, i.e. there exist variables $\{\lambda_i\}_{0\leq i\leq n}$ such that

$$\lambda_0 + \lambda_1 X_1 + \cdots + \lambda_n X_n = 0\,. \tag{50.9}$$

The same concept can be applied to data samples. The data is said to be (multi)collinear if the above equation holds for all entries of the data set. However, in this case, one can also define 'near multicollinearity' if the variables X_i are related as above up to some error term ε. If the variance of ε is small, the matrix X^TX might have an ill-conditioned inverse which might render the algorithms unstable.

Definition 50.2.6 (Variance inflation factor). The VIF is an estimate of how much the variance of a coefficient is inflated by multicollinearity. The VIF of a coefficient β_i is defined as follows:

$$\mathsf{VIF}_i := \frac{1}{1 - R_i^2}\,, \tag{50.10}$$

where R_i^2 is the R^2-value obtained after regressing the predictor $\widehat{X}_i$ on all other predictors. The rule of thumb is that $\mathsf{VIF} \geq 10$ implies that a significant amount of multicollinearity is present in the model.

50.2.2 Gradient descent

The gradient descent algorithm is first introduced in the case of quadratic forms.

Method 50.2.7 (Steepest descent). Consider the quadratic form

$$f(x) = \frac{1}{2}x^T A x - b^T x + c\,. \tag{50.11}$$

Assume that A is symmetric and positive definite, such that $Ax = b$ gives the minimum of f. Like most recursive algorithms, gradient descent starts from an arbitrary guess x_0. It then takes a step in the direction of steepest descent (or largest gradient), i.e. in the direction opposite to $\nabla f(x_0) = Ax_0 - b =: -r_0$:

$$x_{i+1} := x_i + \alpha r_i\,. \tag{50.12}$$

The quantities r_i are called the **residuals**. This procedure is repeated until convergence, i.e. until the residual vanishes up to a fixed numerical tolerance.

A naive gradient descent method would require to fine-tune the step size α. However, a more efficient method is given by the **line search algorithm**, where the value of α is optimized in every step as to minimize f along the line defined by r_i. A standard calculus argument leads to the following form of the step size:

$$\alpha_i = \frac{r_i^T r_i}{r_i^T A r_i}\,. \tag{50.13}$$

This choice forces the descent direction to be orthogonal to the previous one, since

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} f(x_i) = -\nabla f(x_i) \cdot \nabla f(x_{i-1})\,. \tag{50.14}$$

As a consequence, this minimization scheme often results in a chaotic zigzag trajectory through the configuration space. The higher the **condition number** $\kappa = \frac{|\lambda_{\max}|}{|\lambda_{\min}|}$, the worse the zigzag motion will be. A very narrow valley (or some higher-dimensional analogue) will make the trajectory bounce back and forth between the walls, instead of moving towards the minimum.

50.2.3 Conjugate gradient

As noted in the previous section, a common problem with gradient descent is that the direction of steepest descent is often not the same as the direction pointing to the optimal solution and, hence, convergence might only be achieved after many iterations.

A simple solution can be obtained by considering multiple orthogonal directions and taking a suitable step once in every direction. This way, one obtains an algorithm that converges in n steps, where $n \in \mathbb{N}$ is the dimension of the coefficient matrix A. By requiring that the error at step $i+1$ is orthogonal to the direction d_i, it is assured that no direction is used twice. However, the main problem with this idea is that the exact error e_i is not known and, hence, one cannot calculate the required steps.

By modifying the orthogonality condition, one can avoid this problem. This is the idea behind conjugate direction methods.

Definition 50.2.8 (Conjugate vectors). Consider a symmetric, positive-definite matrix A. Any such matrix induces an inner product as follows:

$$\langle v \mid w \rangle_A := v^T A w \,. \tag{50.15}$$

Two vectors v, w are said to be (A-)conjugate if they are orthogonal with respect to this inner product. The general approach to obtain a basis of A-conjugate vectors is a modified version of the Gram–Schmidt procedure (Method 22.3.14) where the ordinary Euclidean inner product is replaced by (50.15). This modification is called the **Arnoldi method**.

By taking the input vectors of the Arnoldi method to be the residuals r_i, one obtains the **conjugate gradient** (CG) algorithm. It is interesting to note that the residuals themselves satisfy a recursion relation:

$$r_{i+1} = r_i - \alpha_i A d_i \,, \tag{50.16}$$

where the step size α_i is defined similar to the step size for ordinary steepest descent:

$$\alpha_i = \frac{d_i^T r_i}{d_i^T A d_i} \,. \tag{50.17}$$

Since the directions are constructed using the residuals, they span the same subspace. By denoting the subspace spanned by the first i directions by $\mathcal{D}_i$, the relation $r_{i+1} \in \mathcal{D}_i + A d_i$ leads to the following expression because of the above recursion relation:

$$\mathcal{D}_i = \operatorname{span}\left\{ r_0, A r_0, \dots, A^{i-1} r_0 \right\} . \tag{50.18}$$

Because of their prominence in the literature on numeric optimization techniques, these subspaces have earned their own name.

Definition 50.2.9 (Krylov subspace). A vector space $\mathcal{K}$ of the form

$$\mathcal{K} := \mathsf{span}\{v, Av, \dots, A^n v\} \tag{50.19}$$

for some matrix A, vector v and natural number $n \in \mathbb{N}$. Given such an A and v, one often denotes the associated Krylov subspace of dimension n by $\mathcal{K}_n(A, v)$.

The fact that the spaces $\mathcal{D}_i$ are Krylov spaces also has an import implication for the numerical complexity of the CG algorithm. The residual r_{i+1} can be shown to be orthogonal to the space $\mathcal{D}_{i+1}$ (this is generally called the **Galerkin condition**). But, since $A\mathcal{D}_i \subset \mathcal{D}_{i+1}$, this also implies that r_{i+1} is A-conjugate to $\mathcal{D}_i$. It follows, that the only relevant contribution in the Arnoldi method is given by the last direction d_i. This reduces the complexity (both timewise and memorywise) per iteration from $O(n^2)$ to $O(n)$.

The steps in the CG algorithm are summarized below.

Method 50.2.10 (Conjugate gradient). Let x_0 be the initial guess with the associated residual $r_0 := b - Ax_0$ acting as the first direction vector d_0. The following scheme gives an iterative n-step ($n \in \mathbb{N}$ being the dimension of the coefficient matrix A) algorithm to obtain the solution to $Ax = b$:

$$\begin{aligned}
\alpha_i &:= \frac{r_i^T r_i}{d_i^T A d_i} \\
x_{i+1} &:= x_i + \alpha_i d_i \\
r_{i+1} &:= r_i - \alpha_i A d_i && (50.20) \\
d_{i+1} &:= r_{i+1} + \frac{r_{i+1}^T r_{i+1}}{r_i^T r_i} d_i\,. && (50.21)
\end{aligned}$$

Remark 50.2.11. In exact arithmetic, the above optimization scheme would result in an exact solution after $n \in \mathbb{N}$ iterations (in fact, the number of iterations is bounded by the number of distinct eigenvalues of A). However, in real life, one is not working in exact arithmetic and the occurrence of floating-point errors has tio be taken into account. These not only ruin the accuracy of the residual recursion relation (50.20), but, more importantly[3], they might result in the search directions not being A-conjugate.

Now, what about general coefficient matrices A such as those resulting in under- or overdetermined systems? For asymmetric or nondefinite square matrices, one can still solve the normal equation using the same methods (Method 50.2.1), since $A^T A$ is both symmetric and positive definite. For underdetermined systems, an exact solution does not always exact, but the numerical methods will always be able to find a solution that minimizes the ℓ^2-error. For overdetermined systems, $A^T A$ will be nonsingular and the

[3]The residual problem can be solved by computing the residual 'exactly', i.e. by the formula $r_i = b - Ax_i$, every $k \in \mathbb{N}$ iterations.

numerical methods can find an exact solution. However, the condition number of A^TA is the square of that of A and, hence, the algorithms will convergence much slower.

A different approach exists where the CG algorithm is not applied to the matrix A^TA, but the individual matrices A, A^T are used directly. This way, not one Krylov space is generated, but two dual 'copies' are constructed.

$$\begin{aligned} \mathcal{D}_i &:= \mathsf{span}\left\{r_0, Ar_0, \dots, A^{i-1}r_0\right\}, \\ \widetilde{\mathcal{D}}_i &:= \mathsf{span}\left\{\tilde{r}_0, A^T\tilde{r}_0, \dots, (A^T)^{i-1}\tilde{r}_0\right\}, \end{aligned} \tag{50.22}$$

where $\tilde{r}_0$ does not have to be related to r_0. In this case, there are two Galerkin conditions: $r_i \perp \mathcal{D}_i$ and $\tilde{r}_i \perp \widetilde{\mathcal{D}}_i$, where only the first one is relevant. The residuals form biorthogonal bases of the Krylov subspaces:

$$\langle r_i \mid r_j \rangle = \|r_i\|^2 \delta_{ij}\,. \tag{50.23}$$

As a consequence, the search directions also form biconjugate bases:

$$\langle d_i \mid d_j \rangle_A = \|d_i\|_A^2 \delta_{ij}\,. \tag{50.24}$$

50.2.4 Nonlinear conjugate gradients

Of course, many real-world applications are determined by nonlinear equations and, hence, it would be pleasant if one could salvage some of the above ideas, even when linear algebra is not the natural language. The main requirement would be that one can calculate the gradient of the function to be minimized.

At the level of the implementation, the structure of the algorithm remains more or less the same. What does change is the form of the Arnoldi method, in particular, the prefactor in Eq. (50.21). For linear CG, there are multiple equivalent formulas, but for nonlinear CG these do not lead to the same algorithm. The two most common choices are given below.

Method 50.2.12 (Nonlinear CG). Since there is no linear equation related to the minimization problem, the residuals are always defined as $r_i := -\nabla f(x_i)$. The algorithm consists of the following iterations:

$$\begin{aligned} \alpha_i &:= \arg\min_{\alpha} f(x_i + \alpha d_i) \\ x_{i+1} &:= x_i + \alpha_i d_i \\ r_{i+1} &:= -\nabla f(x_i) \\ d_{i+1} &:= r_{i+1} + \beta_{i+1} d_i\,, \end{aligned} \tag{50.25}$$

where β_{i+1} is computed by one of the following formulas:

- **Fletcher–Reeves formula**:

$$\beta_{i+1} := \frac{r_{i+1}^T r_{i+1}}{r_i^T r_i} . \tag{50.26}$$

- **Polak–Ribière formula**:

$$\beta_{i+1} := \max\left\{ \frac{r_{i+1}^T (r_{i+1} - r_i)}{r_i^T r_i}, 0 \right\} . \tag{50.27}$$

Some general remarks have to be made concerning the nonlinear CG algorithm.

Remark 50.2.13. As was already mentioned for the linear version, floating-point errors might lead to a loss of conjugacy. For the nonlinear extension, this becomes worse. The more f deviates from a quadratic function, the faster conjugacy is lost (for quadratic formulas, the Hessian is exactly the matrix A, but for higher-degree functions, the Hessian varies from point to point). Another problem, one that did not occur for quadratic functions, is that nonlinear functions might have local minima. The CG method does not care about the difference between local and global minima and, hence, it will not necessarily converge to the global minimum. A last remark concerns the fact that there is no theoretical guarantee that the method will converge in n steps. Since the Gram–Schmidt procedure can only construct n conjugate vectors, the simplest solution is to perform a restart of the algorithm every n iterations.[4]

For linear CG, a simple formula for finding the optimal value of α_i was obtained. However, for nonlinear CG, one cannot solve Eq. (50.25) as easily. The main idea, i.e. that ∇f should be orthogonal to the previous search direction, is still valid. Here, only the **Newton–Raphson approach** is considered:[5]

$$\alpha_i = \frac{\nabla f(x_i)^T d_i}{d_i^T \operatorname{Hess} f(x_i) d_i} . \tag{50.28}$$

To obtain the optimal value for α, one should iteratively apply the Newton–Raphson method in every CG iteration. If the action of the Hessian $\operatorname{Hess} f$ on d_i cannot be simplified, i.e. if the full Hessian has to be computed in every iteration, this can lead to considerable computational overhead. The general rule of thumb is to perform only a few Newton–Raphson iterations and obtain a less accurate but more efficient algorithm. To make sure that the search direction is indeed a direction of descent (and not one of ascent), one can check that $r^T d \geq 0$ and restart the procedure if not.

[4] The max operation in Eq. (50.27) is already a form of restarting, due to the fact that the Polak–Ribière version of nonlinear CG sometimes results in cyclic behaviour.

[5] Another common method is the *secant method*.

50.2.5 Krylov methods

Generally, one starts from an iterative, fixed-point based technique to solve the linear equation $Ax = b$ as before, i.e. one iterates $x_{i+1} = b + (\mathbb{1} - A)x_i$. Using the residuals $r_i = b - Ax_i$, this can be rewritten as

$$x_i = x_0 + \sum_{k=0}^{i-1} r_k = x_0 + \sum_{k=0}^{i-1} (\mathbb{1} - A)^k r_0 \,. \tag{50.29}$$

It is clear that this results in $x_i - x_0 \in \mathcal{K}_i(A, r_0)$. The main idea is, then, to find optimal degree-k polynomials P_k such that

$$x_i - x_0 = \sum_{k=0}^{i-1} P_k(A) r_0 \,. \tag{50.30}$$

Method 50.2.14 (Jacobi method). Consider a linear problem $Ax = b$ where A has spectral radius less than 1. First, decompose A as the sum of a diagonal matrix D and and a matrix E with all diagonal elements equal to 0. If one assumes that D is invertible, the following recursive scheme is obtained:

$$x_{i+1} := D^{-1}(b - Ex_i) \,. \tag{50.31}$$

A sufficient condition for convergence is strict diagonal dominance, i.e. $|D_{ii}| > \sum_{j \neq i} |E_{ij}|$.

@@ COMPLETE (add e.g. Lanczos) @@

50.3 Constrained optimization

50.3.1 Lagrange multipliers

A common generalization of the above optimization problems is the addition of constraints involving equalities:

$$\arg\min_x f(x) \qquad \text{such that} \qquad g_i(x) = 0 \qquad \forall 1 \leq i \leq n \,. \tag{50.32}$$

The general approach to solving such constrained problems is by extending the loss function.

Method 50.3.1 (Lagrange multipliers). Given a constrained optimization problem of the form (50.32), one can construct the enhanced loss function

$$\mathcal{L}(x, \lambda_1, \dots, \lambda_n) := f(x) + \sum_{i=1}^{n} \lambda_i g_i(x) \,. \tag{50.33}$$

The solution to the original problem is obtained by extremizing this loss with respect to x and the Lagrange multipliers λ_i (as usual, this might fail globally for nonconvex

problems):

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial x} = 0 \\ \\ \dfrac{\partial \mathcal{L}}{\partial \lambda_i} = 0 \quad \forall 1 \leq i \leq n\,. \end{cases} \tag{50.34}$$

The situation becomes even more interesting when one also allows constraints involving inequalities:

$$\arg\min_x f(x) \qquad \text{such that} \qquad \begin{cases} g_i(x) = 0 & \forall 1 \leq i \leq m\,, \\ h_j(x) \leq 0 & \forall 1 \leq j \leq n\,. \end{cases} \tag{50.35}$$

Problems of this form are called **primal optimization problems**. By defining an enhanced loss using Lagrange multipliers as before,

$$\mathcal{L}(x, \alpha, \beta) := f(x) + \sum_{i=1}^{m} \alpha_i g_i(x) + \sum_{i=1}^{n} \beta_i h_i(x)\,, \tag{50.36}$$

it is not hard to see that

$$\max_{\alpha,\beta;\beta_j \geq 0} \mathcal{L}(x, \alpha, \beta) = \begin{cases} +\infty & \text{if a constraint is violated,} \\ f(x) & \text{if all constraints are satisfied.} \end{cases} \tag{50.37}$$

Definition 50.3.2 (Primal optimization problem). Denote the maximum of $\mathcal{L}(x, \alpha, \beta)$ by $\theta_P(x)$.

$$p^* := \min_x \theta_P(x) = \min_x \max_{\alpha,\beta|\beta_i \geq 0} \mathcal{L}(x, \alpha, \beta)\,. \tag{50.38}$$

By interchanging the max and min operators in the primal formulation, another problem is obtained.

Definition 50.3.3 (Dual optimization problem).

$$d^* := \max_{\alpha,\beta;\beta_i \geq 0} \theta_D(\alpha, \beta) = \max_{\alpha,\beta|\beta_i \geq 0} \min_x \mathcal{L}(x, \alpha, \beta)\,. \tag{50.39}$$

From basic calculus, it is known that $\max\min \leq \min\max$ and, hence, that $d^* \leq p^*$. The difference $p^* - d^*$ is called the **duality gap** and, if $d^* = p^*$, one says that **strong duality** holds. The real question then becomes: *"When does strong duality hold?"*.

Definition 50.3.4 (Slater conditions). Consider a convex optimization problem, i.e. a problem of the form of Eq. (50.35) where f is convex, the g_i are convex and the h_j are affine. This problem is said to satisfy the Slater condition(s) if there exists an x that is strictly **feasible**, i.e. $h_j(x) < 0$ for all $1 \leq j \leq n$.

Property 50.3.5 (Strong duality). If a convex problem satisfies the Slater conditions, strong duality holds. The solutions x and (α, β) that attain this duality are called primal and dual optima, respectively.

The following property gives a set of sufficient conditions.

Property 50.3.6 (Karush–Kuhn–Tucker conditions). If there exist x, α and β such that strong duality holds, the following conditions are satisfied:

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial x} = 0 \\ \\ \dfrac{\partial \mathcal{L}}{\partial \alpha_i} = 0 \end{cases} \quad \forall 1 \leq i \leq m \qquad \text{and} \qquad \begin{cases} \beta_j h_j(x) = 0 \\ h_j(x) \leq 0 \\ \beta_j \geq 0 \end{cases} \quad \forall 1 \leq j \leq n. \tag{50.40}$$

Conversely, if there exists values x, α and β that satisfy the KKT conditions, they give strongly dual solutions for the primal and dual problems.

Remark 50.3.7 (Complementary slackness). The third equation in the KKT conditions has an important implication. It says that, if there is an index j such that the constraint h_j is not **active**, i.e. $h_j(x) < 0$, the associated Lagrange multiplier is 0 and, conversely, if there is an index j such that the Lagrange multiplier $\beta_j > 0$, the constraint h_j is active.

Remark 50.3.8. It is not hard to see that the KKT conditions reduce to the conditions for Lagrange multipliers when all h_j are identically 0. For this reason, the quantities α and β are called the **KKT multipliers**.

50.3.2 Riemannian gradient descent

In many situations, the full parameter space of an optimization problem is constrained in such a way that the resulting admissible subset admits the structure of a smooth manifold and, in particular, that of a Riemannian manifold. When trying to extend gradient descent algorithms to this setting, one has to take into account that most manifolds are not linear spaces and, hence, that linear updates will often lead outside the manifold.

The first point that has to be treated is the occurrence of the gradient in these algorithms. In ordinary Euclidean space, one simply takes the gradient to be the vector of partial derivatives. However, on general smooth manifolds, this object is actually given by the de Rham differential, which acts on covariant vectors. However, a short proof shows that the Riemannian gradient from Remark 36.1.9 actually gives the direction of steepest ascent. So, even on Riemannian manifolds, the (de Rham) gradient is the correct object to work with.

@@ COMPLETE (also add references) @@

50.4 Approximation theory

50.4.1 Bayes optimality

Definition 50.4.1 (Bayes risk). Let $\mathcal{H} \subseteq \mathcal{Y}^{\mathcal{X}}$ be a collection of models. The Bayes risk (with respect to $\mathcal{H}$) is the minimal risk over all models:

$$R^* := \inf_{f \in \mathcal{H}} R(f)\,. \tag{50.41}$$

'The' Bayes risk is the Bayes risk with respect to $\mathcal{Y}^{\mathcal{X}}$.

Definition 50.4.2 (Bayes classifier). Given a joint probability distribution P over the instance space $\mathcal{X} \times \mathcal{Y}$ and a loss function $l : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$, the pointwise Bayes predictor is defined as follows:

$$f^* : x \mapsto \operatorname*{arg\,min}_{y \in \mathcal{Y}} \int_{\mathcal{Y}} l(y, y')\, dP(y' \mid x)\,. \tag{50.42}$$

Property 50.4.3 (Bayes optimality). The risk of the pointwise Bayes predictor is minimal, i.e. $R(f^*) = R^*$. In practice, however, one cannot achieve Bayes optimality (through the pointwise Bayes predictor), since this would require the knowledge of the distribution.

Definition 50.4.4 (Approximation error). Given a data set $\mathcal{D} \subset \mathcal{X} \times \mathcal{Y}$, the empirical risk R_{emp} is defined as follows:

$$R_{\text{emp}} : \hat{y} \mapsto \frac{1}{|\mathcal{D}|} \sum_{(x,y) \in \mathcal{D}} l(\hat{y}(x), y). \tag{50.43}$$

Consider a hypothesis space $\mathcal{H} \subseteq \mathcal{Y}^{\mathcal{X}}$ (e.g. determined by a choice of model architecture). The minimizer of the true risk in $\mathcal{H}$ is denoted by h^*, while the empirical risk minimizer is denoted by $\hat{h}$. The **approximation error/uncertainty** is defined as the difference $R(\hat{h}) - R(h^*)$, while the **model uncertainty** is defined as the difference $R(h^*) - R^*$. The total error made by $\hat{h}$ (with respect to the Bayes optimum) is then simply the sum of these two.

50.4.2 PAC theory and empirical risk minimization

Property 50.4.5. Note that, by the (strong) law of large numbers (Theorem 48.1.4), the empirical risk converges to the true risk almost surely whenever the data points are sampled i.i.d. However, in practice, the true question becomes how fast this convergence happens.

Definition 50.4.6 (Probably approximately correct). A model $\hat{f}$ is said to be ε-accurate at the confidence level $1 - \delta$ (with respect to the hypothesis space $\mathcal{H}$), or simply (ε, δ)-PAC, if it satisfies

$$\text{Prob}\Big(R(\hat{f}) - \inf_{f \in \mathcal{H}} R(f) > \varepsilon\Big) < \delta\,. \tag{50.44}$$

Corollary 50.4.7. If one can prove a PAC bound, one gets the following bound for free:

$$R(\hat{f}) \leq \inf_{f \in \mathcal{H}} R(f) + 2\varepsilon \tag{50.45}$$

with probability $1 - \delta$, i.e. with high probability the true risk for the minimizer is only slightly higher than the empirical risk and, hence, the minimizer is a reasonably good estimate.

Property 50.4.8 (Bounded losses). Consider a hypothesis space $\mathcal{H}$ with a bounded loss function. The probability that, for a given number of data points $n \in \mathbb{N}_0$, their exists at least one model $f \in \mathcal{H}$ for which the empirical risk deviates significantly from the true risk is bounded as follows:

$$\begin{aligned}
\mathsf{Prob}\big(\exists f \in \mathcal{H} : |R_{\text{emp}}(f) - R(f)| \geq \varepsilon\big) &= \mathsf{Prob}\big(\sup_{f \in \mathcal{H}} |R_{\text{emp}}(f) - R(f)| \geq \varepsilon\big) \\
&\leq \sum_{f \in \mathcal{H}} \mathsf{Prob}\big(|R_{\text{emp}}(f) - R(f)| \geq \varepsilon\big) \\
&\leq \sum_{f \in \mathcal{H}} 2e^{-2n\varepsilon^2} \\
&= 2|\mathcal{H}|e^{-2n\varepsilon^2} ,
\end{aligned}$$

where the second inequality comes from Hoeffding's inequality 45.3.30 (since the empirical risk is an average). In combination with a similar expression for the one-sided tails, using the one-sided Hoeffding inequality, the PAC bound becomes

$$R(f) \leq R_{\text{emp}}(f) + \sqrt{\frac{\ln(|\mathcal{H}|) + \ln(1/\delta)}{2n}} \tag{50.46}$$

with probability at least $1 - \delta$. Note that this bound is distribution-free, i.e. it does not depend on the data-generating distribution.

This bound on the empirical risk also implies a bound on the expected risk of the minimizer $\hat{f}$:

$$\mathsf{E}\big[R(\hat{f})\big] \leq \inf_{f \in \mathcal{H}} R(f) + \sqrt{\frac{\ln(|\mathcal{H}|) + \ln(1/\delta)}{2n}} + \delta , \tag{50.47}$$

where $\delta > 0$ is now arbitrary.

The issue with the above two bounds is that they only apply to the case where the hypothesis space $\mathcal{H}$ is finite. To obtain useful bounds for countable hypothesis spaces, a new tool is required.

Definition 50.4.9 (Complexity regularizer). A function $c : \mathcal{X} \to \mathbb{R}^+$ such that

$$\sum_{x \in \mathcal{X}} e^{-c(x)} \leq 1 . \tag{50.48}$$

A simple example would be the log-probabilities $c(x) := \ln(P(x))$ when a probability distribution P on $\mathcal{X}$ is given.

Using a complexity regularizer on $\mathcal{H}$, the following risk bound can be obtained with probability $1 - \delta$:

$$R(f) \leq R_{\text{emp}}(f) + \sqrt{\frac{c(f) + \ln(1/\delta)}{2n}}\,. \tag{50.49}$$

50.5 Classification problems

50.5.1 Clustering

Probably the most well-known and simplest algorithm for clustering in the unsupervised setting is the k-means algorithm.

Method 50.5.1 (k-means algorithm). Assume that an unlabelled dataset $\mathcal{D} \subset \mathbb{R}^n$ is given. For every integer $k \in \mathbb{N}$, usually satisfying $k \ll |\mathcal{D}|$, and any choice of k distinct **centroids** $\{c_i \in \mathbb{R}^n\}_{i \leq k}$, the k-means algorithm is defined through the following iterative scheme:

1. To every point $d \in \mathcal{D}$, assign a cluster C_i based on the following criterion:

$$i = \operatorname*{arg\,min}_{j \leq k} \|d - c_j\|^2\,. \tag{50.50}$$

2. Update the centroids c_i to represent the center of mass of the associated cluster C_i:

$$c_i \longleftarrow \frac{1}{|C_i|} \sum_{d \in C_i} d\,. \tag{50.51}$$

This algorithm optimizes the following global cost function with respect to the centroids c_i:

$$\mathcal{L}_{k\text{-means}}(c_1, \dots, c_k) = \sum_{i=1}^{k} \sum_{d \in C_i} \|d - c_i\|^2\,. \tag{50.52}$$

Given the above idea, one could ask for a more general algorithm where clustering is performed with respect to a divergence function (Definition 47.1.7). In the case of Bregman divergences (Definition 47.1.10), it can be shown that all one needs to do is replace the Euclidean distance by the divergence D_f.

Property 50.5.2 (Centroid position). Let D_f be a Bregman divergence. The minimizer

$$\operatorname*{arg\,min}_{\kappa} \sum_{i=1}^{k} D_f(x_i \,\|\, \kappa) \tag{50.53}$$

is given by the arithmetic average

$$\kappa = \frac{1}{k}\sum_{i=1}^{k} x_i \,. \tag{50.54}$$

If, instead of a cluster $C = \{x_i \in \mathbb{R}^n\}_{i\leq k}$, one is given a probability distribution p, one simply has to replace the arithmetic average by the expectation value with respect to p. Furthermore, it can be shown that, for any Bregman divergence, the k-means algorithm always converges in a finite number of steps (however, the clustering is not necessarily optimal).

The cluster boundaries $H(c_1, c_2) = \{x \in \mathbb{R}^n \mid D_f(x \,\|\, c_1) = D_f(x \,\|\, c_2)\}$ admit a simple geometric construction.

Property 50.5.3 (Cluster boundaries). Let D_f be a Bregman divergence and consider the k-means problem associated to D_f for $k = 2$ (higher-dimensional problems can be treated similarly). The boundary $H(c_1, c_2)$ is exactly the geodesic hypersurface orthogonal to the dual geodesic connecting c_1 and c_2. This partitioning of the data manifold is a generalization of *Voronoi diagrams* to (Bregman) divergences.[6]

50.5.2 Nearest-neighbour search

@@ COMPLETE @@

50.6 Support-vector machines

50.6.1 Kernel methods

This section will introduce the mathematics of kernel methods. This mainly involves the language of Hilbert spaces (see Chapter 25 for a refresher).

Definition 50.6.1 (Kernel[7]). A function $k : X \times X \to \mathbb{C}$ that is (conjugate) symmetric and for which the **Gram matrix** $K_{ij} := K(x_i, x_j)$ is positive definite for all $n \in \mathbb{N}$ and $\{x_i \in X\}_{i\leq n}$.

Definition 50.6.2 (Reproducing kernel Hilbert space). A Hilbert space $\mathcal{H} \subset \mathsf{Map}(X, \mathbb{C})$ of functions on a set X for which all evaluation functionals $\delta_x : f \mapsto f(x)$ are bounded (or continuous by Property 26.1.8). Reproducing kernel Hilbert spaces are often abbreviated as **RKHS**s.

Using Riesz's representation theorem 25.2.7, one can express every evaluation functional δ_x on $\mathcal{H}$ as a function $K_x \in \mathcal{H}$. This allows for the introduction of a kernel on

[6] See Boissonnat et al. (2010) for more information. This is also introduced in Amari (2016), but there the author has confusingly interchanged the affine and dual coordinates.

[7] Also called a **Mercer kernel**. See Mercer's theorem below for more information.

X.

Definition 50.6.3 (Reproducing kernel). Let $\mathcal{H}$ be an RKHS on a set X. The (reproducing) kernel k on X is defined as follows:

$$k(x,y) := \delta_x(K_y) \overset{\text{Riesz}}{=} \langle K_x \mid K_y \rangle_{\mathcal{H}} . \tag{50.55}$$

Because k is given by an inner product, it is not hard to see that the reproducing kernel is a kernel in the sense of Definition 50.6.1.

Starting from a kernel, one can also characterize an RKHS as follows.

Alternative Definition 50.6.4 (RKHS). A Hilbert space $\mathcal{H} \subset \mathsf{Map}(X, \mathbb{C})$ of functions over a set X such that there exists a kernel k on X with the following properties:

1. **Reproducing property**: For all $x \in X$ and $f \in \mathcal{H}$, the evaluation functional δ_x satisfies $\delta_x(f) = \langle k(\cdot, x) \mid f \rangle_{\mathcal{H}}$.
2. **Density**: The span of $\{k(\cdot, x) \mid x \in X\}$ is dense in $\mathcal{H}$.

The density property is often replaced by the property that $k(\cdot, x) \in \mathcal{H}$ for all $x \in X$.

Property 50.6.5 (Convergence). In an RKHS, convergence in norm implies pointwise convergence.

Theorem 50.6.6 (Moore–Aronszajn). *There exists a bijection between RKHSs and kernels.*

Proof. One direction of the theorem is, as mentioned before, rather simple to see. The other direction is constructive. Given a kernel k, one defines the function $K_x := k(\cdot, x)$ for all $x \in X$. The RKHS is then constructed as the Hilbert completion of $\mathsf{span}\{K_x \mid x \in X\}$, *where the inner product is defined as follows*

$$\left\langle \sum_{x \in X} a_x K_x \,\middle|\, \sum_{y \in X} b_y K_y \right\rangle := \sum_{x,y \in X} \overline{a_x} b_y k(x,y) . \tag{50.56}$$

□

Formula 50.6.7. Let $\mathcal{H}$ be an RKHS with kernel k. If $\{e_i\}_{i \leq \dim(\mathcal{H})}$ is an orthonormal basis for $\mathcal{H}$, then

$$k(x,y) = \sum_{i=1}^{\dim(\mathcal{H})} e_i(x) \overline{e_i(y)} . \tag{50.57}$$

Remark 50.6.8. Note that one can use different conventions in the above definitions, e.g. the definition $k(x,y) = \langle K_y \mid K_x \rangle_{\mathcal{H}}$ is also valid.

Theorem 50.6.9 (Mercer). *Let X be a finite measure space and consider a (conjugate) symmetric function $k \in L^2(X \times X, \mathbb{C})$. If k satisfies the* ***Mercer condition***

$$\iint_{X\times X} k(x,y)\overline{f(x)}f(y)\,dxdy \geq 0 \tag{50.58}$$

for all $f \in L^2(X,\mathbb{C})$, the Hilbert–Schmidt operator

$$T_k : L^2(X,\mathbb{C}) \to L^2(X,\mathbb{C}) : f \mapsto \int_X k(\cdot,x)f(x)\,dx \tag{50.59}$$

admits a countable orthonormal basis $\{e_i\}_{i\in\mathbb{N}}$ with nonnegative eigenvalues $\{\lambda_i\}_{i\in\mathbb{N}}$ such that

$$k(x,y) = \sum_{i=1}^{+\infty} \lambda_i e_i(x)\overline{e_i(y)}\,. \tag{50.60}$$

Theorem 50.6.10 (Bochner). *A continuous function satisfies the Mercer condition if and only if it is a kernel.*

Alternative Definition 50.6.11 (Kernel). Consider a set X. A function $k : X\times X \to \mathbb{C}$ is called a (Mercer) kernel on X if there exists a Hilbert space $\mathcal{H}$ together with a function $\phi : X \to \mathcal{H}$ such that

$$k(x,y) = \langle \phi(x) \mid \phi(y)\rangle_{\mathcal{H}}\,. \tag{50.61}$$

When using Mercer's theorem, the feature maps are given by

$$\phi_i : x \mapsto \sqrt{\lambda_i}e_i(x)\,. \tag{50.62}$$

Remark 50.6.12. The kernel expressions in the Mercer and Moore–Aronszajn theorems are related by the fact that the RKHSs induced by kernels satisfying the assumptions of the Mercer theorem are of the form

$$\mathcal{H} = \left\{ f \in L^2(X,\mathbb{C}) \,\middle|\, \sum_{i=1}^{+\infty} \frac{\langle f, e_i\rangle_{L^2}^2}{\lambda_i} < +\infty \right\}. \tag{50.63}$$

Remark 50.6.13 (Vector-valued functions). Much of this section can be generalized to the setting of vector-valued functions $f : X \to \mathbb{C}^d$. In this case, the kernels $k : X\times X \to \mathbb{C}$ are generalized to a matrix-valued functions $k : X \times X \to \mathbb{C}^{d\times d}$.

50.6.2 Decision boundaries

Consider a linear classification model $y = w^T x + b$. The object x_i is said to belong to the positive (resp. negative) class if $y > 0$ (resp. $y < 0$). This is implemented by the sign activation function

$$\operatorname{sgn}(y) = \begin{cases} 1 & \text{if } y > 0 \\ -1 & \text{if } y < 0 \end{cases} \tag{50.64}$$

to the linear model. The **decision boundary** $y = 0$, where the decision becomes ambiguous, forms a hyperplane in the feature space. However, it should be clear that, in generic situations, there are multiple hyperplanes that can separate the two classes for a finite number of data points. The problem then becomes to obtain the hyperplane with the maximal separation, i.e. the hyperplane for which the distance to the nearest data point is maximal.

The unit vector $\frac{w}{\|w\|}$ defines the normal to the hyperplane and, therefore, one can obtain the distance $d_w(x)$ from a data point x to the decision boundary by projecting onto this unit vector. The point $x - d_w(x)\frac{w}{\|w\|}$ is an element of the decision boundary and, hence, satisfies the hyperplane equation. Rewriting this formula gives an expression for the distance

$$d_w(x) = \frac{w^T x + b}{\|w\|} \,. \tag{50.65}$$

To account for the direction of the vector, this number should be multiplied by the class $\mathrm{sgn}(y) = \pm 1$. This result is called the **geometric margin**:

$$\gamma(x) := \mathrm{sgn}(y) d_w(x) \,. \tag{50.66}$$

The numerator in the geometric margin is called the **functional margin**. The geometric margin is preferable since it is invariant under simultaneous scale transformations of the parameters w, b.

The optimization objective now becomes

$$\max_w \frac{\gamma}{\|w\|} \qquad \text{such that} \qquad y_i(w^T x_i + b) \geq \gamma \|w\| \qquad \forall 1 \leq i \leq n \,, \tag{50.67}$$

where $\gamma = \min_{i \in \{1,\ldots,n\}} \gamma(x_i)$ for x_i ranging over the training set. The problem is formulated in terms of the functional margin $\gamma\|w\|$ to avoid the nonconvex constraint $\|w\| = 1$. This allows the application of the Slater conditions for strong duality. Since the geometric margin is invariant under scale transformations, one can, without loss of generality, work with the assumption $\gamma\|w\| = 1$. The optimization problem is then equivalent to the following minimization problem:

$$\min_w \|w\|^2 \qquad \text{such that} \qquad y_i(w^T x_i + b) \geq 1 \qquad \forall 1 \leq i \leq n \,. \tag{50.68}$$

The KKT conditions (Property 50.3.6) for this problem give the following results:

$$w = \sum_{i=1}^n \beta_i y_i x_i \tag{50.69}$$

and

$$\sum_{i=1}^n \beta_i y_i = 0 \,, \tag{50.70}$$

where the quantities β_i are the KKT multipliers for the affine constraints $1 - y_i(w^T x_i + b) \leq 0$. Using these relations, the quantity y can be expressed for a new data point as follows:

$$y \equiv w^T x + b = \sum_{i=1}^{n} \beta_i y_i \langle x_i \mid x \rangle + b\,. \tag{50.71}$$

Two observations can be made at this point. First of all, complementary slackness (Remark 50.3.7) implies that the only relevant vectors x_i in this calculation are the ones that satisfy $\gamma(x_i) = 0$. These are called the **support vectors** and they give their name to a class of models called **support-vector machines** (SVMs). These are the models that are trained using the above optimization problem. Furthermore, y can be written in terms of an inner product. It is exactly this last observation that allows for the generalization of the above model to nonlinear decision boundaries. The previous section showed that inner products are equivalent to (Mercer) kernels. Hence, by choosing a nonlinear kernel function, one can implicitly work with nonlinear feature maps. This is often called the **kernel trick**. As an example, polynomial kernels represent feature maps from x to monomials in the coefficients of x.

However, as often happens with data analysis algorithms, this procedure is sensitive to outliers. This is especially the case for kernels that are based on feature maps to infinite-dimensional spaces (e.g. the *RBF kernel*). To resolve this problem, one can introduce a regularization term in the cost function. The simplest such term for support-vector machines is a simple ℓ^1-penalty:

$$\min_w \|w\|^2 + C \sum_{i=1}^{n} \xi \qquad \text{such that} \qquad \begin{cases} \xi_i \geq 0 & \forall 1 \leq i \leq n\,, \\ y_i(w^T x_i + b) \geq 1 - \xi_i & \forall 1 \leq i \leq n\,. \end{cases} \tag{50.72}$$

The resulting KKT conditions are as follows:

$$0 \leq \beta_i \leq C \tag{50.73}$$

and

$$\begin{aligned} \beta_i = 0 &\implies y_i(w^T x_i + b) \geq 1\,, \\ \beta_i = C &\implies y_i(w^T x_i + b) \leq 1\,, \\ \beta_i \in]0, C[&\implies y_i(w^T x_i + b) = 1\,. \end{aligned} \tag{50.74}$$

@@ COMPLETE (e.g. geometry) @@

50.7 Complexity theory

50.7.1 VC dimension

Definition 50.7.1 (Shatter coefficient). Let $\mathcal{X}$ denote the universe of discours and consider a set $C \in P(\mathcal{X})$. C **shatters** a set $A \in \mathcal{X}$ if, for every subset $a \subseteq A$, there exists a

$c \in C$ such that

$$a = c \cap A\,. \tag{50.75}$$

The shatter(ing) coefficients of C are defined as follows:

$$S_n(C) := \max_{x_1,\dots,x_n \in \mathcal{X}} \left|\left\{\{x_1,\dots,x_n\} \cap c \mid c \in C\right\}\right|. \tag{50.76}$$

It should be clear that every shatter coefficient S_n is bounded above by 2^n. If $S_n(C) = 2^n$, then C shatters some set of cardinality n.

Given a collection of binary functions $\mathcal{F} \subseteq \{0,1\}^{\mathcal{X}}$, the shatter(ing) coefficients (or **growth functions**) of $\mathcal{F}$ are given by:

$$S_n(\mathcal{F}) := \max_{x_1,\dots,x_n \in \mathcal{X}} \left|\left\{\{f(x_1),\dots,f(x_n)\} \mid f \in \mathcal{F}\right\}\right|. \tag{50.77}$$

The shatter coefficients give a notion of the effective size of $\mathcal{F}$, since they say how many different results the functions can produce given a data set of n points.

Definition 50.7.2 (VC dimension). The Vapnik–Chervonenkis (VC) dimension of a collection of binary functions $\mathcal{F}$ is defined as follows:

$$\mathsf{VC}(\mathcal{F}) := \max\{k \in \mathbb{N} \mid S_k(\mathcal{F}) = 2^k\}\,. \tag{50.78}$$

Note that, if a collection shatters $n \in \mathbb{N}$ points, it also necessarily shatters a subset of these points, therefore, one also has

$$S_n(\mathcal{F}) = 2^n \tag{50.79}$$

for all $n < \mathsf{VC}(\mathcal{F})$. A collection is called a **Vapnik–Chervonenkis class** if its VC dimension is finite.

Property 50.7.3 (Sauer's lemma[8]). The VC dimension bounds shatter coefficients in the following way:

$$S_n(\mathcal{F}) \leq \sum_{k=0}^{\mathsf{VC}(\mathcal{F})} \binom{n}{k} \tag{50.80}$$

for all $n \in \mathbb{N}$. For $n \leq \mathsf{VC}(\mathcal{F})$, the right-hand side is just the full binomial expansion for 2^n and, accordingly, the shatter coefficients grow exponentially. For $n \geq d$, the binomial series is truncated and polynomial behaviour is obtained:

$$\forall n \geq \mathsf{VC}(\mathcal{F}) : S_n(\mathcal{F}) \leq \left(\frac{en}{\mathsf{VC}(\mathcal{F})}\right)^{\mathsf{VC}(\mathcal{F})}, \tag{50.81}$$

where e is the Euler number.

[8]Sometimes called the **Sauer–Shelah lemma**.

The generalization bounds of Section 50.4.2 on empirical risk minimization can be extended to uncountable hypothesis spaces as follows.

Property 50.7.4 (Generalization bound). The expected risk of the empirical risk minimizer $\hat{f}$ is bounded as follows:

$$R(\hat{f}) \leq \inf_{f \in \mathcal{H}} R(f) + 4\sqrt{2\frac{\mathsf{VC}(\mathcal{H})\ln(en/\mathsf{VC}(\mathcal{H})) + \ln(2/\delta)}{n}} \tag{50.82}$$

with probability at least $1 - \delta$.

@@ CHECK ALL THESE BOUNDS @@

50.7.2 Rademacher complexity

Remark 50.7.5 (Real-valued functions). The VC dimension of a collection of arbitrary real-valued functions can be defined as the VC dimension of the corresponding collection of indicator functions (Heaviside functions):

$$\mathsf{VC}(\mathcal{F}) := \mathsf{VC}\big(\{\theta(f(x) - \lambda) \mid f \in \mathcal{F}, \lambda \in \mathbb{R}\}\big). \tag{50.83}$$

This definition is equivalent to the following one based on subgraphs:

$$\mathsf{VC}(\mathcal{F}) = \mathsf{VC}\big(\{C_f := \{(x,\lambda) \in \mathcal{X} \times \mathbb{R} \mid \lambda < f(x)\} \mid f \in \mathcal{F}\}\big). \tag{50.84}$$

Example 50.7.6 (Linear spaces). Every vector space V of real-valued functions has VC dimension at most $\dim(V) + 1$.

Example 50.7.7 (Translations). The set of translates of a real-valued function has VC dimension 1.

Although this remark says that one can, in theory, extend ordinary VC theory to arbitrary (real-valued) functions, this does not mean that the VC bounds obtained before still make sense. To obtain useful bounds, it is necessary to introduce a new notion of effective size.

Definition 50.7.8 (Rademacher complexity). Consider a collection of functions $\mathcal{F} \subseteq \mathbb{R}^{\mathcal{X}}$. The Rademacher complexity of $\mathcal{F}$ is defined as follows:

$$\mathfrak{R}_n(\mathcal{F}) := \mathsf{E}\left[\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \sigma_i f(X_i)\right], \tag{50.85}$$

where the σ_i are Rademacher variables (Example 45.3.2) and the expectation is taken over both the Rademacher variables and the sample $(X_1, \dots, X_n)$.

Property 50.7.9 (Risk bound). Consider a collection of bounded functions $\mathcal{F} \subseteq [a,b]^{\mathcal{X}}$.

$$\mathsf{E}\big[R(f)\big] \leq R_{\text{emp}}(f) + 2\mathfrak{R}_n(\mathcal{F}) + (b-a)\sqrt{\frac{\log(1/\delta)}{2n}} \tag{50.86}$$

with probability at least $1 - \delta$.

Property 50.7.10 (VC dimension). The shatter coefficient bounds the Rademacher complexity in the following way:

$$\mathfrak{R}_n(\mathcal{F}) \leq \sqrt{\frac{2\ln(S_n(\mathcal{F}))}{n}}\,. \tag{50.87}$$

50.7.3 Relation to Glivenko–Cantelli classes

Property 50.7.11. Recall Property 48.2.13. The empirical L^1-norm only depends on the values of the functions $f \in \mathcal{F}$ at the given data points and, therefore, the covering number of $\mathcal{F}$ is bounded above by the covering number of $\{(f(x_1), \dots, f(x_n)) \mid f \in \mathcal{F}\}$. The latter is itself bounded above by the shatter coefficient $S_n(\mathcal{F})$. If $\mathcal{F}$ has finite VC dimension, Sauer's lemma 50.7.3 implies that this coefficient grows polynomial in n, so

$$\frac{1}{n}\ln N_C(\varepsilon, \mathcal{F}_M, \|\cdot\|_1) \xrightarrow{d} 0 \tag{50.88}$$

and, thus, $\mathcal{F}$ is Glivenko–Cantelli (Definition 48.2.10).

Theorem 50.7.12. *A class of sets is Vapnik–Chervonenkis if and only if it is Glivenko–Cantelli.*

50.8 Time series analysis

Definition 50.8.1 (Time series). An $\mathbb{N}$- or $\mathbb{Z}$-indexed stochastic process (Chapter 46). Since $\mathbb{N}$ and $\mathbb{Z}$ are isomorphic, the two conventions for time series will be used interchangeably.

50.8.1 Stationarity

Definition 50.8.2 (Strict stationarity). A time series $(X_n)_{n\in\mathbb{N}}$ is (strictly) stationary if, for any two integers $r, s \in \mathbb{N}$, the joint distribution satisfies the following condition:

$$P(X_{t_1}, \dots, X_{t_r}) = P(X_{t_1+s}, \dots, X_{t_r+s})\,. \tag{50.89}$$

Definition 50.8.3 (Weak stationarity). A time series $(X_n)_{n\in\mathbb{N}}$ is said to be weakly (or **covariance**) stationary if it satisfies the following conditions:

1. **Mean-stationary**: $\mathsf{E}[X_n] = \mathsf{E}[X_0]$ for all $n \in \mathbb{N}$.
2. **Finite covariance**: $\mathsf{Cov}[X_i, X_j] < +\infty$ for all $i, j \in \mathbb{N}$.
3. **Covariance-stationary**: $\mathsf{Cov}[X_i, X_{i+j}] = \mathsf{Cov}[X_0, X_j]$ for all $i, j \in \mathbb{N}$.

The following definition is a reformulation of Birkhoff ergodicity (Theorem 18.2.26).

Definition 50.8.4 (Ergodicity). A time series $\{X_t\}_{t\in\mathbb{Z}}$ is ergodic if, for every measurable function f, the following equation holds for all $t \in \mathbb{Z}$:

$$\lim_{T\to\infty} \frac{1}{2T+1} \sum_{k=-T}^{T} f(X_k) = \mathsf{E}\big[f(X_t)\big] \,. \tag{50.90}$$

Intuitively, this means that state space averages can be evaluated as time averages.

50.8.2 Correlation

Definition 50.8.5 (Autocorrelation function). Consider a time series $(X_n)_{n\in\mathbb{N}}$. The associated autocovariance (resp. autocorrelation) process is defined as the covariance (resp. correlation) function of the random variables $(X_n)_{n\in\mathbb{N}}$.

Remark 50.8.6. Note that some authors use a different approach and omit the standardization of the variables, i.e. they take the autocorrelation (and, by, extension, the cross-correlation) at lag $k \in \mathbb{N}$ to be the inner product $\mathsf{E}\big[X_n \cdot X_{n+k}\big]$.

Definition 50.8.7 (Spectral density). Consider a (weakly) stationary time series $(X_n)_{n\in\mathbb{N}}$. If the associated autocovariance is in ℓ^1, one can define the spectral density as the discrete Fourier transform of the autocovariance function:

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{+\infty} \gamma(k) e^{i\omega k} \,, \tag{50.91}$$

where $\gamma(k)$ is the autocovariance function at lag k.

Under the assumption that the spectral density exists, the process is said to have **short memory** if $f(0)$ is finite. Otherwise, the series is said to have **long memory**.

Definition 50.8.8 (Lag operator[9]). The lag operator sends a variable in a $\mathbb{Z}$-indexed time series to the preceding value:

$$BX_t = X_{t-1} \,. \tag{50.92}$$

An important concept, especially in the context of autoregressive models, is that of a **lag polynomial** (the notation for these is not completely fixed in the literature, but the θ-notation is a common choice):

$$\theta(B) = 1 + \sum_{i=1}^{k} \theta_i B^i \,, \tag{50.93}$$

$$\varphi(B) = 1 - \sum_{i=1}^{k} \varphi_i B^i \,. \tag{50.94}$$

Notation 50.8.9 (Difference operator). The difference operator Δ is defined as follows:

$$\Delta = 1 - B \,. \tag{50.95}$$

[9] Also called the **backshift operator**.

In a similar way, one can define the **seasonal** difference operator:

$$\Delta_s = 1 - B^s \,. \tag{50.96}$$

Method 50.8.10 (Ljung–Box test). A test to see if a given set of autocorrelations of a time series is different from zero. Consider a process of $n \in \mathbb{N}$ elements and let $\{\rho_i\}_{1 \le i \le k}$ be the first $k \in \mathbb{N}$ lagged autocorrelation functions. The test statistic is defined as

$$Q = n(n+2) \sum_{i=1}^{k} \frac{\rho_k}{n-k} \,. \tag{50.97}$$

If the null hypothesis "there is no correlation" is true, the Q-statistic will asymptotically follow a χ^2-distribution with k degrees of freedom.

Method 50.8.11 (Augmented Dickey–Fuller test). Consider a time series $(X_t)_{t \in T}$. The (augmented) Dickey–Fuller test checks if the process is (trend) stationary. For this test, one considers the following regression model (similar to the ARIMA-*models* discussed in the next section):

$$\Delta X_t = \alpha + \beta t + \gamma X_{t-1} + \sum_{i=1}^{p-1} \theta_i \Delta X_{t-i} + \varepsilon_t \,. \tag{50.98}$$

The test statistic is

$$\mathsf{DF} := \frac{\gamma}{\mathsf{SE}(\gamma)} \,, \tag{50.99}$$

where SE denotes the standard error. The null hypothesis states that $\gamma = 0$, i.e. there is a *unit root* $(1 - B)$ present in the model (see also the next section). Comparing the test statistic to tabulated critical values will give an indication whether to reject the hypothesis or not (the more negative the statistic, the more significant the result).

50.8.3 Autoregressive models

Definition 50.8.12 (AR(p)-model). Consider a time series $(X_t)_{t \in T}$. The autogressive model of order $p \in \mathbb{N}$ is defined as the multiple linear regression model of X_t with respect to the first p lagged values $X_{t-1}, \dots, X_{t-p}$ of the process:

$$X_t = \beta_0 + \beta_1 X_{t-1} + \cdots + \beta_p X_{t-p} + \varepsilon_t \,. \tag{50.100}$$

Definition 50.8.13 (Partial autocorrelation function). The pth partial autocorrelation function is defined as the pth coefficient in the AR(p)-model.

Remark 50.8.14. The optimal order $p \in \mathbb{N}$ of an autoregressive model is the one for which all higher partial autocorrelation functions (almost) vanish.

Definition 50.8.15 (MA(p)-model). Consider a time series $(X_t)_{t\in T}$ where every variable X_t contains a white noise contribution $\varepsilon_t \sim \mathcal{N}(0,\sigma^2)$. The moving average model of order $p \in \mathbb{N}$ is defined as the multiple linear regression model of X_t with respect to the first p lagged values $\varepsilon_{t-1},\dots,\varepsilon_{t-p}$ of the error term:

$$X_t = \beta_0 + \beta_1\varepsilon_{t-1} + \cdots + \beta_p\varepsilon_{t-p} + \varepsilon_t\,. \tag{50.101}$$

Since the error terms are assumed to have mean zero, one can see that the intercept term β_0 gives the mean of the process.

Remark 50.8.16. The optimal order $p \in \mathbb{N}$ of a moving average model is the one for which all higher autocorrelation functions (almost) vanish.

Definition 50.8.17 (Invertibility). An MA(q)-model is said to be invertible if all roots of its associated lag polynomial $\theta(B)$ lie outside the unit circle. This condition implies that the lag polynomial is invertible, i.e. $1/\theta(B)$ can be written as a convergent series in the operator B. This in turn implies[10] that one can write the MA(q)-model as an AR(p)-model, where possibly $p = +\infty$. The analogous property for AR(p)-models leads to a definition of **stationarity**.

In practice, it is not always possible to describe a data set using either an autoregressive or a moving average model. However, these two types of models can be combined.

Definition 50.8.18 (ARMA(p,q)-model).

$$X_t = \alpha_0 + \sum_{i=1}^{p}\alpha_i X_{t-i} + \sum_{j=1}^{q}\beta_j\varepsilon_{t-j} + \varepsilon_t \tag{50.102}$$

As above, one can find the optimal values for p and q by analyzing the autocorrelation and partial autocorrelation functions.

Using the lag polynomials, one can rewrite the ARMA(p,q)-model as follows:

$$\varphi(B)X_t = \alpha_0 + \theta(B)\varepsilon_t\,. \tag{50.103}$$

By considering the special case where the lag polynomial has a unit root $1-B$ with multiplicity d, one can obtain a generalization of the model:

$$\varphi(B)(1-B)^d X_t = \alpha_0 + \theta(B)\varepsilon_t\,. \tag{50.104}$$

The interpretation of this additional factor $(1-B)^d$ is related to the stationarity of the process. The operator $1-B$ is a finite difference operator:

$$\begin{aligned}(1-B)X_t &= X_t - X_{t-1}\\ (1-B)^2X_t &= (X_t - X_{t-1}) - (X_{t-1} - X_{t-2})\\ &\dots\end{aligned} \tag{50.105}$$

[10] Sometimes this is used as a definition of invertibility.

By successive applications, one can obtain a stationary process from a nonstationary one. This combination of differencing, autoregression and moving averages is called the **ARIMA**-model[11].

Remark 50.8.19. Including so-called *exogenous* variables, i.e. external predictors, leads to an **ARIMAX**-model.

Remark 50.8.20 (Fitting AR- and MA-models). As is clear from the definition of an AR(p)-model, the parameters θ_i can easily be found using standard techniques for multivariate linear regression such as ordinary least squares. However, in contrast to AR-models, where the predictors are known, the estimation of coefficients in MA-models is harder since the error terms ε_t are, by definition, unknown.

To estimate the coefficients in an MA-model, people have introduced multiple techniques (see for example Sandgren and Stoica (2006)). One of the most famous ones is the following method.

Method 50.8.21 (Durbin). By restricting to invertible MA(q)-models (or by approximating a noninvertible model by an invertible one), one can first fit an AR(p)-model with $p > q$ to obtain estimates for the errors ε_t and then, in a second step, use a least squares-method to solve for the coefficients in the MA-model.

As a last modification, one can introduce seasonal components. Simple trends such as a linear growth are easily removed from the process by detrending or differencing. However, a periodic pattern is harder to remove and, in general, ARIMA-models are not suited to accompany this type of features. Luckily, one can easily modify the ARIMA-model to incorporate seasonal variations. The multiplicative SARIMA-model is obtained by inserting operators similar to the ones of the ordinary ARIMA-model, where the lag operator B is replaced by the seasonal lag operator B^s (where $s \in \mathbb{N}$ is the period of the seasonal variation).

Definition 50.8.22 (ARIMA$(p,q,d)(P,Q,D)_s$-model).

$$\Phi(B^s)\varphi(B)\Delta_s^D\Delta^d X_t = \theta(B)\Theta(B^s)\varepsilon_t \tag{50.106}$$

50.8.4 Granger causality

Definition 50.8.23 (Granger causality). Consider two time series $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$. The process X_n is said to (Granger-)cause Y_n if past values of X_n help to predict future values of Y_n. More formally, this can be stated as follows:

$$P\big(Y_{t+k} \in A \mid \Omega(t)\big) \neq P\big(Y_{t+k} \in A \mid \Omega\backslash X(t)\big) \tag{50.107}$$

[11] The 'I' stands for integrated.

for some $k \in \mathbb{N}$, where $\Omega(t)$ and $\Omega \backslash X(t)$ denote the available information at time t with and without removing the variable X from the universe.

This formulation of causality was introduced by *Granger* under the following two assumptions:

- The cause always happens prior to the effect.
- The cause carries unique information about the effect.

Remark 50.8.24. A slightly different, but, for computational purposes often more useful[12] notion of Granger causality, is as follows. A time series $(X_n)_{n\in\mathbb{N}}$ is said to **Granger-cause** a time series $(Y_n)_{n\in\mathbb{N}}$ if the variance of predictions of Y_n becomes smaller when the information contained in X_n is taken into account.

Remark 50.8.25 (Testing). Assume that two uncorrelated models giving predictions of a time series are given. One way to check if they have the same accuracy is the *Diebold–Mariano test*. However, when testing for Granger causality, one should pay attention. This test is not valid for nested models and, hence, is not applicable to two models that only differ by a set of extra predictors (in this case, an external process).

50.9 Uncertainty quantification

50.9.1 Prediction regions

One of the simplest ways to express uncertainty about predictions or parameter estimates is to give a set of possible values instead of a single value. However, to be meaningful, these sets should satisfy some conditions.

Definition 50.9.1 (Validity). Consider a measurable function $\Gamma : \mathcal{X} \to P(\mathcal{Y})$ and let P be the joint distribution on the instance space $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$. Γ is said to be valid (at **significance level** $\alpha \in [0,1]$ or **confidence level** $1-\alpha$) if it satisfies

$$P\big(Y \in \Gamma^\alpha(X)\big) \geq 1 - \alpha\,. \tag{50.108}$$

One sometimes also distinguishes between exact validity and **conservative** validity, where the former is the subcase of the latter for which the inequality becomes an equality.

In fact, one can define two notions of validity: pointwise and asymptotic. The former is characterized by Eq. (50.108) in the sense that the probability of having an error is given by a Bernoulli process with parameter α. Asymptotic validity is a frequentist

[12] In fact, this was the original definition by *Granger*.

notion in the following sense:

$$\lim_{n\to\infty} \frac{\mathsf{Err}_n(\Gamma)}{n} \leq \alpha\,, \tag{50.109}$$

where $\mathsf{Err}_n(\Gamma)$ is the number of errors made by Γ after $n \in \mathbb{N}$ trials. It should be clear that pointwise validity (both exact and conservative) implies asymptotic validity.

Remark 50.9.2 (Confidence regions). The definition of (valid) confidence predictors above is similar to the definition of confidence regions (Section 48.6). However, in contrast to confidence regions for population parameters, the size of confidence regions for predictive distributions does not go towards zero in the infinite data limit. This follows from the fact that, in general, all observations are subject to noise and, hence, even with perfect knowledge about the data-generating distribution, an exact prediction is impossible.[13]

50.9.2 Classifier calibration

A specific instance of regression problems are classification tasks, where the one tries to model a function $f : \mathcal{X} \to \mathcal{Y}$ with $\mathcal{Y}$ finite (or discrete). For the case of probabilistic classifiers, the notion of validity (Definition 50.9.1) admits the following formulation:

Definition 50.9.3 (Calibration). A probabilistic (multiclass) classifier $\widehat{P}(\cdot \mid \cdot) : \mathcal{Y}\times\mathcal{X} \to [0,1]$ is said to be (well-)calibrated if

$$P\big(Y \,\big|\, \widehat{P}(Y \mid X) = p\big) = p \tag{50.110}$$

for all $p \in [0,1]$. Here, the confidence level $1-\alpha$ is the output of the classifier $\widehat{P}$, i.e. instead of asking for a region satisfying a given confidence level, the model yields the confidence level required to include a given class. A possible way to visually investigate the calibration of a probabilistic classifier is to draw **calibration plots** or **reliability diagrams** for all classes, where, for every class label Y and for a suitable partition $\mathcal{P} := \{0 = p_1, p_2, \dots, p_n = 1\}$ of the interval $[0,1]$, the ith point is the proportion of instances for which Y was predicted with probability between p_i and p_{i+1}. For a calibrated model, these points should lie on the diagonal.

Consider now the case of binary classification: $\mathcal{X} \to \{0,1\}$. Here one can easily apply the conformal prediction framework from the previous section. To this end, define the following nonconformity measure:

$$A_{\text{binary}}\big((x,y)\big) := 1 - \widehat{P}(y \mid x)\,. \tag{50.111}$$

At significance α, the model predicts all classes such that the conformal p-value is greater than α. As before, one can adopt an inductive framework and use the $(1-\alpha)$-quantile of the calibration scores as a threshold.

[13] Note that when observations are sampled deterministically, this is of course not true.

In fact, for classification tasks, where the codomain is finite (or at least discrete), it makes sense to slightly alter the approach. Instead of fixing the significance level α, one could (or should) consider some specific values:

- **Confidence**: $\sup\{1-\alpha \mid |\Gamma^\alpha(x)| \leq 1\}$, and
- **Credibility**: $\inf\{\alpha \mid |\Gamma^\alpha(x)| = 0\}$.

The former is the highest probability with which the model can predict a single class. If one wants a higher probability, more than one class has to be considered. The latter gives a measure of how 'credible' the predictions are. The smaller this number, i.e. the greater the confidence with which the model predicts a vacuously false result (every sample necessarily belongs to a class), the less reliable the model is for a given sample.

Various other approaches exist to calibrate existing probabilistic classifiers. In practice, it has been observed that many models show sigmoidal distortions in their calibration plots. A possible approach is then to modify the final sigmoidal layer in these models (or add an extra sigmoidal layer if the model did not use one). This gives rise to **Platt scaling** and **temperature scaling**. For the former, one takes the output $\widehat{P}$ and fits a sigmoidal layer of the form

$$\widehat{P}_{\text{Platt}}(y \mid x; A, B) = \frac{1}{1 + \exp\big(-A\widehat{P}(y \mid x) - B\big)}\,. \tag{50.112}$$

For the latter, one modifies the existing logits as follows for some parameter $T > 0$:

$$\widehat{P}_{\text{temp}}(y \mid x; T) = \frac{1}{1 + \exp\big(-\hat{z}(y \mid x)/T\big)}\,, \tag{50.113}$$

where $\hat{z}$ represents the logits of the estimator $\widehat{P}$.

Remark 50.9.4 (Accuracy). Because temperature scaling does not change the maximum of the softmax function, the eventual predictions do not change.

50.9.3 Conditionality

The above sections considered confidence predictors that were valid in a global sense, i.e. averaged over the whole instance space $\mathcal{X} \times \mathcal{Y}$. However, as in ordinary probability theory, in many settings it makes more sense to consider conditional statements:

$$\mathsf{Prob}\big(Y \in \Gamma^\alpha(X) \mid \kappa(X, Y)\big) \geq 1 - \alpha\,, \tag{50.114}$$

where the function $\kappa : \mathcal{X} \times \mathcal{Y} \to K$ represents the conditioning statement (the set K can be any set). In the conformal prediction literature it is often called a **taxonomy function**.

In practice, the label set K is often finite and discrete, i.e. one considers a finite subdivision of $\mathcal{X} \times \mathcal{Y}$. The simplest solution to obtain conditional validity in this case is to apply any known algorithm to every conditional class individually. In the case of conformal prediction, this gives rise to the notion of **Mondrian conformal predictors**.

In a perfect world, the ultimate goal should be to have exact objectwise validity:

$$\mathsf{Prob}(Y \in \Gamma^\alpha(X) \mid X) \geq 1 - \alpha\,. \tag{50.115}$$

However, in general, one can show that this cannot be attained:

Property 50.9.5 (No-go theorem). Let $\mathcal{X}$ be a separable metric space equipped with its canonical Borel σ-algebra and consider a confidence predictor Γ^α at conditional significance level $\alpha \in [0,1]$. For every probability distribution P on $\mathcal{X} \times \mathcal{Y}$ and P_X-almost all non-atoms $x \in \mathcal{X}$, one has

$$\mathsf{Prob}\big(\lambda(\Gamma^\alpha(X)) = +\infty\big) \geq 1 - \alpha \tag{50.116}$$

and

$$\mathsf{Prob}\big(\mathsf{co}(\Gamma^\alpha(X)) = \mathbb{R}\big) \geq 1 - 2\alpha\,, \tag{50.117}$$

where λ and co denote the Lebesgue measure and convex hull, respectively.

@@ FINISH @@

50.9.4 Distribution shift

An important problem in the field of data science, in particular that of machine learning, is the change of the data generating process. Consider a classic train-test routine. If the model was trained on a data set sampled from the distribution P_0, but applied to a data set sampled from the distribution P_1, there is no reason the expect that the model will still give reasonable results. Furthermore, even if the crude point predictions are still somewhat sensible, this does not mean that their theoretical properties, such as the validity of confidence regions, is preserved.

A first step to resolve this problem is the detection of a possible distribution shift. Conformal prediction, as introduced in the previous section, can be used to detect such shifts in an online fashion. The main idea of the algorithm is to construct a martingale (Definition 46.3.1) from the p-values produced by a conformal predictor. For every (reasonable) martingale $(X_n)_{n\in\mathbb{N}}$ the Doob-Ville inequality (Property 46.3.5) says that the growth is bounded. However, if the sequence $(X_n)_{n\in\mathbb{N}}$ is constructed in such a way that the martingale property is lost once the data distribution changes, the Doob–Ville inequality can be violated and the crossing of a given threshold can be regarded as evidence for this distribution shift (Vovk et al., 2005).

The general expression of the martingale will be of the form

$$X_i \equiv \prod_{k=0}^{i} f_k(p_k)\,, \tag{50.118}$$

where p_k is the p-value of the kth data point as produced by some conformal predictor. The functions f_k are called the **betting functions**. For $(X_n)_{n\in\mathbb{N}}$ to be a martingale with respect to the natural filtration of the p_k's, the betting functions should satisfy

$$\int_0^1 f_k(p)\,dp = 1\,. \tag{50.119}$$

Method 50.9.6 (Power martingale). For every constant $\varepsilon \in [0,1]$ one defines the power martingale as

$$X_i^\varepsilon := \prod_{k=0}^{i} \varepsilon p_k^{\varepsilon-1}\,. \tag{50.120}$$

One can also construct a **simple mixture** martingale by integrating over ε:

$$X_i := \int_0^1 X_i^\varepsilon\,d\varepsilon\,. \tag{50.121}$$

Because $\varepsilon \leq 1$, the above martingales will start to become very large if the conformal predictor produces small p-values, i.e. when unlikely values are observed.

However, not all distribution shifts will give rise to such behaviour. For example, it is possible that, although the p-values are not distributed uniformly anymore (since the data is not exchangeable), they are concentrated in the upper half of the unit interval and, hence, do not let the martingale grow strongly. For this reason, it is convenient to construct betting functions that take into account the distribution of the p-values.

Method 50.9.7 (Plug-in martingale). Let $\hat{\rho}_i(p)$ denote an estimate of the probability density constructed using the p-values $\{p_1,\dots,p_i\}$. The plug-in martingale is defined by the betting functions

$$f_i := \hat{\rho}_{i-1}\,. \tag{50.122}$$

Now, if the empirical distribution functions of the p-values converge weakly (Definition 18.1.53) to an absolutely continuous distribution and $\log(\hat{\rho}_i(p)) \longrightarrow \log(\rho(p))$ uniformly, where ρ is the limit density of the empirical distributions, it can be shown that plug-in martingale grows quicker than the martingale associated to any other (continuous) betting function.

A different approach is to use p-values that are constructed directly from disitributional data. For inspiration, consider the following property.

Property 50.9.8. Consider a data sample $(x_n)_{n\in\mathbb{N}}$ and let f, g be two possible probability densities describing this sample. If f is the true density, the likelihood process

$$X_i := \prod_{k=0}^{i} \frac{g(x_i)}{f(x_i)} \tag{50.123}$$

is a martingale.

So, likelihood ratios already give rise to test martingales. However, this martingale cannot be used to check for distribution shifts, since it is only a martingale when the true density is used.

Example 50.9.9 (Likelihood nonconformity). For every two probability densities f, g one can define a nonconformity measure as follows:

$$A_{\text{N–P}}(z) := \log f(z) - \log g(z)\,, \tag{50.124}$$

i.e. the log-likelihood ratio.[14] By using this nonconformity measure in combination with the above plug-in approach, one can obtain a likelihood-based changepoint detection algorithm.

50.10 Generative models

Most models considered in classical statistics and data analysis are of the inferential type, i.e. given a data set a model is constructed to estimate some property of the data generating distribution. In statistics this usually is either a moment or the distribution itself, in data analysis it is often the conditional mean or conditional distribution.

However, in certain cases, it is often desirable to have a model that can sample new data points from the data generating distribution. Models of this type are called generative models. (Note that fully distributional estimators can be considered both inferential and generative.)

@@ ADD (e.g. GANs) @@

50.10.1 Normalizing flows

One of the main problems for obtaining valid uncertainty estimates is that the underlying distribution of a given data set is in general unknown. Except for the conformal prediction framework, all other methods make some assumptions about the underlying data generating process (even CP makes the exchangeability assumption). One

[14] The subscript refers to the Neyman–Pearson lemma 48.7.12 for which this function gives the (logarithm) of the test statistic.

way around this problem is by first transforming the data such that it is sampled according to a well-known distribution.

A general normalizing flow for a distribution $P \in \mathbb{P}(\mathbb{R}^n)$ consists of a sequence of invertible (and measurable) transformations[15] $\{\Phi\}_{i \leq k} : \mathbb{R}^n \to \mathbb{R}^n$ such that $(\Phi_k \circ \ldots \circ \Phi_1)_* P$ is (multi)normal.

Consider now the specific case of regression problems. The usefulness of normalizing flows is supported by the following theorem:

Theorem 50.10.1. *Any two absolutely continuous distributions on $\mathbb{R}^n$ are related by an increasing triangular Borel transformations.*

Triangular means that the transformation can be expressed as follows:

$$\psi(x_1, \ldots, x_n) \equiv \begin{pmatrix} \psi_1(x_1) \\ \psi_2(x_2 \mid x_2) \\ \vdots \\ \psi_n(x_n \mid x_1, \ldots, x_{n-1}) \end{pmatrix}, \tag{50.125}$$

with $\psi_i : \mathbb{R}^i \to \mathbb{R}$ for all $i \leq n$. Increasing means that each of the functions $\psi_i : \mathbb{R}^i \to \mathbb{R}$ is increasing with respect to its first argument x_i. This is sufficient for the transformation to be invertible, since one can invert the functions stepwise, keeping the variables $x_1, \ldots, x_{i-1}$ fixed in the ith step (hence the suggestive conditional notation).

The general construction for normalizing flow regression is to train a model $\rho : \mathcal{Y} \to \mathbb{R}^m$ that parametrizes a normalizing flow $\Psi(\cdot\,;\cdot) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ in such a way that for every $y \in \mathcal{Y}$, $\Psi\big(\cdot\,;\rho(y)\big)$ is triangular and increasing. Hence, one passes to 'conditional' normalizing flows.

@@ COMPLETE @@

[15] This makes normalizing flows a subclass of general *transformation models*.

Chapter 51

Fuzzy Sets & Imprecise Probabilities ♣

The main reference for fuzzy set theory is the initial paper by Zadeh (1965). For the basics of (ordered) sets, see Section 1.7 at the beginning of this compendium. The section on imprecise probabilities is mainly based on Augustin, Coolen, De Cooman, and Troffaes (2014).

This chapter begins with a small organizational remark. Although the content of the current chapter fits better in the parts on general set theory and logic, it does use more advanced concepts from, for example, topology and category theory. Furthermore, the main application is the characterization of uncertainty in statistics and machine learning. For that reason, it was added here.

Chapter Contents

51.1 Fuzzy sets **965**

51.2 Imprecise probabilities **966**

51.2.1 Gambles 966

51.2.2 Credal sets 967

51.2.3 Fuzzy measure theory 967

51.1 Fuzzy sets

Definition 51.1.1 (Fuzzy set). Consider a set X (this set corresponds to the universe of discourse in e.g. type theory or category theory). A fuzzy subset of X is a function $A : X \to [0, 1]$. One can interpret the value $A(x)$ at a point $x \in X$ as the degree of membership of x in A. If the function A only takes on values in $\{0, 1\}$, the indicator function of an ordinary subset is recovered. A fuzzy set is said to be **empty** if its membership function is identically zero.

Remark 51.1.2. One can generalize this definition by replacing $[0,1]$ by a more general poset (with the necessary properties).

Definition 51.1.3 (Pullback). Consider two sets X, Y and a fuzzy subset A of Y. Given a function $f : X \to Y$, one can define the pullback f^*A as usual:

$$f^*A := A \circ f\,. \tag{51.1}$$

The following definition is a straightforward generalization of Definition 1.4.6.

Definition 51.1.4 (Fuzzy relation). A fuzzy subset of the product set $X \times X$. This definition can be extended to n-ary relations by considering fuzzy subsets of the n-fold product $X \times \cdots \times X$.

The composition in Definition 1.4.8 can be extended through the following formula:

$$S \circ R(x,z) := \sup_{y \in X} \min\big(R(x,y), S(y,z)\big)\,. \tag{51.2}$$

A more exotic construction for fuzzy sets is the following one (note that this only works if the codomain of fuzzy sets is $[0,1]$).

Definition 51.1.5 (Convex combination). Consider three fuzzy sets A, B, Λ. The convex combination $C \equiv (A, B; \Lambda)$ is defined as follows in analogy to Definition 16.7.1:

$$C(x) := \Lambda(x)A(x) + \big(1 - \Lambda(x)\big)B(x)\,. \tag{51.3}$$

51.2 Imprecise probabilities

51.2.1 Gambles

Definition 51.2.1 (Gamble). Consider a set X. The set of gambles over X is the (Banach) space of bounded real-valued functions

$$\mathcal{B}(X) := \{f : X \to \mathbb{R} \mid \sup_{x \in X} f(x) < +\infty\}\,. \tag{51.4}$$

A subset $\mathcal{D} \subset \mathcal{B}(X)$ of 'desirable' gambles is said to be **coherent** if it satisfies the following condition:

1. **Positivity**: $\lambda > 0 \implies \lambda\mathcal{D} = \mathcal{D}$,
2. **Additivity**: $\mathcal{D} + \mathcal{D} \subseteq \mathcal{D}$,
3. **Accepting partial gains**: $\mathcal{B}^+(X) \subseteq \mathcal{D}$, where $\mathcal{B}^+(X) := \{f \in \mathcal{B}(X) \mid f > 0\}$ denotes the **positive orthant**, and
4. **Avoiding partial losses**: $\mathcal{B}^-(X) \cap \mathcal{D} = \emptyset$, where $\mathcal{B}^-(X) := \{f \in \mathcal{B}(X) \mid f < 0\}$ denotes the negative orthant.

The first two axioms imply that the desirable gambles form a convex cone. It is also clear that the positive orthant $\mathcal{B}^+(X)$ is the smallest coherent set of desirable gambles.

Property 51.2.2 (Order structure). The collection of all coherent sets of desirable gambles over a space X can be given a poset structure by inclusion, with $\mathcal{B}^+(X)$ as its least element. If $\mathcal{D} \subset \mathcal{D}^+$, then $\mathcal{D}$ is said to be less **committal** than $\mathcal{D}'$.

51.2.2 Credal sets

Definition 51.2.3 (Credal set). A subset of the set of probability measures $\mathbb{P}(X)$.

Credal sets are often used to represent the lack of knowledge about a probability distribution. For this reason, it is natural to assume that credal sets are convex. If one is uncertain about both $P_1, P_2 \in K$ with $K \subseteq \mathbb{P}(X)$, one should also be uncertain about the mixtures $\lambda P_1 + (1-\lambda)P_2$ for $\lambda \in [0,1]$.

51.2.3 Fuzzy measure theory

In this section, some of the content of Chapter 18 is generalized to imprecise probability theory. Unless explicitly stated, all concepts will be defined over a general measurable space (X, Σ).

Definition 51.2.4 (Capacity[1]). A set function $\mu : \Sigma \to \mathbb{R}$ satisfying the following conditions:

1. **Grounded**: $\emptyset \in \Sigma \implies \mu(\emptyset) = 0$, and
2. **Monotonicity**: $A \subseteq B \implies \mu(A) \leq \mu(B)$ for all $A, B \in \mathcal{C}$.

If one drops the monotonicity condition, the notion of a **game** is obtained. A capacity is said to be **normalized** (or **regular**) if $\mu(X) = 1$. For infinite sets, the following two conditions are generally added:

3 **Upward continuity**: If $(A_n)_{n\in\mathbb{N}} \subset \Sigma$ is an increasing sequence, then

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} \mu(A_n)\,. \tag{51.5}$$

4 **Downward continuity**: If $(A_n)_{n\in\mathbb{N}} \subset \Sigma$ is a decreasing sequence of compact sets, then

$$\mu\left(\bigcap_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} \mu(A_n)\,. \tag{51.6}$$

[1] Also called a **fuzzy measure**.

Definition 51.2.5 (Alternating capacity). A k-alternating capacity μ satisfies

$$\mu\left(\bigcap_{i=1}^{k} A_i\right) \leq \sum_{I\subset\{1,\dots,k\}} (-1)^{|I|+1}\mu\left(\bigcup_{i\in I} A_i\right) \tag{51.7}$$

for all measurable sets $A_1, \dots, A_k$. A 2-alternating capacity is called a **probability measure** if the inequality is saturated for all A_1, A_2. By interchanging the union and intersection symbols (and the inequality sign), the definition of **k-monotone capacities** is obtained. If a capacity is k-alternating (resp. k-monotone) for all $k \geq 2$, it is also said to be **totally alternating** (resp. **totally monotone**).

Property 51.2.6. A capacity $\mu : \Sigma \to \mathbb{R}$ is k-monotone (resp. k-alternating) if and only if its **dual capacity**

$$\overline{\mu}(A) := \mu(X) - \mu(A^c) \tag{51.8}$$

is k-alternating (resp. k-monotone).

Definition 51.2.7 (Choquet integral). Consider a capacity μ on (X, Σ) and a measurable function $f : X \to \mathbb{R}$, i.e. a function such that $\{y \mid f(y) \geq x\} \in \Sigma$ for all $x \in \mathbb{R}$. The Choquet integral of f is defined as follows:

$$\int_X f\,d\mu := \int_{-\infty}^{0} \big(\mu(\{y \mid f(y) \geq x\}) - \mu(X)\big)\,dx + \int_0^{+\infty} \mu(\{y \mid f(y) \geq x\})\,dx. \tag{51.9}$$

This integral is not additive, but it is monotonic in f. This expression can be obtained as follows, where functions are decomposed in their positive and negative parts: $f = f^+ - f^-$. The original definition by *Choquet* was for nonnegative functions:

$$\int_X f^+\,d\mu := \int_0^{+\infty} \mu(\{y \mid f^+(y) \geq x\})\,dx\,, \tag{51.10}$$

which for simple functions reduces to:

$$\begin{aligned}\int_X f^+\,d\mu &= \int_X \sum_{i=1}^{n} a_i \mathbb{1}_{\{x\in X \mid f^+(x)=a_i\}}\,d\mu \\ &= \int_X \sum_{i=1}^{n} (a_i - a_{i-1}) \mathbb{1}_{\{x\in X \mid f^+(x)\geq a_i\}}\,d\mu \\ &= \sum_{i=1}^{n} (a_i - a_{i-1})\mu(\{x \in X \mid f(x) \geq a_i\}).\end{aligned} \tag{51.11}$$

The general Choquet integral then takes the form

$$\int_X f\,d\mu := \int_X f^+\,d\mu - \int_X f^-\,d\overline{\mu}\,. \tag{51.12}$$

Property 51.2.8. A capacity is 2-alternating if and only if the associated Choquet integral is subadditive.

Example 51.2.9 (Possibility and necessity). A possibility measure is a normalized capacity μ satisfying

$$\mu(A \cup B) = \max\{\mu(A), \mu(B)\} \tag{51.13}$$

for all $A, B \in \Sigma$. If one replaces the maximum by a minimum, the definition of a **necessity measure** is obtained.

Example 51.2.10 (Belief and plausibility). Belief and plausibility measures are, respectively, defined as totally monotone and totally alternating normalized capacities. It is not difficult to show that possibility and necessity measures are particular instances of belief and plausibility measures.

Definition 51.2.11 (Consonant capacity). A capacity $\mu : \Sigma \to \mathbb{R}$ such that the **focal sets**, i.e. the sets $A \in \Sigma$ for which $\mu(A) > 0$, admit a total order (i.e. they are nested).

Example 51.2.12. Necessity and possibility measures are exactly the consonant belief and plausibility measures.

Part VIII

Classical Physics

Table of Contents

52 Classical Mechanics **974**

52.1 Newtonian mechanics . . . 975

52.1.1 Linear motion . . . 975

52.1.2 Rotational motion . . . 976

52.2 Lagrangian mechanics . . . 979

52.2.1 Action . . . 979

52.2.2 Euler–Lagrange equations . . . 980

52.2.3 Kepler problem . . . 985

52.3 Hamiltonian mechanics . . . 986

52.3.1 Poisson bracket . . . 987

52.3.2 Hamilton–Jacobi equation . . . 988

52.4 Continuum mechanics . . . 989

52.4.1 Liouville's theorem . . . 989

52.4.2 Fluid mechanics . . . 991

52.5 Koopman–von Neumann mechanics . . . 992

52.6 Dynamical systems . . . 994

52.7 Geometric mechanics . . . 995

52.7.1 Newton's equations . . . 995

52.7.2 Lagrangian mechanics . . . 997

52.7.3 Hamiltonian mechanics . . . 997

52.7.4 Contact structure . . . 998

53 Classical Field Theory **1000**

53.1 Lagrangian field theory . . . 1000

53.1.1 Noether's theorem . . . 1001

53.2 De Donder–Weyl theory . . . 1003

54 BRST Theory **1004**

54.1 Constrained systems . . . 1005

54.1.1 Constraints . . . 1005

54.1.2 Zero Hamiltonian . . . 1010

54.1.3 Field theories 1011
54.1.4 Constraint surface 1012
54.1.5 Fermionic systems 1013
54.2 Hamiltonian BRST theory 1014
54.2.1 Introduction 1014
54.2.2 Irreducible constraints 1015
54.2.3 Reducible constraints 1019
54.2.4 Observables 1021
54.2.5 Actions and gauge fixing 1022
54.2.6 Nonminimal sectors 1022
54.2.7 Faddeev–Popov action 1023
54.3 Lagrangian BRST theory 1025
54.3.1 Field theory 1025
54.3.2 Gauge algebra 1026
54.3.3 Antifields 1028

55 Electromagnetism 1033
55.1 Electricity 1033
55.1.1 Conductors and resistors 1034
55.1.2 Ohm's law 1036
55.1.3 Electric dipoles 1036
55.2 Magnetism 1038
55.2.1 Lorentz force 1039
55.2.2 Energy and momentum 1039
55.2.3 Electric charges in a magnetic field 1040
55.3 Maxwell equations 1040
55.4 Potentials 1041
55.4.1 Decomposition in potentials 1041
55.4.2 Gauge theory 1041
55.4.3 Differential-geometric perspective 1042

56 Optics 1044
56.1 General 1044
56.1.1 Conservation of energy 1044
56.2 Wave mechanics 1045
56.3 Electromagnetic radiation 1046
56.4 Propagation in media 1048
56.4.1 Reflection 1048
56.4.2 Refraction 1048
56.4.3 Absorption 1049
56.4.4 Interference & Diffraction 1050

Chapter 52

Classical Mechanics

The section about the geometric framework is mainly based on Arnol'd (2013); Palais (1997). For an introduction to differential geometry, see Chapter 31 and onwards.

Chapter Contents

52.1 Newtonian mechanics **975**
52.1.1 Linear motion 975
52.1.2 Rotational motion 976
52.2 Lagrangian mechanics **979**
52.2.1 Action 979
52.2.2 Euler–Lagrange equations 980
52.2.3 Kepler problem 985
52.3 Hamiltonian mechanics **986**
52.3.1 Poisson bracket 987
52.3.2 Hamilton–Jacobi equation 988
52.4 Continuum mechanics **989**
52.4.1 Liouville's theorem 989
52.4.2 Fluid mechanics 991
52.5 Koopman–von Neumann mechanics **992**
52.6 Dynamical systems **994**
52.7 Geometric mechanics **995**
52.7.1 Newton's equations 995
52.7.2 Lagrangian mechanics 997
52.7.3 Hamiltonian mechanics 997
52.7.4 Contact structure 998

52.1 Newtonian mechanics

52.1.1 Linear motion

Axiom 52.1 (Newton's second law). The force acting on a system is related to the change in momentum in the following way:

$$\vec{\mathbf{F}} = \frac{\mathrm{d}\vec{\mathbf{p}}}{\mathrm{d}t}\,. \tag{52.1}$$

Definition 52.1.1 (Work). The work performed along a trajectory γ is given by

$$W := \int_\gamma \vec{\mathbf{F}} \cdot d\vec{\mathbf{l}}\,. \tag{52.2}$$

Definition 52.1.2 (Conservative force). If the work done by a force is independent of the path taken, the force is said to be **conservative**:

$$\oint_C \vec{\mathbf{F}} \cdot d\vec{\mathbf{l}} = 0 \tag{52.3}$$

for all paths $C : S^1 \to \mathbb{R}^3$.

The Kelvin–Stokes theorem 23.2.5 together with Eq. (23.12) allows to rewrite the conservative force as the gradient of a scalar field $V : \mathbb{R}^3 \to \mathbb{R}$:

$$\vec{\mathbf{F}} = -\nabla V\,. \tag{52.4}$$

Note that the **potential (energy)** V is only defined up to a constant.

To generalize this result from position space $\mathbb{R}^3$ to include velocities (or even higher-order derivatives), a variational approach has to be adopted since, in this case, the actual path is of relevance. For velocity-dependent forces, the (generalized) potential $U : \mathbb{R}^6 \to \mathbb{R}$ generates the force as follows:[1]

$$\vec{\mathbf{F}}(\vec{\mathbf{x}}, \vec{\mathbf{v}}) = -\nabla U - \frac{\mathrm{d}}{\mathrm{d}t}(\nabla_{\vec{\mathbf{v}}} U)\,. \tag{52.5}$$

In general, the functional derivative (40.50) is used:

$$\vec{\mathbf{F}}(\vec{\mathbf{r}}, \dot{\vec{\mathbf{r}}}, \dots, t) = -\delta_{\mathrm{EL}} U(\vec{\mathbf{r}}, \dot{\vec{\mathbf{r}}}, \dots, t)\,. \tag{52.6}$$

Definition 52.1.3 (Central force). A force that only depends on the relative position of two objects:

$$\vec{\mathbf{F}}_c = F\big(\|\vec{\mathbf{r}}_2 - \vec{\mathbf{r}}_1\|\big)\hat{\mathbf{e}}_{\mathbf{r}}\,. \tag{52.7}$$

[1] Note that, as with ordinary forces, not all velocity-dependent forces are conservative (in this generalized sense).

Formula 52.1.4 (Momentum). Consider a mass m with velocity $\vec{\mathbf{v}}$. Its momentum is given by

$$\vec{\mathbf{p}} = m\vec{\mathbf{v}}\,. \tag{52.8}$$

If the mass remains constant along the trajectory, Newton's second law (Axiom 52.1) can be rewritten as follows:

$$\vec{\mathbf{F}} = m\frac{\mathrm{d}\vec{\mathbf{v}}}{\mathrm{d}t} = m\vec{\mathbf{a}}\,. \tag{52.9}$$

Formula 52.1.5 (Kinetic energy). For a free particle with total momentum p, the kinetic energy is given by the following formula:

$$E_{\text{kin}} := \frac{p^2}{2m}\,. \tag{52.10}$$

52.1.2 Rotational motion

In this section, $r \in \mathbb{R}^+$ always denotes the distance from the object's center of mass to the axis around which the object rotates. More generally, the vector $\vec{\mathbf{r}} \in \mathbb{R}^3$ will denote the position with respect to the axis of rotation.

Definition 52.1.6 (Angular velocity).

$$\vec{\omega} := \frac{\vec{\mathbf{r}} \times \vec{\mathbf{v}}}{r^2} \tag{52.11}$$

Definition 52.1.7 (Angular frequency).

$$\nu := \frac{\|\vec{\omega}\|}{2\pi} \tag{52.12}$$

Definition 52.1.8 (Moment of inertia). For a (spherically) symmetric object, the moment of inertia is given by

$$I := \int_V r^2\rho(r)\,dV\,, \tag{52.13}$$

where $\rho : \mathbb{R}^3 \to \mathbb{R}^+$ denotes the mass density function. For a general body, the moment-of-inertia tensor is given by

$$\mathcal{J} := \int_V \rho(\vec{\mathbf{r}})\big(r^2\mathbb{1}_{3\times3} - \vec{\mathbf{r}} \otimes \vec{\mathbf{r}}\big)\,dV\,. \tag{52.14}$$

Example 52.1.9 (Azimuthal symmetry). Let M, R denote the mass and the radius of the object, respectively. The (scalar) inertia of some common objects is given:

- Solid disk: $I = \frac{1}{2}MR^2$,
- Cylindrical shell: $I = MR^2$,

- Hollow sphere: $I = \frac{2}{3}MR^2$, and
- Solid sphere: $I = \frac{2}{5}MR^2$.

A proof for the solid disk and sphere are given:

Proof. The volume of a (solid) disk is given by

$$V_{\text{disk}} = \pi R^2 d\,,$$

where R denotes the radius and d denotes the thickness. The mass density is then given by

$$\rho = \frac{M}{\pi R^2 d}\,.$$

Using cylindrical coordinates, the moment of inertia becomes

$$\begin{aligned} I &= \frac{M}{\pi R^2 d}\int_0^{2\pi} d\varphi \int_0^d dz \int_0^R r^3\,dr \\ &= \frac{M}{\pi R^2 d} 2\pi d \frac{R^4}{4} \\ &= \frac{1}{2}MR^2\,. \end{aligned}$$

The volume of a solid sphere is given by

$$V_{\text{sphere}} = \frac{4}{3}\pi R^3\,,$$

where R denotes the radius. The mass density is then given by

$$\rho = \frac{M}{\frac{4}{3}\pi R^3}\,.$$

Spherical coordinates will be used to derive the moment of inertia, but one has to be careful. The r in Definition 52.1.8 is the distance between a point in the body and the axis of rotation. Hence, it is not the same as the radial coordinate r' in spherical coordinates, which is the distance between a point and the origin. However, the relation between these two quantities is easily found using basic geometry:

$$r = r' \sin\theta\,.$$

Now, one can calculate the moment of inertia as follows:

$$\begin{aligned} I &= \frac{M}{\frac{4}{3}\pi R^3}\int_0^{2\pi} d\varphi \int_0^R r'^4\, dr' \int_0^{\pi} \sin^3\theta\, d\theta \\ &= \frac{M}{\frac{4}{3}\pi R^3} 2\pi \frac{R^5}{5}\frac{4}{3} \\ &= \frac{2}{5}MR^2\,. \end{aligned}$$

Definition 52.1.10 (Principal axes of inertia). Let I be the matrix of inertia, i.e. the matrix associated with the inertia tensor in Eq. (52.14). This is a real symmetric matrix and, by Property 22.5.17, admits an eigendecomposition of the form

$$I = Q\Lambda Q^T\,. \tag{52.15}$$

The columns of Q determine the principal axes of inertia. The eigenvalues are called the **principal moments of inertia**.

Theorem 52.1.11 (Parallel axis theorem[2]). *Consider a rotation about an axis ψ through a point A and let ψ_{CM} be a parallel axis through the center of mass. The moment of inertia about ψ is related to the moment of inertia about ψ_{CM} in the following way:*

$$I_\psi = I_{CM} + m\|\vec{\mathbf{r}}_A - \vec{\mathbf{r}}_{CM}\|^2\,, \tag{52.16}$$

where m is the mass of the rotating body.

Definition 52.1.12 (Angular momentum).

$$\vec{\mathbf{L}} := \vec{\mathbf{r}} \times \vec{\mathbf{p}} \tag{52.17}$$

Given the angular velocity vector $\vec{\omega}$, one can compute the angular momentum as follows:

$$\vec{\mathbf{L}} = \mathcal{J}(\vec{\omega})\,, \tag{52.18}$$

where $\mathcal{J}$ is the inertia tensor. If $\vec{\omega}$ is parallel to a principal axis, the formula reduces to

$$\vec{\mathbf{L}} = I\vec{\omega}\,, \tag{52.19}$$

where I is the corresponding principal moment of inertia.

Formula 52.1.13 (Torque). For angular momenta, there exists a formula analogous to Newton's second law:

$$\vec{\tau} := \frac{\mathrm{d}\vec{\mathbf{L}}}{\mathrm{d}t}\,. \tag{52.20}$$

[2] Also called **Steiner's theorem**.

For a constant mass, this formula can be rewritten as follows:

$$\vec{\boldsymbol{\tau}} = \mathcal{J}(\vec{\boldsymbol{\alpha}}) = \vec{\mathbf{r}} \times \vec{\mathbf{F}}\,, \tag{52.21}$$

where

$$\vec{\boldsymbol{\alpha}} := \frac{\mathrm{d}\vec{\boldsymbol{\omega}}}{\mathrm{d}t} \tag{52.22}$$

is the **angular acceleration**.

Remark 52.1.14. From the previous definitions, it follows that both the angular momentum and torque vectors are in fact pseudovectors and, accordingly, change sign under coordinate transformations with negative determinant.

Formula 52.1.15 (Rotational energy).

$$E_{\mathrm{rot}} := \frac{1}{2}\mathcal{J}(\vec{\boldsymbol{\omega}}) \cdot \vec{\boldsymbol{\omega}} \tag{52.23}$$

52.2 Lagrangian mechanics

52.2.1 Action

Definition 52.2.1 (Generalized coordinates). The generalized coordinates are mutually independent coordinates that completely characterize the configuration of a system (relative to a reference configuration).

When a system is characterized by N parameters and n_c constraints, there are $N - n_c$ generalized coordinates. Furthermore, every set of generalized coordinates describing the same system contains exactly $N - n_c$ coordinates. (In Chapter 54, it is explained how the relevant degrees of freedom can be extracted.)

Definition 52.2.2 (Generalized velocities). The generalized velocities $\dot{q}^k$ are the derivatives of the generalized coordinates q^k with respect to time.

Definition 52.2.3 (Conjugate momentum).

$$p_k := \frac{\partial L}{\partial \dot{q}^k} \tag{52.24}$$

Notation 52.2.4. Given a Lagrangian function $L : \mathbb{R}^{2n+1} \to \mathbb{R}$, depending on n generalized coordinates and their associated velocities, the following shorthand notation is often used:

$$L(q, \dot{q}, t) \equiv L\big(q_1(t), \dots, q_n(t), \dot{q}_1(t), \dots, \dot{q}_n(t), t\big)\,. \tag{52.25}$$

Definition 52.2.5 (Action). Given a Lagrangian function L, the action is a functional on the space of paths in configuration space defined by integrating L:

$$S[q] := \int_{t_1}^{t_2} L(q, \dot{q}, t)\, dt\,. \tag{52.26}$$

Especially in continuum (fluid) physics, the total time derivative receives a proper name.

Definition 52.2.6 (Lagrangian derivative[3]). Consider a quantity $A(q, \dot{q}, t)$. The total time derivative along a path $\big(q(t), \dot{q}(t)\big)$ is given by

$$\begin{aligned} \frac{\mathrm{D}A}{\mathrm{D}t} :\equiv \left.\frac{\mathrm{d}A}{\mathrm{d}t}\right|_{\left(q(t),\dot{q}(t)\right)} &= \lim_{\Delta t \to 0} \frac{A(q + \Delta q, \dot{q} + \Delta \dot{q}, t + \Delta t) - A(q, \dot{q}, t)}{\Delta t} \\ &= \frac{\partial A}{\partial t} + \dot{q}^i \frac{\partial A}{\partial q^i} + \frac{\mathrm{d}\dot{q}^i}{\mathrm{d}t} \frac{\partial A}{\partial \dot{q}^i} \,. \end{aligned} \tag{52.27}$$

The second term $\dot{\vec{\mathbf{q}}} \cdot \nabla a$ in this equation is called the **advective term**. In most treatments of continuum mechanics, the last term is not included, since one usually follows particles through space and not phase space.

Remark 52.2.7 (Convection). When $A(q, \dot{q}, t)$ is a tensor field, the gradient ∇ has to be replaced by the covariant derivative. The advective term is then sometimes called the **convective** term.

Example 52.2.8. For $A(q, \dot{q}, t) = \vec{\mathbf{q}}(t)$, one obtains

$$\frac{\mathrm{D}\vec{\mathbf{q}}}{\mathrm{D}t} = \dot{\vec{\mathbf{q}}} \,. \tag{52.28}$$

52.2.2 Euler–Lagrange equations

Axiom 52.2 (d'Alembert's principle).

$$\sum_i (\vec{\mathbf{F}}_i - \dot{\vec{\mathbf{p}}}_i) \cdot \delta \vec{\mathbf{q}}^i = 0 \,, \tag{52.29}$$

where the $\delta \vec{\mathbf{q}}^i$ denote the **virtual displacement** vectors, i.e. the infinitesimal variations consistent with the contraints. In the spirit of calculus of variations (Chapter 40), these are defined as follows. Consider a trajectory $\vec{\mathbf{q}} : [0, 1] \to \mathbb{R}^n$ and a variation $\vec{\gamma}$ of $\vec{\mathbf{q}}$, i.e. a smooth function $\vec{\gamma} : [0, 1] \times [-\varepsilon, \varepsilon] \to \mathbb{R}^n$ such that $\vec{\gamma}(t, 0) = \vec{\mathbf{q}}(t)$. This function encodes how $\vec{\mathbf{q}}$ can vary given the constraints of the system. The virtual displacement vector is then defined as the tangent vector

$$\delta \vec{\mathbf{q}} := \left.\frac{\mathrm{d}\vec{\gamma}}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} \,. \tag{52.30}$$

Formula 52.2.9 (Euler–Lagrange equation of the first kind).

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial \dot{q}^k}\right) - \frac{\partial T}{\partial q^k} = Q_k \,, \tag{52.31}$$

[3] Also known as the **material derivative**.

where T is the total kinetic energy and Q_k are the **generalized forces**:

$$Q_k := \sum_i \vec{\mathbf{F}}_i \cdot \frac{\partial \vec{\mathbf{r}}_i}{\partial q^k} \,. \tag{52.32}$$

Note that the constraint forces do not contribute to this quantity because they are always perpendicular to the motion.

Remark 52.2.10 (Conservative forces). Recall Definition 52.1.2. With the above definition of generalized forces, it is clear that a force is conservative if and only if its induced generalized force defines a closed 1-form.

Formula 52.2.11 (Euler–Lagrange equation of the second kind).

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}^k}\right) - \frac{\partial L}{\partial q^k} = 0 \tag{52.33}$$

Two proofs are provided:

Proof (Proof based on d'Alembert's principle). In the following derivation, the mass is assumed to be constant.

$$\begin{aligned}
&\sum_k \left(\vec{\mathbf{F}}_k - \dot{\vec{\mathbf{p}}}_k\right) \cdot \dot{\vec{\mathbf{r}}}_k = 0 \\
\iff &\sum_k \left(\vec{\mathbf{F}}_k - \dot{\vec{\mathbf{p}}}_k\right) \cdot \left(\sum_l \frac{\partial \vec{\mathbf{r}}_k}{\partial q^l} \dot{q}^l\right) = 0 \\
\iff &\sum_l \left(\sum_k \vec{\mathbf{F}}_k \cdot \frac{\partial \vec{\mathbf{r}}_k}{\partial q^l} - \sum_k m \ddot{\vec{\mathbf{r}}}_k \cdot \frac{\partial \vec{\mathbf{r}}_k}{\partial q^l}\right) \dot{q}^l = 0 \\
\iff &\sum_l \left(Q_l - \sum_k m \ddot{\vec{\mathbf{r}}}_k \cdot \frac{\partial \vec{\mathbf{r}}_k}{\partial q^l}\right) \dot{q}^l = 0 \,.
\end{aligned}$$

Now, consider the following derivative:

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\left(\dot{\vec{\mathbf{r}}} \cdot \frac{\partial \vec{\mathbf{r}}}{\partial q^l}\right) = \ddot{\vec{\mathbf{r}}} \cdot \frac{\partial \vec{\mathbf{r}}}{\partial q^l} + \dot{\vec{\mathbf{r}}} \cdot \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \vec{\mathbf{r}}}{\partial q^l}\right) \\
\iff &\ddot{\vec{\mathbf{r}}} \cdot \frac{\partial \vec{\mathbf{r}}}{\partial q^l} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\dot{\vec{\mathbf{r}}} \cdot \frac{\partial \vec{\mathbf{r}}}{\partial q^l}\right) - \dot{\vec{\mathbf{r}}} \cdot \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \vec{\mathbf{r}}}{\partial q^l}\right) \\
\iff &\ddot{\vec{\mathbf{r}}} \cdot \frac{\partial \vec{\mathbf{r}}}{\partial q^l} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\dot{\vec{\mathbf{r}}} \cdot \underbrace{\frac{\partial \vec{\mathbf{r}}}{\partial q^l}}_{A}\right) - \dot{\vec{\mathbf{r}}} \cdot \frac{\partial \dot{\vec{\mathbf{r}}}}{\partial q^l} \,.
\end{aligned}$$

To evaluate the factor indicated by A, one can consider another derivative:

$$\begin{aligned}\frac{\partial \dot{\vec{\mathbf{r}}}}{\partial \dot{q}^l} &= \frac{\partial}{\partial \dot{q}^l}\left(\sum_k \frac{\partial r}{\partial q^k}\dot{q}^k\right)\\ &= \sum_k \frac{\partial r}{\partial q^k}\delta_{kl}\\ &= \frac{\partial \vec{\mathbf{r}}}{\partial q^l}\\ &= A\,.\end{aligned}$$

Substituting this in the previous equation gives

$$\begin{aligned}\ddot{\vec{\mathbf{r}}}\cdot\frac{\partial \vec{\mathbf{r}}}{\partial q^l} &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\dot{\vec{\mathbf{r}}}\cdot\frac{\partial \dot{\vec{\mathbf{r}}}}{\partial \dot{q}^l}\right) - \dot{\vec{\mathbf{r}}}\cdot\left(\frac{\partial \dot{\vec{\mathbf{r}}}}{\partial q^l}\right)\\ &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}\frac{\partial \dot{\vec{\mathbf{r}}}^2}{\partial \dot{q}^l}\right) - \frac{1}{2}\frac{\partial \dot{\vec{\mathbf{r}}}^2}{\partial q^l}\,.\end{aligned}$$

If one multiplies this by the mass m and sums over all masses, the following expression is obtained:

$$\begin{aligned}\sum_k m_k\ddot{\vec{\mathbf{r}}}_k\cdot\frac{\partial \vec{\mathbf{r}}_k}{\partial q^l} &= \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial}{\partial \dot{q}^l}\left(\sum_k \frac{1}{2}m\dot{\vec{\mathbf{r}}}_k^2\right) - \frac{\partial}{\partial q^l}\left(\sum_k \frac{1}{2}m\dot{\vec{\mathbf{r}}}_k^2\right)\\ &= \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial T}{\partial \dot{q}^l} - \frac{\partial T}{\partial q^l}\end{aligned}$$

where the total kinetic energy is denoted by T in the last line. Plugging this result into the first equation gives

$$\sum_l\left(Q_l - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial T}{\partial \dot{q}^l} - \frac{\partial T}{\partial q^l}\right)\dot{q}^l = 0\,.$$

Because all the coordinates q^l are independent, the following relation should hold for all l:

$$\begin{aligned}&Q_l - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial \dot{q}^l}\right) - \frac{\partial T}{\partial q^l} = 0\\ \Longleftrightarrow\ &\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial \dot{q}^l}\right) - \frac{\partial T}{\partial q^l} = Q_l\,.\end{aligned}$$

This last equation is the Euler–Lagrange equation of the first kind. If the system only contains conservative forces, the force on the ith mass can be written as

$$F_i = -\nabla_i V\,.$$

With this in mind, one can relate the partial derivatives of the potential to the generalized forces:

$$\begin{aligned}\frac{\partial V}{\partial q^l} &= \sum_i \left(\nabla_i V\right) \cdot \frac{\partial \vec{\mathbf{r}}_i}{\partial q^l} \\ &= -Q_l \,.\end{aligned}$$

Furthermore, the derivative of V with respect to the generalized velocities vanishes. This, combined with the Euler–Lagrange equation of the first kind, gives

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial \dot{q}^l}\right) - \frac{\partial T}{\partial q^l} = Q_l \\ \Longleftrightarrow\ &\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial \dot{q}^l}\right) - \frac{\partial T}{\partial q^l} = -\frac{\partial V}{\partial q^l} + \frac{\partial V}{\partial \dot{q}^l} \\ \Longleftrightarrow\ &\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial T}{\partial \dot{q}^l} - \frac{\partial V}{\partial \dot{q}^l}\right) - \frac{\partial}{\partial q^l}(T - V) = 0 \,.\end{aligned}$$

If one introduces the Lagrangian $L := T - V$, one gets the Euler–Lagrange equation of the second kind:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}^l}\right) - \frac{\partial L}{\partial q^l} = 0 \,.$$

□

Proof (Proof based on the principle of least action). First, recall Definition 52.2.5 of the action:

$$S[q] := \int_{t_1}^{t_2} L\big(q(t), \dot{q}(t), t\big)\, dt \,.$$

The principle of least action (also called **Hamilton's principle**) postulates that the action is minimal for the physically relevant path. To this end, define a family of paths

$$q(t, \alpha) = q(t) + \alpha \eta(t) \,,$$

where $\eta(t)$ is an arbitrary function satisfying the following boundary conditions:

$$\begin{cases} \eta(t_1) = 0 \,, \\ \eta(t_2) = 0 \,. \end{cases}$$

If the action integral is extended to such a family, the integral becomes a function

of α:

$$S(\alpha) := \int_{t_1}^{t_2} L\big(q(t,\alpha), \dot{q}(t,\alpha), t\big)\, dt\,.$$

Requiring that the action integral is stationary for the physical path $q(t)$, where $\alpha = 0$, is equivalent to requiring that the derivative at $\alpha = 0$ vanishes:

$$\left.\frac{\mathrm{d}S}{\mathrm{d}\alpha}\right|_{\alpha=0} = 0\,.$$

As one evaluates this derivative at $\alpha = 0$, $q(t,\alpha)$ coincides with $q(t)$ in this expression:

$$\begin{aligned}\frac{\mathrm{d}S}{\mathrm{d}\alpha} &= \int_{t_1}^{t_2} \sum_l \left[\frac{\partial L}{\partial q^l}\frac{\partial q^l}{\partial \alpha} + \frac{\partial L}{\partial \dot{q}^l}\frac{\partial \dot{q}^l}{\partial \alpha}\right] dt \\ &= \int_{t_1}^{t_2} \sum_l \left[\frac{\partial L}{\partial q^l}\eta^l(t) + \frac{\partial L}{\partial \dot{q}^l}\dot{\eta}^l(t)\right] dt\,.\end{aligned}$$

By applying integration by parts to the second term in this integral, one obtains

$$\begin{aligned}\frac{\mathrm{d}S}{\mathrm{d}\alpha} &= \int_{t_1}^{t_2} \sum_l \left[\frac{\partial L}{\partial q^l}(t)\eta^l(t) + \frac{\partial L}{\partial \dot{q}^l}(t)\dot{\eta}^l(t)\right] dt \\ &= \int_{t_1}^{t_2} \sum_l \frac{\partial L}{\partial q^l}(t)\eta^l(t)\, dt + \frac{\partial L}{\partial \dot{q}^l}(t_2)\eta^l(t_2) - \frac{\partial L}{\partial \dot{q}^l}(t_1)\eta^l(t_1) - \int_{t_1}^{t_2} \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}^l}\right)\eta^l(t)\, dt\,.\end{aligned}$$

Due to the initial conditions for the function η, the second and third term vanish:

$$\frac{\mathrm{d}S}{\mathrm{d}\alpha} = \int_{t_1}^{t_2} \sum_l \left[\frac{\partial L}{\partial q^l} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}^l}\right)\right]\eta^l(t)\, dt\,.$$

Furthermore, because the function η was arbitrary, the only possible way that this derivative can be zero is when the integrand itself is zero:

$$\frac{\partial L}{\partial q^l} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}^l}\right) = 0\,.$$

Comparing this result to Eq. (52.33) shows that one can also obtain the Lagrangian equations of the second kind by starting from the principle of least action. □

Definition 52.2.12 (Cyclic coordinate). If the Lagrangian L does not explicitly depend on a coordinate q^k, the coordinate is said to be cyclic.

Theorem 52.2.13 (Noether). *The conjugate momentum of a cyclic coordinate is a conserved*

quantity:

$$\dot{p}_k \overset{52.2.3}{=} \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}^k}\right) \overset{(52.33)}{=} \frac{\partial L}{\partial q^k} = 0\,. \tag{52.34}$$

52.2.3 Kepler problem

Formula 52.2.14 (Gravitational potential for a point mass).

$$V = -G\frac{M}{r}\,, \tag{52.35}$$

where $G = 6,67\mathrm{e}{-11}\frac{\mathrm{Nm}^2}{\mathrm{kg}^2}$ is the **gravitational constant**.

To solve the Kepler problem, it is useful to perform a coordinate transformation. Simultaneously, one passes to the **center-of-mass reference frame**

$$\vec{\mathbf{R}} := \frac{m_1\vec{\mathbf{r}}_1 + m_2\vec{\mathbf{r}}_2}{m_1 + m_2} \qquad M := m_1 + m_2 \tag{52.36}$$

and also introduces the **reduced mass**:

$$\vec{\mathbf{r}} := \vec{\mathbf{r}}_1 - \vec{\mathbf{r}}_2 \qquad \mu := \frac{m_1 m_2}{m_1 + m_2}\,. \tag{52.37}$$

With these coordinates, the Kepler Lagrangian becomes:

$$L_{\text{Kepler}} = \frac{1}{2}M\dot{R}^2 + \frac{1}{2}\mu\dot{r}^2 + \frac{GM}{r}\,. \tag{52.38}$$

The first consequence of this transformation is that the equations of motion decouple. The center of mass behaves as a free particle:

$$M\ddot{\vec{\mathbf{R}}} = 0\,, \tag{52.39}$$

whereas the displacement vector and reduced mass are those of a particle influenced by a gravitational force:

$$\mu\ddot{\vec{\mathbf{r}}} = \frac{GM}{r^2}\hat{\mathbf{e}}_{\mathbf{r}}\,. \tag{52.40}$$

Theorem 52.2.15 (Bertrand). *The only central force problems with closed, bounded orbits, are given by the inverse potential*

$$V(r) \sim \frac{1}{r} \tag{52.41}$$

and harmonic potential

$$V(r) \sim r^2\,. \tag{52.42}$$

Definition 52.2.16 (Laplace–Runge–Lenz vector). Although all conservative central force problems have some common symmetries (time translations and rotations), the central force problems with a $1/r$-potential have an extra symmetry. However, in contrast to the ordinary Euclidean symmetries, this new symmetry is harder to understand. It is a dynamical symmetry (Definition 37.4.6) instead of a kinematical one (Definition 37.4.5), i.e. it cannot simply be deduced from the Lagrangian.

Before explaining the symmetry itself, the associated Noether charge (Theorem 37.4.4), the Laplace–Runge–Lenz vector, is given:

$$\vec{\mathbf{A}} := \vec{\mathbf{p}} \times \vec{\mathbf{L}} - mk\vec{\mathbf{r}}\,, \tag{52.43}$$

where m denotes the mass and k fixes the scale of the potential:

$$V(r) = \frac{k}{r}\,. \tag{52.44}$$

The reason why this conserved quantity cannot be obtained from a cyclic coordinate of the ordinary Lagrangian is that it is actually associated to a 4-dimensional problem. The central problem with an inverse-square force law can be reformulated as the free motion of a particle in 4 dimensions and only in this description can the Noether charge be obtained from a cyclic coordinate.

52.3 Hamiltonian mechanics

Definition 52.3.1 (Canonical coordinates). Consider the generalized coordinates $(q, \dot{q}, t)$. From these, one can derive a new set of coordinates, called canonical coordinates, by exchanging the time derivatives $\dot{q}^i$ in favour of the conjugate momenta p_i.

This is only equivalent if the transformation $(q, \dot{q}) \leftrightarrow (q, p)$ is invertible. A sufficient condition is that the Hessian of L with respect to the generalized velocities is invertible.

Definition 52.3.2 (Phase space). The set of all possible $2n$-tuples (q^i, p_i) of generalized coordinates and associated momenta. Note that this set also includes the tuples that are not solutions of the equations of motion.

Definition 52.3.3 (Libration). A closed trajectory in phase space for which the coordinates take on only a subset of the allowed values. It is the generalization of an oscillation. Topologically, it is characterized by a contractible, closed trajectory.

Definition 52.3.4 (Rotation). A closed trajectory in phase space for which at least one of the variables takes on all possible values. Topologically, it is characterized by a non-contractible, closed trajectory.

@@ CORRECT (for an ordinary rotation in spherical coordinates, this indeed holds. However, for rotations in Cartesian coordinates, this fails) @@

Definition 52.3.5 (Separatrix). When plotting (closed) trajectories in phase space, the curve that separates regions of librations and rotations is called the separatrix.[4]

Definition 52.3.6 (Hamiltonian function). Given a Lagrangian L, the Hamiltonian function is defined by the following Legendre transformation (Definition 16.7.7):

$$H(q,p,t) := \sum_i p_i \dot{q}^i - L(q,p,t)\,. \tag{52.45}$$

Formula 52.3.7 (Hamilton's equations). Inserting the above definition in the action (Definition 52.2.5) and applying the variational principle results in the following equations:

$$\begin{cases} \dot{q}^k = \dfrac{\partial H}{\partial p_k}\,, \\ \dot{p}_k = -\dfrac{\partial H}{\partial q^k}\,. \end{cases} \tag{52.46}$$

Systems obeying these equations are called **Hamiltonian systems**. (See Section 37.4 for the mathematical background).

52.3.1 Poisson bracket

Definition 52.3.8 (Poisson bracket). To stick with the conventions of Definition 37.2.1, the Poisson bracket is defined as

$$\{A,B\} := \sum_k \frac{\partial A}{\partial p_k}\frac{\partial B}{\partial q^k} - \frac{\partial A}{\partial q^k}\frac{\partial B}{\partial p_k}\,, \tag{52.47}$$

where q, p are the canonical coordinates in the Hamiltonian formalism.

Remark. As noted before, some authors define the Poisson bracket with the opposite sign. One should always pay attention to which convention is used.

Formula 52.3.9 (Total time derivative). Hamilton's equations imply the following expression for the total time derivative:

$$\frac{\mathrm{d}F}{\mathrm{d}t} = \frac{\partial F}{\partial t} + \{H,F\}\,, \tag{52.48}$$

which is the evolution equation for a nonautonomous system (Definition 37.4.2).

Definition 52.3.10 (Liouville operator). The operator $\hat{L} : C^\infty(M) \to C^\infty(M)$ defined as follows:

$$\hat{L}f := \{H,f\}\,. \tag{52.49}$$

[4] In general, the separatrix of a dynamical system is a curve that separates regions with different behaviour.

52.3.2 Hamilton–Jacobi equation

For a geometrical interpretation, see Section 37.4.2.

Definition 52.3.11 (Canonical transformations). A transformation that leaves the Hamiltonian equations of motion unchanged. This means that the transformation leaves the action invariant up to a constant or, equivalently, it leaves the Lagrangian invariant up to a complete time-derivative:

$$\sum_i p_i \dot{q}^i - H(q,p,t) = \sum_i P_i \dot{Q}^i - K(Q,P,t) - \frac{\mathrm{d}S}{\mathrm{d}t}(Q,P,t)\,. \tag{52.50}$$

The function S is called the **generating function** of the canonical transformation. The choice of generating function uniquely determines the transformation (the converse is, however, not true).

Formula 52.3.12 (Hamilton–Jacobi equation). Sufficient conditions for a function S to be a generating function of canonical transformations are:

$$\begin{aligned} p_i &= \frac{\partial S}{\partial q^i}\,, \\ Q^i &= \frac{\partial S}{\partial P_i}\,, \end{aligned} \tag{52.51}$$

and

$$K = H + \frac{\partial S}{\partial t}\,. \tag{52.52}$$

Choosing the new Hamiltonian function K, sometimes called the **Kamiltonian**, to be 0 gives the Hamilton–Jacobi equation:

$$H\left(q, \frac{\partial S}{\partial q}, t\right) + \frac{\partial S}{\partial t} = 0\,. \tag{52.53}$$

In this case the function S is called **Hamilton's principal function**.

Property 52.3.13. The new coordinates Q^i and P_i are all constants of motion. This immediately follows from the choice $K = 0$.

Definition 52.3.14 (Hamilton's characteristic function). For time-independent systems, the Hamilton–Jacobi equation can be rewritten as follows:

$$H\left(q, \frac{\partial S}{\partial q}\right) = -\frac{\partial S}{\partial t} =: E\,. \tag{52.54}$$

After a redefinition of H, this is the same as Eq. (37.55). One thus obtains the classical result that, for time-independent systems, the Hamiltonian function is a constant of motion (often called the **energy**[5]). Integration with respect to time gives the following form of the principal function:

$$S(q,p,t) = W(q,p) - Et\,. \tag{52.55}$$

The time-independent function W is called Hamilton's characteristic function.

[5]Note that, in general, this is not the physical energy.

Property 52.3.15 (Stäckel condition). The Hamilton–Jacobi equation is separable if and only if the potential is of the form

$$V(q) = \sum_{i=1}^{n} \frac{W_i(q^i)}{G_i^2(q)} \tag{52.56}$$

whenever the Hamiltonian function can be written as

$$H(q,p) = \frac{1}{2}\sum_i \frac{p_i^2}{G_i^2(q)} + V(q)\,. \tag{52.57}$$

Potentials of this form are called **Stäckel potentials**.

52.4 Continuum mechanics

52.4.1 Liouville's theorem

Let $\rho : \mathbb{R}^{2n} \to \mathbb{R}_+$ be the phase space distribution function of some extensive quantity F and Vol be the canonical measure, i.e. the quantity F evaluated on the volume $V \subseteq \mathbb{R}^{2n}$ is given by

$$F(V) = \int_V \rho\,\mathsf{Vol} \equiv \int_V \rho\, dV\,. \tag{52.58}$$

Formula 52.4.1 (Liouville's lemma). Consider a time-dependent phase space volume $V : \mathbb{R} \to \mathcal{B}(\mathbb{R}^{2n})$, evolving along a path $x(t) \equiv \big(q(t), \dot{q}(t)\big)$. The Jacobian $J\big(x(t)\big)$ associated with this motion is given by

$$J\big(x(t)\big) = \det\left(\frac{\partial x(t)}{\partial x(0)}\right) = \sum_{ijklmn} \varepsilon^{ijklmn} \frac{\partial x^1}{\partial x_0^i}\frac{\partial x^2}{\partial x_0^j}\frac{\partial x^3}{\partial x_0^k}\frac{\partial x^4}{\partial x_0^l}\frac{\partial x^5}{\partial x_0^m}\frac{\partial x^6}{\partial x_0^n}\,. \tag{52.59}$$

The Lagrangian derivative of this Jacobian is

$$\frac{\mathrm{D}J}{\mathrm{D}t} = (\nabla\cdot\dot{x})J\,, \tag{52.60}$$

where ∇ denotes the gradient in phase space. Flows for which the right-hand side vanishes, i.e. for which $\nabla\cdot\dot{x} = 0$, are said to be **incompressible**.

Formula 52.4.2 (Reynolds' transport theorem). Consider a quantity

$$F(t) := \int_{V(t)} f(q,\dot{q},t)\, dV\,, \tag{52.61}$$

where $V : \mathbb{R} \to \mathcal{B}(\mathbb{R}^{2n})$ is a time-dependent volume in phase space. Combining Eq. (52.60) with the divergence theorem 23.2.6 gives

$$\frac{\mathrm{D}F}{\mathrm{D}t} = \int_{V(t)} \left(\frac{\partial f}{\partial t} + \nabla\cdot(f\dot{\vec{\mathbf{q}}})\right) dV = \int_{V(t)} \frac{\partial f}{\partial t}\, dV + \oint_{\partial V(t)} f\dot{\vec{\mathbf{q}}}\cdot d\vec{\mathbf{S}}\,, \tag{52.62}$$

where $\dot{\vec{\mathbf{q}}}$ denotes the velocity of the infinitesimal area element. This formula can be interpreted as a three-dimensional generalization of the Leibniz integral rule (Formula 16.6.5).

Definition 52.4.3 (Material volume). A time-dependent volume $V : \mathbb{R} \to \mathcal{B}(\mathbb{R}^{2n})$ such that the velocity $\dot{\vec{\mathbf{q}}}$ at every point matches the flow velocity.

Remark. In most parts of the literature, Reynolds' transport theorem is stated for material volumes. For simplicity, this will be assumed from here on.

Corollary 52.4.4. Note that Reynolds' transport theorem can be rewritten as

$$\frac{\mathrm{D}F}{\mathrm{D}t} = \int_{V(t)} \left(\frac{\mathrm{D}f}{\mathrm{D}t} + f \nabla \cdot \dot{\vec{\mathbf{q}}} \right) dV \,. \tag{52.63}$$

Hence, for incompressible fluids, the material derivative can be taken inside the integral.

Formula 52.4.5 (Continuity equations). For a conserved quantity F, Reynolds' transport theorem becomes:

$$\frac{\partial f}{\partial t} + \nabla \cdot (f \dot{\vec{\mathbf{q}}}) = 0 \tag{52.64}$$

or, after using Definition 52.2.6,

$$\frac{\mathrm{D}f}{\mathrm{D}t} + (\nabla \cdot \dot{\vec{\mathbf{q}}}) f = 0 \,. \tag{52.65}$$

If one sets $f = \rho$ (the phase space density), the equations are called the **Eulerian continuity equation** and **Lagrangian continuity equation**, respectively.

The difference between these two equations corresponds to the way the system is observed. In the Eulerian approach, one observes a fixed point in space and measures how a given quantity at that point evolves. In the Lagrangian approach, one observes a fixed element in the system and measures how a given quantity evolves around the chosen point as it moves throughout space.

Corollary 52.4.6. Combining Reynolds' transport theorem with the continuity equation, and assuming that the velocity of $V(t)$ matches the flow velocity, gives the following identity for an arbitrary function $f : \mathbb{R}^{2n+1} \to \mathbb{R}$:

$$\frac{\mathrm{D}}{\mathrm{D}t} \int_{V(t)} \rho f \, dV = \int_{V(t)} \rho \frac{\mathrm{D}f}{\mathrm{D}t} \, dV \,. \tag{52.66}$$

Although this result does imply that the total probability in any material volume is preserved, the preservation of the probability density function requires a slightly stronger assumption.

Theorem 52.4.7 (Liouville). *If the flow is incompressible, the density is constant along the flow:*[6]

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} = 0 \,. \tag{52.67}$$

[6] This equation is also called the **Liouville equation**.

Moreover, the volume of any material volume is also constant:

$$\frac{\mathrm{D}\,\mathrm{Vol}(V)}{\mathrm{D}t} = 0 \tag{52.68}$$

for all material volumes $V : \mathbb{R} \to \mathcal{B}(\mathbb{R}^{2n})$.

Remark 52.4.8 (♣). By Section 37.4, every phase space admits a symplectic form ω and the induced canonical volume form on phase space is $\mathrm{Vol}_\omega \sim \omega^n$. Accordingly, the above theorem (more specifically, the second part) easily follows from the fact that time evolution is the flow of the Hamiltonian vector field, which preserves the symplectic form.

Formula 52.4.9 (Boltzmann's transport equation). Let $F(q, \dot{q}, t)$ be the mass distribution function:

$$M_{\text{tot}} = \int_{V(t)} F(q, \dot{q}, t)\, dV\,. \tag{52.69}$$

From the conservation of mass, one can derive the following formula:

$$\frac{\mathrm{D}F}{\mathrm{D}t} = \frac{\partial F}{\partial t} + \dot{q}^i \cdot \frac{\partial F}{\partial q^i} - \nabla V^i \frac{\partial F}{\partial \dot{q}^i} = \left[\frac{\partial F}{\partial t}\right]_{\text{col}}, \tag{52.70}$$

where the right-hand side gives the change of F due to collisions.[7] This partial differential equation in 7 variables can be solved to obtain F.

A prime example of flows are given by time evolution in a Hamiltonian system. The Hamiltonian equations of motion can be shown to imply the incompressibility of the Hamiltonian flow and, accordingly, by Liouville's theorem, this flow gives rise to a volume- (or measure-)preserving map $g : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ (Definition 18.1.39). This gives rise to the following theorem.

Theorem 52.4.10 (Poincaré recurrence theorem). *For every point $x_0 \in \mathbb{R}^{2n}$ and for every neighbourhood $U \subseteq \mathbb{R}^{2n}$ of x_0, there exists a point $y \in U$ such that $g^n(y) \in U$ for every $n \in \mathbb{N}$.*

Theorem 52.4.11 (Strong Jeans's theorem[8]). *In a time-independent system for which almost all orbits are regular, the distribution function can be expressed in terms of three integrals of motion.*

The constants in Jeans's theorem are called the **isolating integrals** of the system.

52.4.2 Fluid mechanics

@@ ADD (e.g. stress vector) @@

[7] The collisionless form of this equation is sometimes called the **Vlasov equation**.
[8] Actually due to *Lynden-Bell*.

Theorem 52.4.12 (Cauchy's stress theorem[9]). *The stress vectors acting on the coordinate planes through a point determine the stress vector acting on an arbitrary plane passing through that point.*

Cauchy's stress theorem is equivalent to the existence of the following tensor.

Definition 52.4.13 (Cauchy stress tensor). The Cauchy stress tensor is a $(0,2)$-tensor $\mathbf{T}$ that gives the relation between a stress vector associated to a plane and the normal vector $\vec{\mathbf{n}}$ to that plane:

$$\vec{\mathbf{t}}_{(\vec{\mathbf{n}})} = \mathbf{T}(\vec{\mathbf{n}})\,. \tag{52.71}$$

Example 52.4.14. For identical particles, the stress tensor is given by

$$\mathbf{T} = -\rho\langle \vec{\mathbf{w}} \otimes \vec{\mathbf{w}}\rangle\,, \tag{52.72}$$

where $\vec{\mathbf{w}}$ is the random component of the velocity vector and $\langle\cdot\rangle$ denotes the expectation value (Definition 45.3.17).

Theorem 52.4.15 (Cauchy's lemma). *The stress vectors acting on opposite planes are equal in magnitude but opposite in direction:*

$$\vec{\mathbf{t}}_{(-\vec{\mathbf{n}})} = -\vec{\mathbf{t}}_{(\vec{\mathbf{n}})}\,. \tag{52.73}$$

Formula 52.4.16 (Cauchy momentum equation). From Newton's second law (Axiom 52.1), it follows that

$$\frac{\mathrm{D}\vec{\mathbf{P}}}{\mathrm{D}t} = \int_V \vec{\mathbf{f}}(x,t)\,dV + \oint_{\partial V} \vec{\mathbf{t}}(x,t)\,dS\,, \tag{52.74}$$

where $\vec{\mathbf{P}}$ is the momentum density, $\vec{\mathbf{f}}$ are the 'body' forces and $\vec{\mathbf{t}}$ are the 'surface' forces (such as *shear stress*). Using Cauchy's stress theorem and the divergence theorem 23.2.6, one obtains

$$\frac{\mathrm{D}\vec{\mathbf{P}}}{\mathrm{D}t} = \int_V \left[\vec{\mathbf{f}}(x,t) + \nabla\cdot\mathbf{T}(x,t)\right] dV\,. \tag{52.75}$$

The left-hand side can be rewritten using Eq. (52.66) as

$$\int_V \rho\frac{\mathrm{D}\vec{\mathbf{v}}}{\mathrm{D}t}\,dV = \int_V \left[\vec{\mathbf{f}}(x,t) + \nabla\cdot\mathbf{T}(x,t)\right] dV\,. \tag{52.76}$$

52.5 Koopman–von Neumann mechanics

After the invention and formalization of quantum mechanics (see Part IX), some authors tried to reformulate classical mechanics in terms of operators and Hilbert spaces.

[9] Also known as **Cauchy's fundamental theorem**.

A first step towards this goal was made by, amongst others, *Koopman* and *von Neumann*, who expressed the measure-preserving transformations (i.e. the canonical transformation) on phase space as operators, which form a Hilbert space under the Hilbert–Schmidt inner product (Formula 22.4.7). However, notwithstanding this reformulation, the content of this section is actually of a different nature and the historical attribution is, therefore, incorrect. The idea of this section is to rephrase the classical probability distribution of phase space as a wave function.

Recall the Liouville equation (Theorem 52.4.7)

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} = 0 \tag{52.77}$$

for the phase space density function. Using Definition 52.3.10, this can be rewritten as

$$\frac{\partial\rho}{\partial t} = -\hat{L}\rho\,. \tag{52.78}$$

Definition 52.5.1 (Koopman–von Neumann wave function). The KvN wave function $\psi : \mathbb{R}^{2n+1} \to \mathbb{C}$ is required to factorize the density function:

$$\rho = \overline{\psi}\psi \tag{52.79}$$

and satisfy the Liouville equation:

$$\frac{\partial\psi}{\partial t} = \hat{L}\psi\,. \tag{52.80}$$

An alternative approach start from the operator formulation of quantum mechanics, where the observables are represented as self-adjoint operators on a Hilbert space, as mentioned above. The axioms are as follows:

1. **Normalization**: $\langle\psi\,|\,\psi\rangle = 1$.
2. **Expectation**: $\mathsf{E}\big[A\big] = \langle\psi \mid \widehat{A} \mid \psi\rangle$.
3. **Born rule**: The probability of measuring the value $a \in \sigma(\widehat{A})$ is given by
$$p_a = |\langle a\,|\,\psi\rangle|^2\,. \tag{52.81}$$
4. **Compositionality**: The disjoint union of systems is represented by the tensor product of their wave functions.

Property 52.5.2 (Generality). The above set of axioms are, in fact, very general, in that they allow to recover both classical mechanics and quantum mechanics, depending on an additional set of relations. First of all, if Newton's second law (Axiom 52.1) is postulated to hold on average (in an Ehrenfest-like manner, cf. Theorem 57.3.3), i.e.

$$m\frac{\mathrm{d}}{\mathrm{d}t}\langle x\rangle = \langle p\rangle \qquad \text{and} \qquad \frac{\mathrm{d}}{\mathrm{d}t}\langle p\rangle = -\langle\nabla U(x)\rangle \tag{52.82}$$

for some function $U : \mathbb{R}^n \to \mathbb{R}$, the following conditions are recovered:

$$m[\hat{L}, \hat{x}] = \hat{p} \qquad \text{and} \qquad [\hat{L}, \hat{p}] = -\nabla U(\hat{x}) \,, \tag{52.83}$$

where $\hat{L}$ is the generator of time evolution according to Stone's theorem 26.1.26.

Now, if the classical commutation relation $[\hat{x}, \hat{p}] = 0$ is required to hold, the above equations cannot hold for any operator $\hat{L} \equiv L(\hat{x}, \hat{p})$. To solve this issue, an additional set of operators $\{\hat{\lambda}_x, \hat{\lambda}_p\}$ has to be introduced. These are required to form a Lie algebra, called the **KvN algebra**, with the following commutation relations:

$$\begin{aligned} [\hat{\lambda}_x, \hat{x}] = [\hat{\lambda}_p, \hat{p}] = i \,, \\ [\hat{x}, \hat{p}] = [\hat{\lambda}_x, \hat{p}] = [\hat{\lambda}_p, \hat{x}] = [\hat{\lambda}_x, \hat{\lambda}_p] = 0 \,. \end{aligned} \tag{52.84}$$

A form of $\hat{L}$ that makes the above conditions hold is:

$$-i\hat{L} = \frac{\hat{p}}{m}\hat{\lambda}_x - \nabla U(\hat{x})\hat{\lambda}_p \,. \tag{52.85}$$

Using this expression, the classical rules of statistical mechanics can be recovered.

To recover *quantum mechanics* (see Chapter 57), the *canonical commutation relations* (see Definition 57.4.1) are postulated: $[\hat{x}, \hat{p}] = i$. In this case, no additional operators have to be introduced and (up to a factor $\hbar$), the Liouville operator can be shown to coincide with standard Hamiltonian:

$$\hbar\hat{L} = \frac{\hat{p}^2}{2m} + U(\hat{x}) \,, \tag{52.86}$$

for some function $U : \mathbb{R}^n \to \mathbb{R}$.

Remark 52.5.3 (Linearity). Note that, in contrast to the *Schrödinger equation* (see, e.g., Eq. (57.10)), both the Liouville and KdV equations are linear PDEs in classical mechanics.

@@ COMPLETE @@

52.6 Dynamical systems

The following property, although seemingly innocuous, is very important for the study of mechanical systems.

Property 52.6.1. For dynamical systems governed by ODEs satisfying the Picard–Lindelöf conditions (Theorem 20.2.1), different trajectories never intersect.

The above property has an important (but nontrivial) consequence.

Theorem 52.6.2 (Poincaré–Bendixson). *In a two-dimensional phase space, the only trajectories inside of a closed bounded subregion without fixed points are either closed orbits or trajectories spiralling into closed orbits.*

Corollary 52.6.3. In two-dimensional (Cartesian) phase spaces, there cannot exist chaos, i.e. no *strange attractors* can exist.

Definition 52.6.4 (Lypanuov exponents). Consider two trajectories of a system and denote the distance between them at a time t_0 by $s_0 := s(t_0)$. (The initial time can be taken to be 0 without loss of generality.) If, after some time t, these trajectories satisfy

$$s(t) \approx e^{\lambda t} s_0 \,, \tag{52.87}$$

the parameter λ is called the Lyapunov exponent of the system.

Definition 52.6.5 (Limit cycle). Consider a closed trajectory C. If there exist curves that asymptotically ($t \longrightarrow \pm\infty$) converge to C, i.e. their **limit set** is C, one calls C a limit cycle.

Definition 52.6.6 (Poincaré map). Consider a dynamical system determined by a phase flow ϕ and let S be a codimension-1 hypersurface in the phase space Q that is transversal to ϕ, i.e. all trajectories intersect S at isolated points. Intuitively, the Poincaré map $P : S \to S$ is defined as the 'map of first return', i.e. the image of every point $x \in S$, if it exists, is given by $\phi_T(x)$ with $T =: \min\{t \in \mathbb{R}^+ \mid \phi_t(x) \in S\}$.

One can give a more formal definition (one that also avoids the fact that P would only be partially defined). The Poincaré map P is defined as follows:

- $P : U \to S$ is a differentiable map, where $U \subseteq S$ is open and connected.
- $P|_{P(U)}$ is a diffeomorphism.
- For every point $u \in U$, the positive semi-orbit of u intersects S for the first time at $P(u)$.

The usefulness of this map lies in the fact that it preserves (quasi)periodicity whilst reducing the dimensionality of the space. It is especially useful for the study of 3-dimensional spaces, where the section S is 2-dimensional and, hence, easily visualized.

Property 52.6.7 (Fixed points). Fixed points of the Poincaré map correspond to closed orbits.

52.7 Geometric mechanics

52.7.1 Newton's equations

In this section, the current chapter is reformulated in a differential-geometric framework (in the spirit of Section 37.4). The first step is to reformulate ordinary Newtonian mechanics. The general setting here is a Riemannian manifold (M, g), where the metric

is mainly used for defining the kinetic energy

$$E_{\text{kin}} := \frac{1}{2} g(v, v)\,, \tag{52.88}$$

together with a second-order ODE in the form of a vector field $X \in \mathfrak{X}(TM)$ such that $\pi_*(X_v) = v$, where $\pi : TM \to M$ denotes the tangent bundle projection.

For integral curves γ in TM of second-order ODEs, it is easy to show that they are the tangent vector fields of their projections. In particular, if $q^i(t)$ are the local coordinates of a curve $\sigma := \pi(\gamma)$ on M, it can be shown that the tangent coordinates $\dot{q}^i$ of γ are exactly the derivatives of the coordinates q^i:

$$\dot{q}^i(t) = \frac{\mathrm{d}q^i}{\mathrm{d}t}(t)\,. \tag{52.89}$$

As such, the abuse of notation $\dot{q}^i$ is justified. Conversely, it can be shown that a vector field on TM is second order if and only if this is true for any local chart, i.e. if the vector field $X \in \mathfrak{X}(TM)$ can be expressed as follows:

$$X = \dot{q}^i \frac{\partial}{\partial q^i} + F^i(q, \dot{q}) \frac{\partial}{\partial \dot{q}^i}\,. \tag{52.90}$$

By writing this vector field as a system of differential equations, a second-order ODE is obtained (hence the terminology):

$$\frac{\mathrm{d}^2 q^i}{\mathrm{d}t^2} = F^i(q, \dot{q})\,. \tag{52.91}$$

The primary example of such a second-order ODE is the vector field generating the geodesic flow on TM, i.e. the integral curves are the tangent curves of geodesics on M. Therefore, it should not be surprising that the above equation is similar to Formula 30.2.47. By adopting the notation of Eq. (36.22), one can generalize the geodesic equation to obtain Newton's equations of motion for an arbitrary smooth potential $U : M \to \mathbb{R}$.

Formula 52.7.1 (Newton's equation). Let (M, g) be a Riemannian manifold and let $U : M \to \mathbb{R}$ be a smooth function. Newton's equations for a curve $\sigma : [a, b] \to M$ read

$$\nabla_{\dot{\sigma}} \dot{\sigma} = -\mathsf{grad}(U)\,, \tag{52.92}$$

where ∇ indicates the Levi-Civita connection and grad denotes the gradient operator (Definition 23.1.1), which is here not denoted by ∇ to avoid confusion with the covariant derivative.

Property 52.7.2 (Cogeodesic flow). Consider the kinetic Hamiltonian (52.88) expressed on the cotangent bundle T^*M of a Riemannian manifold:

$$H_{\text{kin}}(q, p) := \frac{1}{2} \sum_{i,j} g^{ij}(q) p_i p_j\,, \tag{52.93}$$

where g^{ij} is the inverse of the (local) matrix representative of the metric g. The integral curves of the associated Hamiltonian flow project down to geodesics on M.

52.7.2 Lagrangian mechanics

Now it is time to turn to Noether's theorem and, in particular, the version concerning cyclic coordinates (Theorem 52.2.13). Any diffeomorphism of M induces a diffeomorphism on TM by pushforward.

Definition 52.7.3 (Lagrangian symmetry). A symmetry of a Lagrangian $L : TM \to \mathbb{R}$ is a diffeomorphism ϕ of M such that $\phi^* L = L$. Infinitesimal symmetries, i.e. the generators of symmetries, are vector fields for which the associated flow is a symmetry.

Given a complete vector field X, one can define the conjugate momentum $\widehat{X} : TM \to \mathbb{R}$ as follows:

$$\widehat{X}(v) := g(X_{\pi(v)}, v)\,. \tag{52.94}$$

Using this definition, Noether's theorem 52.2.13 can be reformulated as follows.

Theorem 52.7.4 (Noether's theorem). *The conjugate momentum of an infinitesimal symmetry is a constant of motion.*

If the conjugate momenta of the coordinate-induced vector fields ∂_i are denoted by p_i, the nondegeneracy of g implies that the set $\{q^i, p_i\}_{i \leq n}$ gives well-defined coordinate functions on T^*M. The equivalence of the Lagrangian action principle and the Newtonian equations of motion imply that the second-order ODE associated to the potential U takes the following form:

$$X_U := \dot{q}^i \frac{\partial}{\partial q^i} + \frac{\partial L}{\partial q^i}\frac{\partial}{\partial p_i}\,. \tag{52.95}$$

After performing the Legendre transformation $E := p_i \dot{q}^i - L$ to obtain the (Hamiltonian) energy function, Newton's equations can be rewritten in Hamiltonian form:

$$X_E = \frac{\partial E}{\partial p_i}\frac{\partial}{\partial q^i} - \frac{\partial E}{\partial q^i}\frac{\partial}{\partial p_i}\,. \tag{52.96}$$

The function E will not be called a Hamiltonian function because this will be reserved for functions on the cotangent bundle.

52.7.3 Hamiltonian mechanics

The procedure of mapping a (complete) vector field to its conjugate momentum can be generalized to an isomorphism $TM \to T^*M$ as follows.

Definition 52.7.5 (Fibre derivative). Let $L : TM \to \mathbb{R}$ be a smooth Lagrangian. The fibre derivative of L is defined as the directional (Gâteaux) derivative (Definition 25.3.12) of L:

$$\langle \mathbb{F}L(v), w\rangle := \frac{\mathrm{d}}{\mathrm{d}t} L(v + tw)\bigg|_{t=0}\,. \tag{52.97}$$

Because $\mathbb{F}L(v) \in \mathcal{L}(TM, \mathbb{R}) = T^*M$ by definition of the derivative, one can see that $\mathbb{F}L$ defines a map $TM \to T^*M$. In local coordinates $(q^i, \dot{q}^i)$, the fiber derivative is given by

$$\mathbb{F}L : (q^i, \dot{q}^i) \mapsto \left(q^i, \frac{\partial L}{\partial \dot{q}^i}\right) \equiv (q^i, p_i) \,. \tag{52.98}$$

As such, the classical definition 52.2.3 for conjugate momenta is obtained. In the case of kinetic Lagrangians defined by a metric g, it is not hard to see that this boils down to Eq. (52.94).

Remark 52.7.6 (Legendre transform). The fibre derivative $\mathbb{F}L$ is often called the Legendre transform of L. Although this does not exactly coincide with Definition 16.7.7 or Eq. (47.17), the relation is simple enough. The Legendre transformation $L \mapsto L^*$ (on the tangent bundle) is implemented as

$$L^*(x) = \langle \mathbb{F}L(x), x \rangle - L(x) \,. \tag{52.99}$$

Lagrangians for which the Legendre transformation is invertible, i.e. for which $\mathbb{F}L$ is a diffeomorphism, give rise to equivalent mechanics encoded in the cotangent bundle. Given such a Lagrangian, we can construct the associated energy function E by a Legendre transformation and map it to a Hamiltonian H on the cotangent bundle as $H := E \circ \mathbb{F}L^{-1}$. By abuse of notation, write $L \equiv L \circ \mathbb{F}L^{-1}$. This transformation also induces a Hamiltonian vector field $X_H := \mathbb{F}L_* X_E$ on T^*M that can alternatively be encoded as $X_E = \mathbb{F}L^* X_H$. If the cotangent coordinates $p_i(\alpha) := \alpha(\partial_i)$ are chosen, it can easily be seen that $p_i \circ \mathbb{F}L \equiv p_i$. This way the Hamiltonian equations remain virtually unchanged when transporting them to the cotangent bundle.

For any choice of coordinates for which the symplectic form on T^*M takes the standard Darboux form $\omega = dp_i \wedge dq^i$, the Newtonian equations of motion take on a Hamiltonian form. If there exists a coordinate chart in which the Hamiltonian function H does not depend on any of the base coordinates q^i, the coordinates are called **action-angle variables** and the system is said to be **completely integrable** (cf. Section 37.4.3).

52.7.4 Contact structure

By extending the cotangent bundle of the configuration space with a time variable, one can also (re)obtain some interesting features. Consider the following one-form on the trivial line bundle $T^*M \times \mathbb{R}$:

$$\alpha := p_i \mathrm{d}q^i - H \mathrm{d}t \,. \tag{52.100}$$

First of all, this endows the extended cotangent bundle with a contact structure (Definition 38.1.3). Moreover, it can be shown to be a relative integral invariant (Definition 34.6.11) of the following Pfaffian system:

$$\frac{\mathrm{d}q^i}{\partial H / \partial p_i} = \mathrm{d}t = \frac{-\mathrm{d}p_i}{\partial H / \partial q^i} \,. \tag{52.101}$$

The vector fields that leave the contact form invariant are exactly those generated by the Hamiltonian vector field X_H. Liouville's theorem 52.4.7 is then simply a consequence of the absolute invariance of $\mathrm{d}\alpha$.

A canonical transformation (Definition 52.3.11) was originally defined as a transformation that leaves the Hamiltonian equations of motion invariant. In the contact setting, it can be defined as follows.

Definition 52.7.7 (Canonical transformation). A transformation of the trivial line bundle $T^*M \times \mathbb{R}$ that leaves both the fibres and the **symplectization**[10]

$$\Omega := \mathrm{d}(e^{\lambda}\alpha) \tag{52.102}$$

invariant.

@@ CHECK THIS STATEMENT (cannot find source) @@

[10] This form turns $T^*M \times \mathbb{R}^2$ into a symplectic manifold.

Chapter 53

Classical Field Theory

Rigorous definitions and statements about the mathematical concepts used in this chapter can be found in Chapters 33, 34, 36 and 40.

Chapter Contents

53.1 Lagrangian field theory **1000**

53.1.1 Noether's theorem 1001

53.2 De Donder–Weyl theory **1003**

53.1 Lagrangian field theory

The physical space will be assumed to be a (pseudo-)Riemannian, n-dimensional manifold (M, g) with the fields being sections of a vector bundle $E \to M$. In general, an **action** is a function $S : \Gamma_c(E) \to \mathbb{R}$ from the space of (compactly supported) sections of E to the real numbers. This is often given by local functionals for a local Lagrangian (Definition 40.3.2):

$$S : \Gamma_c(E) \to \mathbb{R} : \phi \mapsto \int_M (j^\infty \phi)^* L \, \mathsf{Vol}_M \, . \tag{53.1}$$

Associated to the manifold M, one can construct a cochain complex similar to the de Rham complex $\Omega^\bullet(M)$. This structure takes two geometric features into account. On the one hand, one has the ordinary de Rham differential d on the base manifold M, whereas, on the other hand, one has a differential δ induced by the variation of fields along the jet fibres. The total differential will be the sum of these as is standard in the context of bicomplexes. This defines the variational bicomplex.

Some key concepts from the calculus of variations are recalled here. The variational

derivative or Euler–Lagrange derivative (40.50) is defined as follows (partial derivatives are denoted by subscripted commas, e.g. $\partial_\mu \partial_\nu \phi \equiv \phi_{,\mu\nu}$):[1]

$$\frac{\delta L}{\delta \phi} := \frac{\partial L}{\partial \phi} - \frac{\partial}{\partial x^\mu}\left(\frac{\partial L}{\partial \phi_{,\mu}}\right) + \frac{\partial^2}{\partial x^\mu \partial x^\nu}\left(\frac{\partial L}{\partial \phi_{,\mu\nu}}\right) - \cdots . \tag{53.2}$$

By comparing this formula to the formula for the variation of the Lagrangian density, one obtains the first variational formula (40.83):

$$\delta L = \frac{\delta L}{\delta \phi^I}\delta\phi^I - \mathrm{d}\Theta[\phi] . \tag{53.3}$$

The first term vanishes on-shell because it is proportional to the Euler–Lagrange equation associated to the field ϕ^I. The last term contains the boundary terms obtained after performing integration by parts. The $(n-1,1)$-form Θ is called the **presymplectic potential**.

The **presymplectic current** ω is obtained by taking the variation of the presymplectic potential:

$$\omega[\phi] := \delta\Theta[\phi] . \tag{53.4}$$

On-shell, this form is closed, i.e. $\mathbf{d}\omega \approx 0$. Off-shell, this does not necessarily hold and, hence, the form is not properly symplectic. It can be shown that, if the variations $\delta\phi^I$ satisfy the linearised equations of motion, then for every gauge transformation ξ, there exists an $(n-2,1)$-form $k_\xi[\phi]$ such that $\omega \approx \mathrm{d}k_\xi[\phi]$.[2]

@@ EXPLAIN this last statement better @@

53.1.1 Noether's theorem

Theorem 53.1.1 (Noether's first theorem). *Consider an infinitesimal field transformation*

$$\phi \longrightarrow \phi + \alpha\delta\phi , \tag{53.5}$$

where the Lagrangian L depends on the fields and their first-order derivatives.[3] *In case of a symmetry, one obtains a conservation law of the following form:*

$$\partial_\mu\left(\frac{\partial L}{\partial \phi_{,\mu}}\delta\phi - \mathcal{J}^\mu\right) = 0 . \tag{53.6}$$

The factor between parentheses can be interpreted as a conserved current $j^\mu(x)$.

[1] Total derivatives with respect to spatial coordinates are still denoted by ∂, in contrast to Eq. (40.50), to allign with most works in the literature.

[2] More details can be found in e.g. Compère (2019).

[3] An extension to higher-order derivatives can be obtained by including further boundary terms.

Proof. *The general transformation rule for the Lagrangian is*

$$L \longrightarrow L + \alpha \delta L \,. \tag{53.7}$$

To have a symmetry, i.e. to keep the action invariant, the deformation factor has to be a divergence:

$$L \longrightarrow L + \alpha \partial_\mu \mathcal{J}^\mu \,. \tag{53.8}$$

To obtain the conservation law (53.6)*, the Lagrangian is varied explicitly:*

$$\begin{aligned}
\delta L &= \frac{\partial L}{\partial \phi} \delta \phi + \frac{\partial L}{\partial \phi_{,\mu}} \delta \phi_{,\mu} \\
&= \frac{\partial L}{\partial \phi} \delta \phi + \partial_\mu \left(\frac{\partial L}{\partial \phi_{,\mu}} \delta \phi \right) - \partial_\mu \left(\frac{\partial L}{\partial \phi_{,\mu}} \right) \delta \phi \\
&= \partial_\mu \left(\frac{\partial L}{\partial \phi_{,\mu}} \delta \phi \right) + \left[\frac{\partial L}{\partial \phi} - \frac{\partial L}{\partial \phi_{,\mu}} \right] \delta \phi \,.
\end{aligned}$$

The second term vanishes due to the Euler–Lagrange equation (52.33)*. Combining these formulas gives*

$$\partial_\mu \left(\frac{\partial L}{\partial \phi_{,\mu}} \delta \phi \right) - \partial_\mu \mathcal{J}^\mu(x) = 0 \,. \tag{53.9}$$

From this equation, one can conclude that the current

$$j^\mu(x) = \frac{\partial L}{\partial \phi_{,\mu}} \delta \phi - \mathcal{J}^\mu(x) \tag{53.10}$$

is conserved. □

The above conservation law can also be expressed in terms of a charge (such a current and its associated charge are generally called the **Noether current** and **Noether charge**):

$$Q[\Sigma] := \int_\Sigma j^0 \, d^{n-1}x \,, \tag{53.11}$$

where Σ is a spacelike hypersurface. The conservation law can then simply be restated as

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = 0 \,. \tag{53.12}$$

Definition 53.1.2 (Stress-energy tensor). Consider the translation of a scalar field:

$$\phi(x) \longrightarrow \phi(x+a) = \phi(x) + a^\mu \partial_\mu \phi(x) \,. \tag{53.13}$$

Because the Lagrangian is a scalar quantity, it transforms in the same way as the fields:

$$L \longrightarrow L + a^\mu \partial_\mu L = L + a^\nu \partial_\mu (L\delta^\mu{}_\nu)\,. \tag{53.14}$$

On an n-dimensional manifold, this leads to the existence of $n \in \mathbb{N}$ conserved currents. These can be used to define the stress-energy tensor:

$$T^\mu{}_\nu = \frac{\partial L}{\partial \phi_{,\mu}} \partial_\nu \phi - L\delta^\mu{}_\nu\,. \tag{53.15}$$

53.2 De Donder–Weyl theory

Consider a Lagrangian L on a Lorentzian spacetime M, i.e. there are $d+1$ coordinates x^μ with one timelike coordinate x^0. In the general approach to converting Lagrangian mechanics to Hamiltonian mechanics, one takes the Legendre transformation with respect to time (Definition 52.3.6). However, this would clearly violate covariance!

The De Donder–Weyl approach proceeds by not just performing a Legendre transformation with respect to the timelike coordinate, but with respect to all coordinates, thereby preserving covariance.

Definition 53.2.1 (Polymomentum). Definition 52.2.3 of conjugate momenta is generalized as follows:

$$\pi^\mu_I := \frac{\partial L}{\partial(\partial_\mu \phi^I)}\,. \tag{53.16}$$

Given these canonical (poly)momenta, Hamilton's equations 52.3.7 are then readily generalized:

$$\begin{cases} \dfrac{\partial \phi^I}{\partial x^\mu} = \dfrac{\partial H}{\partial \pi^\mu_I}\,, \\[2ex] \dfrac{\partial \pi^\mu_I}{\partial x^\mu} = -\dfrac{\partial H}{\partial \phi^I}\,. \end{cases} \tag{53.17}$$

Remark 53.2.2 (Bracket). The Poisson bracket can be generalized to De Donder–Weyl theory through the structure of a Gerstenhaber algebra (Definition 4.9.4).

Chapter 54

BRST Theory

The foundations for this subject were laid down by Dirac (1950). By introducing constraints, the coordinates and their momenta become dependent. This implies, for example, that the Hamiltonian equations of motion have to be modified.

Chapter Contents

54.1 Constrained systems **1005**
54.1.1 Constraints 1005
54.1.2 Zero Hamiltonian 1010
54.1.3 Field theories 1011
54.1.4 Constraint surface 1012
54.1.5 Fermionic systems 1013
54.2 Hamiltonian BRST theory **1014**
54.2.1 Introduction 1014
54.2.2 Irreducible constraints 1015
54.2.3 Reducible constraints 1019
54.2.4 Observables 1021
54.2.5 Actions and gauge fixing 1022
54.2.6 Nonminimal sectors 1022
54.2.7 Faddeev–Popov action 1023
54.3 Lagrangian BRST theory **1025**
54.3.1 Field theory 1025
54.3.2 Gauge algebra 1026
54.3.3 Antifields 1028

54.1 Constrained systems

54.1.1 Constraints

Definition 54.1.1 (Holonomic constraint). A constraint $f(q,t) = 0$ that only depends on the coordinates q^i and t and not on the derivatives.

Method 54.1.2 (Holonomic constraints). The Euler–Lagrange equations of a system with $k \in \mathbb{N}$ holonomic constraints

$$\zeta_k(q,t) = 0 \tag{54.1}$$

can be obtained from the extended action functional

$$S_E[q,\lambda] := \int_{t_1}^{t_2} \left[L(q,\dot{q},t) - \sum_{i=1}^{k} \lambda_i(t)\zeta_i(q,t) \right] dt\,, \tag{54.2}$$

where the λ_i are Lagrange multipliers (Section 50.3.1). Extremizing with respect to these multipliers induces the constraints:

$$\frac{\delta S_E}{\delta \lambda_i} = \zeta_i = 0\,. \tag{54.3}$$

First, recall the Lagrangian equations of motion (52.33). By expanding these equations, it can be shown that the accelerations $\ddot{q}$ are uniquely determined by the coordinates and velocities $(q,\dot{q})$ if and only if the Hessian of the Lagrangian is invertible. If the Hessian is not invertible, the definition of the conjugate momenta (Definition 52.2.3) cannot be inverted to express velocities in terms of momenta. Alternatively, the coordinates q and momenta p are not independent and there must exist relations of the form

$$\zeta(q,p) = 0\,. \tag{54.4}$$

Constraints of this type are called **primary constraints**. They do not serve to constrain the range of the coordinates q, they only couple the coordinates and the momenta.

Axiom 54.1 (Regularity conditions). It will always be assumed that the independent constraints, i.e. the minimal generating set, satisfy the following (equivalent) conditions:

- The constraints can locally serve as the first coordinates of a (regular) coordinate system.
- The differentials (gradients) $\mathrm{d}\zeta_m$ are locally linearly independent.
- The variations $\delta\zeta_m$ are of the order ε whenever the variations $\delta q^i, \delta p_i$ are of the order ε. (This is the original condition due to *Dirac*.)

A constrained dynamical system consists of a dynamic system (M, ω, H) together with a finite collection of constraints $\zeta_m(q,p) = 0, m \in I$. If this system is derived from a Lagrangian $L(q,\dot{q})$, the calculus of variations easily extends to these constrained systems, where it gives the following modified Hamiltonian equations:

$$\dot{q}^i = \frac{\partial H}{\partial p_i} + \sum_{m\in I} u_m \frac{\partial \zeta_m}{\partial p_i}\,, \tag{54.5}$$

$$\dot{p}_i = -\frac{\partial H}{\partial q^i} - \sum_{m\in I} u_m \frac{\partial \zeta_m}{\partial q^i}\,, \tag{54.6}$$

where the u_m are functions of the coordinates and velocities that play a role similar to ordinary Lagrange multipliers. In terms of Poisson brackets, the constrained time evolution of a (time-independent) function is given by:

$$\dot{f} = \{H, f\} + \sum_{m\in I} u_m \{\zeta_m, f\}\,. \tag{54.7}$$

Remark 54.1.3. The above relations follow from the property that the general solutions to $\lambda_i \delta q^i + \mu_i \delta p^i = 0$ for variations $\delta q^i, \delta p_i$ tangent to the constraint surface are of the form

$$\begin{cases} \lambda_i = \sum_{m\in I} u_m \frac{\partial \zeta_m}{\partial q^i}\,, \\ \mu^i = \sum_{m\in I} u_m \frac{\partial \zeta_m}{\partial p_i}\,. \end{cases} \tag{54.8}$$

Combining this result with the usual derivation of Hamilton's equations from a Lagrangian action principle gives the above modified equations.

Method 54.1.4 (General Poisson brackets). Until now, Poisson brackets were only defined for functions depending on the canonical coordinates (q,p). This definition can be generalized to arbitrary functions through the Poisson algebra properties (Definition 32.2.5). Furthermore, after working out the Poisson brackets, one can use the constraint equations to drop all terms that are proportional to ζ_m.

For example, Eq. (54.7) can be rewritten as

$$\dot{f} = \left\{ H + \sum_{m\in I} u_m \zeta_m, f \right\}. \tag{54.9}$$

To prove the equivalence, one can use the linearity and Leibniz properties. This involves the following equality

$$\{u_m \zeta_m, f\} = \{u_m, f\}\zeta_m + u_m \{\zeta_m, f\}\,. \tag{54.10}$$

The Poisson brackets in the second term only involve functions depending on (q,p) and can be calculated in the usual way. The first term, however, involves a Poisson bracket of the Lagrange multiplier u_m. In general, these do not simply depend on q and p. Luckily, this does not pose a problem because the term is proportional to the constraints and, as such, vanishes on-shell. It is important that the constraints are only applied after the Poisson brackets have been fully worked out.

Notation 54.1.5 (Weak equality). The constraints ζ_m only vanish on shell. To distinguish between functional equalities, i.e. equalities that also hold off-shell, and on-shell equalities, also called **weak equalities**, the latter are often denoted by the $\approx$ symbol. For example, the condition $\zeta_m \approx 0$ is only a weak equality.

Using the above definitions, one can write an arbitrary time derivative as

$$\dot{f} \approx \{H_T, f\}\,, \tag{54.11}$$

where $H_T := H + \sum_{m\in I} u_m \zeta_m$ is the total Hamiltonian.

Remark 54.1.6 (Closure). An important insight regarding weak equalities can be obtained by calculating the Poisson bracket with a function f that is strongly zero, i.e. a function that vanishes on shell and whose variation also vanishes. In this case, $\{f, g\} \approx 0$ for all functions g, i.e. the brackets only vanish weakly. Furthermore, if f is only weakly zero, then $\{f, g\}$ does not even have to vanish at all.

Property 54.1.7 (Weakly vanishing functions). Assuming that the constraints satisfy Axiom 54.1, a function that vanishes on shell is equal to some combination of the constraints (the coefficients might be functions themselves) (Henneaux & Teitelboim, 1992, p. 70).

Property 54.1.8 (Consistency conditions). By taking $f = \zeta_n$ for any $n \in I$ in Eq. (54.7), a set of consistency conditions is obtained:

$$\{H, \zeta_n\} + \sum_{m\in I} u_m \{\zeta_m, \zeta_n\} \approx 0\,. \tag{54.12}$$

It is possible that this condition reduces to an inconsistency of the type $1 \approx 0$. In this case, the equations of motion are inconsistent and the theory is not physical. If this is not the case, multiple possibilities can arise:

- After imposing the primary constraints, a tautology $0 = 0$ is found. This gives no new information.
- The equation reduces to an equation not involving the Lagrange multipliers u_m. This gives an additional constraint

  $$\chi(q, p) = 0\,. \tag{54.13}$$

  These are called **secondary constraints**.
- A condition on the coefficients u_m is obtained.

After having found a set of secondary constraints, this procedure can be iterated until no new contraints or conditions are found. Because the consistency conditions are linear in the coefficients u_m, the general solution can be written as

$$u_m = U_m + v_a V^a_m\,, \tag{54.14}$$

where U_m is a solution of the inhomogeneous equation (54.12) and the V^a_m are linearly independent solutions of the homogeneous equation

$$\sum_{m\in I} u_m\{\zeta_m, \zeta_n\} = 0\,. \tag{54.15}$$

The resulting coefficients v_a are completely arbitrary functions of time and will be referred to as the **Lagrange multipliers**. Therefore, the total Hamiltonian can be written in the form

$$H_T = H'(q,p) + v^a(t)\zeta_a(q,p)\,, \tag{54.16}$$

where $\zeta_a := \sum_{m\in I} V^a_m \zeta_m$. The occurence of arbitrary functions in the Hamiltonian implies that the evolution of the phase space variables is not unique and, accordingly, that the theory has a gauge freedom.

Definition 54.1.9 (First- and second-class). A function $f(q,p)$ is said to be first class if its Poisson bracket with every constraint (both primary and secondary) is weakly zero. The function is said to be second class otherwise. It can be shown that both the total Hamiltonian H_T and the primary constraints ζ_a are first class. The number of Lagrange multipliers v^a is equal to the number of primary, first-class constraints.

For first-class constraints $\{\zeta_a\}_{a\in I}$, this leads to the following structure:

$$\{\zeta_a, \zeta_b\} = C^c_{ab}\zeta_c + T^i_{ab}\frac{\delta S}{\delta x^i}\,, \tag{54.17}$$

where $C^c_{ab} : M \to \mathbb{R}$ are the **structure functions** and the $T^i_{ab} : M \to \mathbb{R}$ are antisymmetric under $a \leftrightarrow b$. The structure of the constraint algebra depends on the properties of these coefficients:

- $T_{ab} = 0$ and C^c_{ab} constant: **closed algebra** (this corresponds to Lie or L_∞-algebras),
- $T_{ab} = 0$: **soft algebra** (this corresponds to Lie or L_∞- algebroids 44.6.4), and
- in general: **open algebra** (this corresponds to *central extensions of L_∞-algebroids*) .

Notation 54.1.10. To distinguish between first- and second-class constraints, the latter are often denoted by a separate symbol χ.

Property 54.1.11 (Closure). The Poisson bracket of two primary, first-class functions is first-class. So is the Poisson bracket of the total Hamiltonian and a primary, first-class constraint.

Corollary 54.1.12 (Lie algebra). The first-class constraints form a Lie algebra with respect to the Poisson bracket and the associated gauge transformations define a submanifold in phase space by Frobenius' theorem 34.6.6.

Property 54.1.13. If the constraint algebra is closed, every gauge-invariant function is gauge equivalent to a strongly gauge-invariant function.

Remark 54.1.14 (Dirac conjecture). The primary, first-class constraints generate gauge transformations in the sense that variations in the coefficients, which are arbitrary, give rise to phase space variations that leave the physical state invariant. Some secondary constraints might also generate gauge transformations and *Dirac* even conjectured that this was the case for all constraints. However, counterexamples have been found. A common workaround is simply to restrict to systems where the conjecture is true and, from here on, the distinction between primary and secondary will be dropped. From this point of view, it makes sense to define the extended Hamiltonian

$$H_E := H_T + v^b(t)\zeta_b(q,p)\,, \tag{54.18}$$

where b ranges over all secondary, first-class constraints. For gauge-invariant functions, i.e. those functions whose Poisson bracket with all first-class constraints vanishes, evolution with all three Hamiltonians H, H_T and H_E is identical. For general functions, only H_E takes into account the full gauge freedom.[1]

Formula 54.1.15 (Degrees of freedom). The number of degrees of freedom is given by the following formula:

$$\begin{aligned} 2 \times \#\text{d.o.f.} = {} & \#\text{canonical coordinates} \\ & - \#\text{second-class constraints} \\ & - 2 \times \#\text{first-class constraints.} \end{aligned}$$

Definition 54.1.16 (Dirac bracket). To take care of second-class constraints, *Dirac* introduced a modification of the Poisson bracket:

$$\{f,g\}_D := \{f,g\} - \{f,\chi_a\}C^{ab}\{\chi_b,g\}\,, \tag{54.19}$$

where the χ_a are the second-class constraints and the (invertible) matrix C^{ab} is the inverse of the matrix $C_{ab} := \{\chi_a,\chi_b\}$.

The benefit of using the Dirac bracket (after the Poisson bracket has been used to separate constraints in first-class and second-class constraints) is that second-class constraints become strong equalities, i.e. they can be used even before evaluating further (Dirac) brackets. The Dirac bracket satisfies the same algebraic properties as the Poisson bracket, i.e. it also defines a Poisson algebra. From here on, all constraints will be assumed to be first class, i.e. the Poisson bracket will be assumed to be the one obtained after applying the Dirac procedure to all second-class constraints.

Remark 54.1.17. Instead of splitting the constraints in first- and second-class instances and having to work with the nontrivial Dirac bracket, one can also try to remove second-class constraints in a different way. In the above formula for the degrees of freedom, the

[1] Note that H_T is the Hamiltonian that corresponds to a Lagrangian approach. H_E gives a more general theory.

factor 2 on the right-hand side is obtained by the introduction of gauge-fixing conditions. What these actually do is turning first-class constraints into second-class ones. In fact, the converse is also possible. One can obtain all second-class constraints as gauge-fixed first-class constraints after enlarging the system (although this procedure is not unique). After doing this, there is no need for the Dirac bracket anymore and one can simply work with the Poisson bracket (with the added complexity that all constraints now only hold weakly).

Definition 54.1.18 (Gauge-invariant functions). Consider the algebra $C^\infty(M)$ of smooth functions on phase space. In the spirit of algebraic geometry, the space of functions on the constraint surface Σ is given by the quotient algebra $C^\infty(\Sigma) := C^\infty(M)/\mathcal{N}$, where $\mathcal{N}$ is the ideal having Σ as its zero locus, i.e. $\mathcal{N}$ is the ideal generated by the constraints. The elements of $C^\infty(\Sigma)$ that are gauge-invariant, i.e. first class with respect to first-class constraints, should be considered as the **classical observables**. Passing to this extension effectively amounts to solving a(n integrable) system through constants of motion.

The restriction to gauge-invariant functions is also imperative if one wants to extend the Dirac bracket to $C^\infty(\Sigma)$. In general, the ideal $\mathcal{N}$ is not an ideal of the Dirac bracket. The gauge-invariant subalgebra is, in fact, the maximal subalgebra of $C^\infty(\Sigma)$ for which $\mathcal{N}$ is again an ideal.

54.1.2 Zero Hamiltonian

In the previous section, dynamical systems with constraints were considered. Using these tools, one can turn any system evolving under a physical, but nondynamical or external, time parameter t into a system having time as a canonical coordinate. In this setting, the time variable is treated on the same footing as the other coordinates. Such a system is said to be **generally covariant**.

If one starts from the single-particle action

$$S[q,p] = \int \left(p_i \dot{q}^i - H_0\right) dt, \tag{54.20}$$

one can introduce time as a generalized coordinate with momentum p_0 by modifying the action as follows:

$$S[q,p,t,p_0,u] = \int \left[p_0 t' + p_i q'^i - u(p_0 + H_0)\right] d\tau\,, \tag{54.21}$$

where the primes indicate derivatives with respect to the parameter τ and u acts as a Lagrange multiplier. It is easily checked that the resulting equations of motion are the same as for the original action.

The system now involves a single constraint $H_0 = -p_0$, which is first class. It is often called the **Hamiltonian constraint**. Aside from this constraint, the extended action

contains no first-class Hamiltonian. Evolution is solely determined by a (first-class) constraint and, therefore, is given by a gauge transformation.

Remark 54.1.19 (Nonholonomicity). All constraints were assumed to be holonomic, i.e. they did not explicitly depend on time. The presence of time derivatives is not compatible with the Poisson/Dirac bracket. However, when passing to a generally covariant system as above, the time variable loses its peculiar character and all constraints can be handled in the same way.

Property 54.1.20 (Vanishing Hamiltonian). If the canonical coordinates (q, p) transform as scalars under τ-reparametrizations, the Hamiltonian is weakly zero. Conversely, zero Hamiltonians give rise to reparametrization-invariant systems.

Although in many cases, it is reasonable to introduce a certain split of the coordinates and break general covariance — gauge-invariant observables and constants of motion are usually treated differently — this is not required. For a (weakly) zero Hamiltonian, first class functions are exactly constant of motion.

54.1.3 Field theories

Up until now, the theories were assumed to have a finite number of degrees of freedom. However, in field theory, the phase space is (a subspace of) the module of sections of some vector bundle and, as such, is infinite dimensional. To ensure the existence of a well-behaved BRST theory, one has to restrict to 'local theories', i.e. all relevant functions/functionals should only depend on the fields and their derivatives up to some finite order.

Axiom 54.2 (Locality). A Hamiltonian gauge theory is local if the following conditions are satisfied:

1. The Hamiltonian action is local:

$$S[z, \zeta] := \int_{t_1}^{t_2} \int_M \left(\dot{\phi}^I \pi_I - h_0 - \lambda^a \zeta_a \right) d^n x \, dt \,, \tag{54.22}$$

   where the field momenta π_I, the Hamiltonian density h_0 and the constraints ζ_a are functions of the fields ϕ^I and their derivatives up to a finite order $k \in \mathbb{N}$, i.e. they are elements of a finite jet bundle $J^k(M)$.

2. The bracket $\{\phi^I(x), \phi^J(y)\}$ is local for all I, J and $x, y \in M$.

3. Any reducibility relation is again local.

The second condition ensures that taking iterated brackets respects the locality of the theory.

Axiom 54.3 (Local completeness). Any function that vanishes weakly is zero on the constraint surface by means differential identities of the constraints only, i.e. it does not depend on the boundary conditions.

Remark 54.1.21. Local completeness can always be ensured by adding constraints or by replacing constraints by their (integral) primitives.

Axiom 54.1 can be extended to local field theories as follows.

Axiom 54.4 (Regularity condition). Let $V := \{\zeta_a\}$ be a set of k-local constraints, with $k \in \mathbb{N}$. For every $n \in \mathbb{N}$, the jet prolongation $j^n V$ cuts out a submanifold of $J^{n+k}(M)$. It is assumed that there exists a generating set of $j^n V$ such that these independent constraints form a regular coordinate system of $j^n V$.

If the above axioms are satisfied, Property 54.1.7 can be generalized to field theories.

Property 54.1.22. If a local function vanishes on the constraint surface, it is an element of some finite prolongation of the constraint algebra.

54.1.4 Constraint surface

Property 54.1.23 (Geometric characterization). Restricted to a first-class constraint surface Σ, the 'symplectic' form becomes maximally degenerate with rank

$$\mathrm{rk}(\omega) = \dim(M) - 2\dim(\Sigma)\,. \tag{54.23}$$

This essentially says that the constraint surface is coisotropic and, as a consequence, that the Poisson bracket is ill-defined (since this would involve the inverse of the symplectic form). After passing to the **reduced phase space**, i.e. the leaf space of the Hamiltonian foliation generated by the constraints, one again obtains a well-defined Poisson bracket that coincides with the ordinary Poisson bracket without any constraints.

The opposite situation arises for constraint surfaces that only involve second-class constraints. Here, the induced symplectic form is of maximal rank

$$\mathrm{rk}(\omega) = \dim(M) - \dim(\Sigma)\,, \tag{54.24}$$

which implies that the surface is isotropic. Furthermore, the induced Poisson bracket coincides with the restriction of the Dirac bracket.

Formula 54.1.24 (Reduced action). Let $y^j(q^i)$ be regular coordinates on Σ. Let $\theta_j \mathrm{d}y^j$ be a symplectic potential for the associated symplectic form (on Σ). The reduced action reads

$$S[y] = \int_{t_1}^{t_2} (\theta_j y^j - H|_{\zeta=0})\, dt\,. \tag{54.25}$$

The reduced potential satisfies

$$\theta_j \mathrm{d}y^j \approx p_i \mathrm{d}q^i + \mathrm{d}M\,. \tag{54.26}$$

Remark 54.1.25 (Algebraic characterization). The fact that first-class constraints define a coisotropic submanifold is not a peculiarity. A multiplicative ideal of a Poisson algebra that is closed under the Poisson bracket, is often called a **coisotrope** (or **coisotropic ideal**). Coisotropic submanifolds of a Poisson manifold (Definition 37.2.3) are exactly the zero loci of such coisotropes. In fact, one can restate the above constructions in purely algebraic terms. Given a Poisson algebra P and a coisotrope $\mathcal{J}$, one can pass to the quotient $N(\mathcal{J})/\mathcal{J}$, where N denotes the normalizer in P. This quotient is again a Poisson algebra, called the **reduced Poisson algebra**. This construction is strictly more general than the symplectic case considered above. (The sections further on could also be generalized to this setting.)

Theorem 54.1.26 (Abelianization). *Locally, one can always find a gauge transformation such that the constraint algebra becomes Abelian:*

$$\{\zeta, \zeta'\} = 0\,. \tag{54.27}$$

54.1.5 Fermionic systems

In this section, the study of constrained systems with 'fermionic' or odd statistics is considered. For an introduction to Grassmann numbers, see Section 23.5.2. In general, the phase space will be assumed to be a supermanifold (Example 44.1.1).

First, the ordinary Poisson bracket is extended to Grassmann-odd coordinates as follows:

$$\{\theta^i, \theta^j\} = 0 = \{\pi_i, \pi_j\} \tag{54.28}$$

and

$$\{\theta^i, \pi_j\} = \delta^i_j = \{\pi_j, \theta^i\}\,. \tag{54.29}$$

By defining the matrix $\sigma^{ij} := \{z^i, z^j\}$, where z can be any of the q, p, θ or π, one can then succintly write the Poisson bracket of superfunctions as follows:

$$\{f, g\} := \frac{\partial^R f}{\partial z^i} \sigma^{ij} \frac{\partial^L g}{\partial z^j}\,. \tag{54.30}$$

Note that this is virtually the same expression as the ordinary Poisson bracket, where the matrix σ was the inverse of the symplectic matrix. Writing out all terms gives

$$\{f, g\} = \left(\frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q^i} - \frac{\partial f}{\partial q^i}\frac{\partial g}{\partial p_i}\right) + (-1)^{\deg(f)}\left(\frac{\partial f}{\partial \pi_i}\frac{\partial g}{\partial \theta^i} + \frac{\partial f}{\partial \theta^i}\frac{\partial g}{\partial \pi_i}\right). \tag{54.31}$$

The algebraic properties of this generalized Poisson bracket are graded generalizations of those of the ordinary one:

$$\{f,g\} = -(-1)^{\deg(f)\deg(g)}\{g,f\} \tag{54.32}$$

$$\begin{aligned} 0 = \{f,\{g,h\}\} &+ (-1)^{[\deg(f)+\deg(g)]\deg(h)}\{h,\{f,g\}\} \\ &+ (-1)^{\deg(f)[\deg(g)+\deg(h)]}\{g,\{h,f\}\} \end{aligned} \tag{54.33}$$

$$\{f,gh\} = \{f,g\}h + (-1)^{\deg(f)\deg(g)} f\{g,h\} \tag{54.34}$$

$$\deg(\{f,g\}) = \deg(f) + \deg(g)\,. \tag{54.35}$$

The first two properties state that the generalized Poisson bracket gives rise to a Lie superalgebra (Example 4.9.2) and the third property states that it is a *Poisson superalgebra*. In fact, this is the example that lends its name to the algebraic structure. Geometrically, the matrix σ gives rise to a *supersymplectic structure*.

54.2 Hamiltonian BRST theory

54.2.1 Introduction

Consider a dynamical system (M,ω,H) together with a set of first-class constraints $\{\zeta_a\}_{a\in I}$. As was shown before, these constraints generate an algebra under the Poisson bracket. However, even more structure is present. To explore this structure, the phase space can be enlarged by introducing, for every constraint and every relation between constraints, a Grassmann-odd[2] **ghost variable** η^a and its canonical conjugate $\mathcal{P}_a$, i.e.

$$\{\mathcal{P}_a,\eta^b\} := -\delta_a^b\,. \tag{54.36}$$

These come with two types of grading:

1. the **pure ghost number** generated by $\text{pure gh}(\mathcal{P}_a) = 0$ and $\text{pure gh}(\eta^a) = 1$, and
2. the **antighost number** generated by $\text{antigh}(\mathcal{P}_a) = 1$ and $\text{antigh}(\eta^a) = 0$.

The (total) **ghost number** is defined as the difference of these ghost numbers:

$$\text{gh}(f) := \text{pure gh}(f) - \text{antigh}(f)\,. \tag{54.37}$$

Moreover, both the ghosts and ghost momenta have Grassmann parity $\varepsilon_a + 1$.

On this extended phase space, one can then construct a nilpotent function Ω, the **BRST generator**, that induces a cohomology theory (Definition 5.1.3) through the Poisson

[2]In fact, one can generalize this section to phase spaces that already contain odd variables. In that case, the ghost variables should have the opposite parity of the associated constraints.

bracket:

$$s := \{\cdot, \Omega\}\,. \tag{54.38}$$

This cohomology theory characterizes the gauge structure (such as gauge-invariant functions).

Remark 54.2.1 (Nonminimal sectors). In certain situations, it is useful to extend the phase space even further by introducing additional conjugate pairs, e.g. the Lagrange multipliers in an action principle. Such descriptions are said to belong to the nonminimal sector. An example would be the *Nakanishi–Lautrup field B*, introduced when quantizing Yang–Mills theory (see Section 69.6.1), which is conjugate to the *Faddeev–Popov antighost field* $\bar{c}$, i.e. $\{\bar{c}, \Omega\} = B$. This will be covered further down in Section 54.2.6.

54.2.2 Irreducible constraints

In this section, only systems with irreducible constraints are considered, i.e. it will be assumed that no relations between the constraints exist. To formulate the BRST complex in terms of invariant geometric notions, a homological and differential-geometric approach will be adopted.

A first step is to express the algebra of on-shell functions $C^\infty(\Sigma)$ in an invariant way. The idea is to rewrite the quotient $C^\infty(M)/\mathcal{N}$ as a homology group (which is invariant by its very nature). To this end, one passes to the Koszul complex (Property 5.4.16) associated to the first-class constraints ζ_a: $C^\infty(M) \otimes \mathbb{C}[\mathcal{P}_a]$. (Independence of the constraints exactly says that they form a regular sequence and, hence, the complex gives a homological resolution.) One then finds that $H_0(\delta) \cong C^\infty(M)/\mathcal{N} \cong C^\infty(\Sigma)$, where δ is the Koszul differential.

Definition 54.2.2 (Antighost number). The degree corresponding to δ is exactly the antighost number and the Koszul generators are the ghost momenta $\mathcal{P}_a$ associated to the constraints ζ_a. This also implies that $\mathsf{antigh}(\delta) = -1$.

A second step is to characterize the gauge structure of the constraint surface Σ. To this end, a modified exterior derivative on the phase space is introduced. Although the gauge algebra spanned by the constraints ζ_a does not necessarily generate a closed gauge group on the full phase space, it does so when restricted to the constraint surface. The $|I|$-dimensional leaves of the foliation generated by the constraints are called the **gauge orbits** and the Hamiltonian vector fields associated to the first-class constraints are tangent to these orbits. Furthermore, by the irreducibility of the constraints, the vector fields form a frame field for the tangent bundle of the gauge orbits.

Definition 54.2.3 (Longitudinal complex). Vector fields that are tangent to the gauge orbits are said to be **longitudinal** or **vertical** (not to be confused with the vertical vector

fields from Section 35.3). A frame for these vector fields is given by the Hamiltonian vector fields

$$X_a := \{\zeta_a, \cdot\}\,. \tag{54.39}$$

The tangent bundle of Σ in M can be decomposed as follows:

$$TM|_\Sigma = T\mathcal{F} \oplus N\Sigma\,, \tag{54.40}$$

where $T\mathcal{F}$ denotes the tangent bundle of the foliation generated by the constraints and $N\Sigma$ denotes the normal bundle to the foliation.

This decomposition also turns the de Rham complex $\Omega(\Sigma) := \Omega(M)|_\Sigma$ into a bicomplex $\Omega^{\bullet,\bullet}(\Sigma) = \Lambda^\bullet T^*\mathcal{F} \otimes \Lambda^\bullet N^*\Sigma$. The longitudinal (or vertical) complex is given by the subcomplex $\Omega^{\bullet,0}(\Sigma)$, i.e. the longitudinal forms η^a are the multilinear duals of the longitudinal vector fields. The **longitudinal derivative** is the projection of the total de Rham differential on the longitudinal subcomplex, i.e. the operator $\mathbf{d} : \Omega^{\bullet,0}(\Sigma) \to \Omega^{\bullet+1,0}(\Sigma)$ given by

$$\mathbf{d}f := \mathrm{d}f = \{\zeta_a, f\}\,\eta^a\,, \tag{54.41}$$

$$\mathbf{d}\eta^a := -\frac{1}{2}C^a_{bc}(q,p)\eta^b \wedge \eta^c\,, \tag{54.42}$$

where the $C^a_{bc}(q,p)$ are the structure functions of the constraint algebra. The Grassmann parity of the forms η^a is taken to be $\varepsilon_a + 1$. Note that the action of $\mathbf{d}$ is exactly the action of the Chevalley–Eilenberg differential from Section 32.6. (This is not a coincidence as will be explained further on.)

The longitudinal complex can be extended to all of M by taking its elements to be the forms

$$A \equiv A_{i_1\ldots i_k}(q,p)\eta^{i_1} \wedge \cdots \wedge \eta^{i_k}\,, \tag{54.43}$$

where the coefficients are equivalence classes of weakly equal functions in $C^\infty(M)$.

Definition 54.2.4 (Pure ghost number). The longitudinal differential forms correspond to the ghosts in Section 54.2.1 and, accordingly, their degree corresponds to the (pure) ghost number. This also implies that $\mathsf{pure\ gh}(\mathbf{d}) = 1$.

Note that the number of Koszul generators is the same as the number of ghost fields, since both are induced by the constraints ζ_a (as foretold in Section 54.2.1). To extend the Poisson bracket to the **extended phase space** containing phase space functions, ghost fields and ghost momenta, the following convention is introduced:

$$\{\mathcal{P}_a, \eta^b\} = -(-1)^{(\varepsilon_a+1)(\varepsilon_b+1)}\{\eta^b, \mathcal{P}_a\} := -\delta^b_a\,. \tag{54.44}$$

Definition 54.2.5 (Ghost number). The total ghost number of an element in the coordinate superalgebra $C^\infty(M) \otimes \mathbb{C}[\eta^a] \otimes \mathbb{C}[\mathcal{P}_a]$ on the extended phase space is defined as follows:

$$\mathsf{gh}(A) := \mathsf{pure\ gh}(A) - \mathsf{antigh}(A)\,. \tag{54.45}$$

It satisfies

$$\mathsf{gh}(AB) = \mathsf{gh}(A) + \mathsf{gh}(B)\,. \tag{54.46}$$

Property 54.2.6. The ghost number is equal to the eigenvalue of the operator

$$\mathcal{G} := i\eta^a \mathcal{P}_a\,. \tag{54.47}$$

When passing to longitudinal forms on all of M, the operator $\mathbf{d}$ fails to be a differential, i.e. $\mathbf{d}^2 \neq 0$ on M. It is only weakly zero outside Σ. Furthermore, when extending the longitudinal derivative $\mathbf{d}$ to the extended phase space, one has the freedom to choose the action on the ghost momenta under the constraint that $\mathsf{gh}(\mathbf{d}) = 1$ and $\mathsf{antigh}(\mathbf{d}) = 0$.[3] By making the choice

$$\mathbf{d}\mathcal{P}_a := (-1)^{\varepsilon_a} \eta^c C^b_{ca} \mathcal{P}_b\,, \tag{54.48}$$

the Koszul differential and longitudinal derivative satisfy $[\delta, \mathbf{d}] = 0$. This also turns $\mathbf{d}$ into a differential modulo δ (Definition 5.1.8). The homology of δ can be generalized to the full extended phase space by tensoring the Koszul resolution of $C^\infty(\Sigma)$ by $\mathbb{C}[\eta^a]$, since the latter is a free and, in particular, projective module. The cohomology of $\mathbf{d}$ modulo δ on M can be shown to coincide with the cohomology of $\mathbf{d}$ on Σ. This is exactly the **BRST cohomology** from the introduction.

Homological perturbation theory (Theorem 5.3.3) now also says that there exists a true differential s on the extended phase space generating BRST cohomology.

Definition 54.2.7 (BRST operator). To any dynamical system governed by first-class constraints $\{\zeta_a\}_{a\in I}$, one can associate a BRST differential[4]

$$\begin{aligned} s &= \delta + \mathbf{d} + \cdots\,, \\ \varepsilon(s) &= 1\,, \\ \mathsf{gh}(s) &= 1\,. \end{aligned} \tag{54.49}$$

given by the Poisson bracket with a **BRST function** Ω

$$sA = \{A, \Omega\} \tag{54.50}$$

that satisfies the following conditions:

[3] Because δ is a boundary operator, i.e. it decreases the degree, there is less freedom in defining $\delta\eta^a$. Only $\delta\eta^a = 0$ is allowed.

[4] BRST stands for *Becchi, Rouet, Stora* and *Tyutin*.

1. It is of ghost degree 1:

$$\mathsf{gh}(\Omega) = 1\,. \tag{54.51}$$

2. It is nilpotent with respect to the Poisson bracket:

$$\{\Omega, \Omega\} = 0\,. \tag{54.52}$$

3. It is real/Hermitian:

$$\Omega^* = \Omega\,. \tag{54.53}$$

4. It is proportional to the constraints to lowest order:

$$\Omega = \eta^a \zeta_a + \text{terms of nonzero pure ghost number}^{5}\,. \tag{54.54}$$

Remark 54.2.8 (Off-shell closure). It should be noted that the nilpotency of the BRST operator not only holds on-shell, but everywhere on M. In this sense, the ghost momenta appearing in its definition are the fields necessary to close the algebra outside the constraint surface. In fact, it is important to work in the Hamiltonian formalism if one wants to achieve this off-shell nilpotency. It has been shown that in the Lagrangian formalism, this property cannot hold for gauge transformations that only close on-shell. (This latter property is related to the fact that the structure coefficients are generally functions of the canonical variables. Only when they are constants does the algebra of canonical transformations generated by the constraints close off-shell.) In fact, 'the BRST complex' historically only refers to the longitudinal complex. It only gives a resolution for the quotient by the constraint algebra. The full complex as considered here is called the **Batalin–Fradkin–Vilkovisky complex** (BFV theory).

It can be shown that the BRST operator only depends on the constraint surface and not on the choice of a local description.

Property 54.2.9 (Uniqueness). Any two BRST generators Ω, Ω' for the same constraint surface are related by a canonical transformation in the extended phase space.

Example 54.2.10 (Abelian constraint algebra). When the constraints form an Abelian algebra, i.e. $\{\zeta_a, \zeta_b\} = 0$, the terms involving higher ghost momenta vanish:

$$\Omega = \eta^a \zeta_a\,. \tag{54.55}$$

Example 54.2.11 (Closed constraint algebra). When the constraints form a closed algebra, i.e. $\{\zeta_a, \zeta_b\} = C^c_{ab}\zeta_c$ with C^c_{ab} constant, the BRST operator has a slightly more

[5] To ensure that the ghost number of Ω is 1, this means that the extra terms are at least quadratic in the ghost fields.

complex expression since the zeroth order term in the BRST expansion is not nilpotent on its own. However, since the structure functions C^c_{ab} are constants, all higher order terms still vanish:

$$\Omega = \eta^a \zeta_a - \frac{(-1)^{\varepsilon_b}}{2} \eta^b \eta^c C^a_{bc} \mathcal{P}_a \,. \tag{54.56}$$

More generally, the BRST generator has this form in lowest degree:

$$\Omega = \eta^a \zeta_a - \frac{(-1)^{\varepsilon_b}}{2} \eta^b \eta^c C^a_{cb} \mathcal{P}_a + \text{higher-order terms} \,. \tag{54.57}$$

Property 54.2.12 (BRST cohomology). For all functions f on the extended phase space, one has the following equality:

$$s^2 f = \{\{f, \Omega\}, \Omega\} = 0 \,, \tag{54.58}$$

i.e. s is a proper differential. In view of this structure, one says that a function is BRST closed function if

$$\{f, \Omega\} = 0 \tag{54.59}$$

and that it is BRST exactwhen it can be written as

$$f = \{g, \Omega\} \tag{54.60}$$

for some function g. It is clear that the resulting BRST cohomology theory is gauge-invariant, since Ω is gauge-invariant.

54.2.3 Reducible constraints

In this section, the irreducibility requirement for the first-class constraints is relaxed. To recover the BRST complex as a homological object, one has to modify the construction from the previous section. First of all, the Koszul complex is not a resolution of $C^\infty(\Sigma)$ anymore. Because higher-order relations exist among the constraints, the higher-degree homology groups do not vanish. Mathematically, the issue is that the constraints do not form a regular sequence anymore. However, they still generate a module and so a Koszul–Tate resolution exists (Property 5.4.17). Instead of only introducing ghost momenta corresponding to constraints, one also has to introduce ghost-of-ghosts.

A second problem occurs when trying to define the longitudinal complex and trying to combine it with the Koszul–Tate complex. The number of ghost momenta is now greater than the number of (longitudinal) ghost fields and, furthermore, the longitudinal algebra is not a tensor product $C^\infty(\Sigma) \otimes \mathbb{C}[\eta^a]$ due to the existence of relations among the constraints. The solution here is again to pass to a larger structure that has the correct 'homotopical' structure. In this case, this will be a Sullivan model (Definition 7.1.74), i.e. the Chevalley–Eilenberg algebra associated to an L_∞-algebroid (Definition 44.6.6).

Formula 54.2.13 (Ghost number). Due to the introduction of ghost-of-ghosts, Eq. (54.47) has to be modified. Let η^{a_1} denote the ghost fields, i.e. fields of (pure) ghost number 1, η^{a_2} the ghost-of-ghosts, i.e. fields of (pure) ghost number 2, etc. The total ghost number operator is then given by

$$\mathcal{G} := i \sum_{n=1}^{+\infty} (n+1)\eta^{a_n}\mathcal{P}_{a_n} + \text{constants due to operator ordering}\,, \tag{54.61}$$

where $\mathcal{P}_{a_n}$ are the ghost-of-ghost momenta of antighost number i.

Construction 54.2.14 (Longitudinal complex). Whereas there were as many Koszul generators as longitudinal basis forms in the case of irreducible algebras, this is not the case anymore for reducible algebras. Moreover, for reducible theories, the extended phase space can only locally be given a tensor product space. Sullivan models will allow to consider an equivalent dgca, with a global tensor product structure and as many longitudinal generators as Koszul–Tate generators, such that the associated cohomology coincides with the longitudinal cohomology.

In reducible theories, the constraints are related as follows

$$Z^{a_0}_{a_1}\zeta_{a_0} = 0 \tag{54.62}$$

for some functions $Z^{a_0}_{a_1} : M \to \mathbb{R}$. The coefficients in the overcomplete 'basis' X_a of longitudinal vector fields (54.39) then satisfy the following relations:

$$(-1)^{\varepsilon_{a_0}(\varepsilon_{a_1}+1)} Z^{a_0}_{a_1} X_{a_0} = 0\,. \tag{54.63}$$

Of course, higher-order reducibility then corresponds to further relations $Z^{a_1}_{a_2}$ among these relations. For every reducibility identity of order $k \in \mathbb{N}$, one then adds a ghost(-of-ghost) generator η^{a_k} satisfying

$$\Delta\eta^{a_k} := (-1)^{\varepsilon_{a_k}+k+1}\eta^{a_k+1}Z^{a_k}_{a_k+1}\,. \tag{54.64}$$

and

$$\text{pure gh}(\eta^{a_k}) = k+1 \qquad \text{and} \qquad \varepsilon(\eta^{a_k}) = \varepsilon_{a_k} + k + 1\,. \tag{54.65}$$

Note that the η^{a_k} are not given by differential forms as they were in the irreducible case!

Equation (54.57) can be generalized to the reducible setting by including the reducibiliy relations.

Formula 54.2.15. The BRST generator can be expanded as follows:

$$\Omega = \eta^{a_0}\zeta_{a_0} + \eta^{a_k}Z^{a_k-1}_{a_k}\mathcal{P}_{a_k-1} + \text{higher-order terms}\,. \tag{54.66}$$

Remark 54.2.16 (Chevalley–Eilenberg complex). As has been noted before, there are some relations between BRST complexes and Chevalley–Eilenberg algebras. In fact, these relations are no mere coincidences. The constraint algebra

$$\{\zeta_a, \zeta_b\} = C^c_{ab}\zeta_c \tag{54.67}$$

defines a Lie algebroid in the case of irreducible constraints and a (potentially truncated) L_∞-algebroid in the case of reducible constraints. Similar to Eq. (32.81) in Lie algebra cohomology, one can characterize invariants of Lie algebroids in terms of their Chevalley–Eilenberg cohomology. From this point of view, gauge-invariant functions do not just resemble elements of the zeroth cohomology group of a Chevalley–Eilenberg differential, they are exactly that.

54.2.4 Observables

By construction of the BRST complex, the s-cohomology coincides with the cohomology of the longitudinal differential. In negative ghost degree, it can be shown that BRST cohomology vanishes. In degree 0, one finds the physical observables.

Property 54.2.17 (Gauge-invariant functions). $H^0(s)$ is isomorphic to the set of equivalence classes of weakly equal, gauge-invariant functions.

Given a BRST-closed function f of ghost degree 0, the associated **classical observable** is obtained as the term of antighost number 0 in its BRST expansion. Conversely, any BRST-closed function of ghost number 0 is called a **BRST-invariant extension** of its term of antighost number 0.

Remark. The interpretation of higher cohomology groups will be adressed in Chapter 59.

Property 54.2.18 (Poisson algebra). The Poisson bracket descends to $H^0(s)$ and defines a (graded) Poisson algebra structure. Furthermore, if f and g are BRST-invariant extensions of f_0 and g_0 respectively, the functions fg and $\{f, g\}$ are BRST-invariant extensions of $f_0 g_0$ and $\{f_0, g_0\}$, respectively.

Example 54.2.19 (Extension of constraints). Consider a first-class constraint ζ_a. A BRST-invariant extension is given by the Poisson bracket

$$G_a := \{-\mathcal{P}_a, \Omega\}\,. \tag{54.68}$$

This immediately shows that the extension G_a corresponds to the observable 0, since it is s-exact and, hence, vanishes in cohomology. If the constraints form a closed algebra, so do their extensions (due to the property above). However, in general, the BRST extensions do not obey any kind of algebra-like condition.

Remark 54.2.20 (Higher cohomology). The higher cohomology groups $H^{\bullet\geq 1}(s)$ do not have a straightforward physical explanation. $H^1(s)$ and $H^2(s)$ are related to symmetry breaking and anomalies.

54.2.5 Actions and gauge fixing

Consider a classical Hamiltonian H_0 and its BRST extension H. Because of Property 54.2.18, Hamiltonian dynamics can be defined on the entire extended phase space:

$$\frac{\mathrm{d}F}{\mathrm{d}t} := \{F, H\}\,. \tag{54.69}$$

For BRST-invariant functions, this is equivalent to the ordinary equations of motion for their associated classical observables. Because these equations do not have any gauge redundancies, they are said to be **gauge fixed**.

Definition 54.2.21 (Gauge fixing). One can change the Hamiltonian H by a BRST-exact term $\{K, \Omega\}$ without changing the cohomology, i.e. without changing the dynamics of BRST-invariant functions. However, the dynamics of noninvariant functions is modified. For this reason, the function K is called the **gauge-fixing fermion** (since it has to be odd for $\{K, \Omega\}$ to be even).

If the gauge-fixing fermion has the form

$$K = K_0 + k^{a_1}(z, \eta)\mathcal{P}_{a_1} + \text{higher antighost terms}\,, \tag{54.70}$$

the transformation is given by $H \to H + k^{a_1}(z, \eta)\mathcal{P}_{a_1}$. Note that the gauge-fixed Hamiltonian might involve multighost terms.

Given a BRST invariant Hamiltonian and gauge-fixing fermion K, a BRST invariant action is given by

$$S_K[z, \eta, \mathcal{P}] := \int_{t_1}^{t_2} \left(\dot{z}^\mu \theta_\mu(z) + \sum_{k=0}^{+\infty} \dot{\eta}^{a_k} \mathcal{P}_{a_k} - H - \{K, \Omega\} \right) dt\,, \tag{54.71}$$

where θ_μ are the coefficients of a symplectic potential (as in Formula 54.1.24).

54.2.6 Nonminimal sectors

In certain situations, it is useful (or even necessary) to enlarge the extended phase space even more, without modifying $H^0(s)$. An example would be generally covariant gauges for relativistic (field) theories.

In their most basic form, nonminimal sectors arise from including cohomologically trivial canonical pairs $(\alpha, \beta, \pi_\alpha, \pi_\beta)$:

$$\begin{aligned} s\alpha &= \beta \\ s\pi_\beta &= \pi_\alpha\,. \end{aligned} \tag{54.72}$$

The BRST generator in this nonminimal sector is then given by $\pi_\alpha \beta$.

Remark 54.2.22. The distinction between minimal and nonminimal is mostly artificial. For example, the trivial condition $\zeta \equiv 0$, characterizing an unconstrained system, can be implemented by either $\Omega = 0$ or $\Omega = \mathcal{P}\pi$. When ζ is not included as an explicit constraint, the latter choice is nonminimal, whereas it becomes minimal when ζ is included.

The nonminimal sectors can also be chosen in such a way that they contain the 'Lagrange multipliers' of the constraints in the extended Hamiltonian (54.18). The function v^a has as canonical momentum a weakly vanishing function b_a. Including these pairs does, accordingly, not change the dynamics of the theory. To preserve BRST invariance, however, one should also include appropriate ghost and antighost fields (and momenta) $(\rho^a, \overline{C}_a)$.

54.2.7 Faddeev–Popov action

Consider a closed constraint algebra:

$$\{\zeta_a, \zeta_b\} = C^c_{ab}\zeta_c \,. \tag{54.73}$$

In this case, the minimal BRST generator was given by

$$\Omega_0 = \eta^a \zeta_a - \frac{1}{2}\eta^a \eta^b C^c_{ba} \mathcal{P}_c \,. \tag{54.74}$$

For scalar functions, the BRST differential acts through gauge variations:

$$\{f, \Omega_0\} = (-1)^{\varepsilon_a} \delta_\eta f \,. \tag{54.75}$$

Note that, whereas the gauge variations usually have an infinitesimal parameter, this is here replaced by the ghosts η. A common (nonminimal) gauge-fixing fermion for Lagrangian field theory is

$$K = i\overline{C}_a \chi^a - \mathcal{P}_a v^a + \frac{i}{2}\overline{C}_a g^{ab} b_b \,, \tag{54.76}$$

where g^{ab} is a 'metric' (it is a proper metric for bosonic theories) and the χ^a's are real functions. This way, the gauge-fixed Hamiltonian contains a quadratic (kinetic) term in the momenta b_a which allows to recover a Lagrangian formulation.[6] Such gauges are called **propagating gauges**, i.e. the ghosts admit a well-defined propagator.

Definition 54.2.23 (Propagating gauge). Propagating gauges, which are quadratic in the conjugate momenta, allow to eliminate these momenta through their own equations of motion. The resulting gauge-fixed action only has BRST invariance on shell.

[6] If the last term were note present, the gauge-fixed Hamiltonian would be linear in b_a and, hence, the momenta can not be eliminated.

This type of BRST transformations are called **Lagrangian BRST transformations.** The Lagrangian BRST invariance has an associated Noether current: the BRST generator Ω.

The equations of motion derived from the nonminimally extended action for the antighosts $\mathcal{P}_a$ and ρ^a are:

$$\dot{\overline{C}}_a - i\mathcal{P}_a i^{\varepsilon_a} = 0 \tag{54.77}$$

and

$$\dot{\eta}^a + (-1)^{\varepsilon_c} v^c \eta^b C^a_{bc} + i\rho^a (-i)^{\varepsilon_a} = 0\,. \tag{54.78}$$

These can be substituted back into the action to obtain the following gauge-fixed expression:

$$\begin{aligned} S'_K[z,v,b,\eta,\rho] &= S_E + S_{\text{gauge breaking}} + S_{\text{ghost}} \\ &= \int_{t_1}^{t_2} \left[\dot{z}^\mu \theta_\mu(z) - H_0 - v^a \zeta_a\right] dt \\ &\quad + \int_{t_1}^{t_2} \left[\dot{v}^a + i^{\varepsilon_a}\chi^a + \frac{(-i)^{\varepsilon_a}}{2} b^a\right] b_a\, dt \\ &\quad - i^{\varepsilon_a+1} \int_{t_1}^{t_2} \overline{C}_a \delta_\eta \left[\dot{v}^a + i^{\varepsilon_a}\chi^a\right] dt\,. \end{aligned} \tag{54.79}$$

This action, in turn, gives the following equations of motion for the momenta b_a conjugate to the Lagrange multipliers v^a:

$$\dot{v}^a + i^{\varepsilon_a}\chi^a + (-i)^{\varepsilon_a} b^a = 0\,. \tag{54.80}$$

Substituting this again, gives the following gauge-fixed ghost action:

$$S''_{\text{ghost}} = -\frac{i^{\varepsilon_a}}{2} \int_{t_1}^{t_2} \left[\left(\dot{v}^a + i^{\varepsilon_a}\chi^a\right)\left(\dot{v}_a + i^{\varepsilon_a}\chi_a\right)\right] dt. \tag{54.81}$$

The above expressions contain three important ingredients:

- the original gauge-invariant action,
- a gauge-symmetry breaking term that determines the Lagrange multipliers and, accordingly, freezes the gauge invariance of the action, and
- a Faddeev–Popov term that is quadratic in the ghosts.

The gauge symmetry-breaking term leads to a derivative gauge:

$$\dot{v}^a + i^{\varepsilon_a}\chi_a = 0\,. \tag{54.82}$$

Property 54.2.24 (Derivative gauges). Derivative gauge can be taken into account in the BRST-BV complex by including the Lagrange multipliers and their momenta as canonical coordinates.

Comparing Eq. (54.79) and Eq. (54.82), one can see that the ghost action contains the 'ghost variation' of the gauge-fixing condition.

Formula 54.2.25 (Ghost action). When the constraint algebra is closed, the Faddeev–Popov ghost Lagangian is of the form

$$L_{\text{ghost}} = \overline{C}_a \delta_\eta F^a \,, \tag{54.83}$$

where F^a are the gauge-fixing conditions. For soft and open algebras, nonquadratic ghost vertices ruin this form. In fact, this is also the case for closed algebras for nonlinear gauges, i.e. when the gauge-fixing fermion is nonlinear in $\mathcal{P}$ and $\overline{C}$.

54.3 Lagrangian BRST theory

When considering (classical) constrained Hamiltonian systems (Section 54.2) and their *quantization* (see Section 59.5), the phase space is extended by both ghosts and antighosts. The former corresponded to differential forms along the gauge orbits (giving a resolution of the quotient by the gauge group) and the antighosts correspond to the Koszul–Tate generators characterizing the zero locus of the field equations (i.e. taking the intersection with the equations of motion). Now, what about Lagrangian field theories, where the solutions are not functions in $C^\infty(\mathbb{R}, M)$, but sections of a vector bundle $E \to M$ that solve the field equations $\frac{\delta S}{\delta \phi} = 0$? As shown below, such theories admit symmetries generated by the equations of motion themselves: the zilch symmetries. One can then play the same game as before, with the ordinary phase space replaced by the covariant phase space Σ and the symmetry group replaced by the full gauge group (including true gauge symmetries and zilch symmetries). As a start, the formalism for finite-dimensional systems will be introduced, i.e. the fields reduce to functions of time. Afterwards, the formalism, originally developed by *Peierls*, will be introduced in the absence of local gauge symmetries.

54.3.1 Field theory

Since the classical notion of phase space as the set of (q, p)-points in coordinate-momentum space, at a given time t, is clearly not covariant (the choice of a time slice ruins any form of relativistic invariance). In the previous section, one possible solution was considered: including time as a canonical coordinate. Here, another approach is embraced which is more convenient in field theory.

Definition 54.3.1 (Covariant phase space). Let S be a local action functional. The covariant phase space Σ associated to S is the set of solutions of the equations of motion

$$\frac{\delta S}{\delta \phi^I} = 0\,, \tag{54.84}$$

i.e. it is exactly the constraint surface.

As before, the physical observables are defined as the smooth functions on this new phase space Σ. These can be described as before. Let $\mathcal{E} \supseteq \Sigma$ be the set of all field histories, e.g. the space of sections $\Gamma(E)$ of a vector bundle. The ring of physical observables $C^\infty(\Sigma)$ is obtained as the quotient of $C^\infty(\mathcal{E})$ by the ideal of functions vanishing on-shell and by the gauge transformations.

54.3.2 Gauge algebra

Consider a local action S on a space M. A **gauge transformation** of S is a general field transformation, depending on M, that leaves the action invariant, i.e. it is a vertical automorphism (Definition 35.1.9) of the field bundle. The most general form of such transformations is

$$\delta_\varepsilon \phi^I = \overline{R}^I_{(0),\alpha}\varepsilon^\alpha + \overline{R}^{I,\mu}_{(1),\alpha}\partial_\mu \varepsilon^\alpha + \cdots + \overline{R}^{I,\mu_1\cdots\mu_s}_{(s),\alpha}\partial_{\mu_1\cdots\mu_s}\varepsilon^\alpha \equiv R^I_\alpha \varepsilon^\alpha\,, \tag{54.85}$$

where the coefficients $\overline{R}_{(i)}$ are general functions of the coordinates and, in the last step, a new shorthand was introduced where the summation over α also includes an integral over x (the **DeWitt convention**):

$$\begin{aligned} R^I_\alpha \varepsilon^\alpha &:= \int R^I_a(x,x')\varepsilon^a(x')\,dx' \qquad (54.86)\\ &= \int \sum_{j,a_j}\Big(\overline{R}^I_{(0),a_j}(x)\delta(x-x') + \overline{R}^I_{(1),a_j}(x)\delta'(x-x') + \overline{R}^I_{(2),a_j}(x)\delta^{(2)}(x-x') + \cdots\Big)\varepsilon^{a_j}(x')\,dx'\,. \end{aligned}$$

Invariance of the action implies that

$$\delta_\varepsilon S = \frac{\delta S}{\delta\phi^I}\delta_\varepsilon\phi^I = \frac{\delta S}{\delta\phi^I}R^I_\alpha\varepsilon^\alpha = 0\,. \tag{54.87}$$

Because this should hold for every value of the transformation parameters ε^α, one immediately obtains the variational Noether identities.

Property 54.3.2 (Noether identities). If a local action is invariant under the transformation (54.85), then

$$\frac{\delta S}{\delta\phi^I}R^I_\alpha = 0 \tag{54.88}$$

for all 'indices' α. In contrast to Noether's theorem 52.2.13, these identities do not imply conserved quantities. Instead, they imply that the equations of motion are not independent (cf. Noether's second theorem 40.1.6).

The structure of the infinitesimal gauge transformations is easily seen to be that of a (real) Lie algebra $\overline{\mathcal{G}}$, whilst that of finite (exponentiated) transformations is a Lie group. However, the gauge algebra is very large (in fact, it is infinite-dimensional) and contains a lot of physically irrelevant information. The simplest example is that of the **zilch symmetries** as referred to by Van Proeyen and Freedman (2012).

Definition 54.3.3 (Zilch symmetry). All transformations of the form

$$\delta_\varepsilon \phi^I = \varepsilon^{IJ} \frac{\delta S}{\delta \phi^J}, \tag{54.89}$$

where ε^{IJ} is antisymmetric, are physically irrelevant since they vanish on-shell by the equations of motion. The trivial gauge transformations form an ideal $\mathcal{N}$ of the gauge algebra and the physically relevant algebra is the quotient $\mathcal{G} := \overline{\mathcal{G}}/\mathcal{N}$. However, for some reasons it might be convenient to retain the full gauge algebra.

Another problem with the gauge algebra is that independent transformations might lead to dependent Noether identities, wich implies that there is still some redundancy. To analyze this issue, one first finds a minimal set of generating transformations.

Definition 54.3.4 (Generating set). A generating set of the gauge algebra (or **complete set of transformations**) is a set of transformations $\delta_\varepsilon \phi^I = R^I_\alpha \varepsilon^\alpha$ such that every gauge transformation can be written as follows:

$$\delta \phi^I = R^I_\alpha \mu^\alpha + M^{IJ} \frac{\delta S}{\delta \phi^J}, \tag{54.90}$$

where μ^α and $M^{IJ} = -M^{JI}$ may depend on the fields. Because the coefficients might be functions of the fields and their derivatives, the generating set is, in general, not a basis for the gauge algebra. However, due to the Lie algebra structure, there must exist structure functions $C^\gamma_{\alpha\beta}$ and $M^{IJ}_{\alpha\beta}$ such that

$$R^J_\alpha \frac{\delta R^I_\beta}{\delta \phi^J} - R^I_\beta \frac{\delta R^J_\alpha}{\delta \phi^I} = C^\gamma_{\alpha\beta} R^I_\gamma + M^{IJ}_{\alpha\beta} \frac{\delta S}{\delta \phi^J}, \tag{54.91}$$

where $M^{IJ}_{\alpha\beta} = -M^{JI}_{\alpha\beta}$.

In Definition 54.1.9, the structure of the contraint algebra was considered. Comparing Eq. (54.17) to Eq. (54.91), it is clear that the constraint algebra and the generating algebra of the gauge transformations play a similar role:

- If all M are zero, the algebra is said to be **closed**. (Even though the generating set itself might not be closed as a Lie algebra because the C's generally are functions of the fields. The algebra is said to be **soft** in this case.)
- Otherwise, it is said to be **open**.

Similarly, A generating set is said to be **irreducible** if there exist no nontrivial combinations of elements:

$$R^I_\alpha \varepsilon^\alpha = M^{IJ} \frac{\delta S}{\delta \phi^J} \implies \varepsilon^\alpha = N^{\alpha I} \frac{\delta S}{\delta \phi^I} . \tag{54.92}$$

The following remark is the Lagrangian counterpart of Remark 54.2.16 in the Hamiltonian treatment of constrained systems.

Remark 54.3.5 (Lie algebroids). If one restricts to closed gauge algebras, i.e. ignores zilch symmetries, Eq. (54.91) is exactly the closure condition for a Lie algebroid (Definition 44.6.4). Higher relations between the generators, i.e. a reducible theory, turn the gauge algebra into a Lie n-algebroid or even a L_∞-algebroid.

54.3.3 Antifields

In the Hamiltonian BRST formalism of Section 54.2, two (co)homological constructions were combined. On the one hand, the quotient of phase space by the constraint algebra, which was modeled by a Chevalley–Eilenberg complex, and, on the other hand, the critical locus of the constraints, modeled by a Koszul–Tate resolution. This led to the extended phase space containing three types of objects:

1. Canonical coordinates: (q, p),
2. Ghosts: η (the Chevalley–Eilenberg generators), and
3. Antighosts: $\mathcal{P}$ (the Koszul–Tate generators).

In the Lagrangian setting, a very similar structure is obtained. However, now, the critical locus is not just the one induced by the Noether identities. One also has to model the equations of motion themselves. This implies that one also has to add antifields, i.e. Koszul–Tate generators corresponding to the field equation $\frac{\delta S}{\delta \phi^I} = 0$. This gives the **BV-BRST complex**.[7]

Definition 54.3.6 (Antifield). For every gauge symmetry $\widehat{G}_a$ in a generating set of the (proper) gauge group, one also introduces a set of antifields $\mathcal{P}_I$. These carry the following cohomological grading:

$$\begin{aligned} \varepsilon(\mathcal{P}_I) &= \varepsilon_I + 1 , \\ \mathsf{antigh}(\mathcal{P}_I) &= \mathsf{antigh}(\phi^I) + 1 = 1 . \end{aligned} \tag{54.93}$$

The antifields $\mathcal{P}_I$ are the Koszul–Tate generators associated to the zilch symmetries:

$$\delta \mathcal{P}_I := -\frac{\delta S}{\delta \phi^I} . \tag{54.94}$$

[7] BV stands for Batalin–Vilkovisky as it did in the case of BFV theory.

The ghost antifields (i.e. the antighosts from Section 54.2) arise as the Koszul–Tate generators induced by the Noether identities (Property 54.3.2), where the higher antifields correspond to reducibility relations among the symmetries, since the Noether identities $\delta(R^I_a \mathcal{P}_I) = 0$ induce elements of $H^1(\delta)$.

If one extends the above grading properties to the antighost fields, one obtains

$$\mathsf{antigh}(\mathcal{P}_a) = 2\,, \tag{54.95}$$

so the antifields are shifted in degree by 1 when compared to the Hamiltonian setting. The Grassmann parity is also shifted compared to the Hamiltonian case:

$$\varepsilon(\mathcal{P}_I) = \varepsilon_I + 1 \qquad \varepsilon(\mathcal{P}_a) = \varepsilon_a + 1 \qquad \cdots \tag{54.96}$$

Notation 54.3.7 (Antifields). A common alternative notation for antifields is ϕ^*_I.

Remark 54.3.8 (Interpretation). An interpretation for this parity shift and the extra antifields is that, in the Lagrangian framework, the field equations can be seen as the fundamental constraints and the Noether identities as reducibility relations. Hence, even without true gauge symmetry, the Lagrangian Koszul–Tate resolution is nontrivial even though the Hamiltonian one was trivial without constraints.

In the Hamiltonian BRST complex, the natural Poisson bracket on phase space could be extended in a convenient way. The ghosts and antighosts paired up to form canonical pairs. However, in the Lagrangian setting, there is no natural Poisson structure. Luckily, there exists a modified bracket structure and, with respect to this new bracket, the BRST generator will still be the generator of canonical transformations.

Definition 54.3.9 (Antibracket). Instead of pairing ghosts and antighosts, there is a new natural pairing due to the apparent symmetry between fields and antifields.

Fields	Antifields
ϕ^I	$\mathcal{P}_I$
η^{a_0}	$\mathcal{P}_{a_0}$
η^{a_1}	$\mathcal{P}_{a_1}$
$\vdots$	$\vdots$

The antibracket is then defined to have the canonical form with respect to this ghost degree pairing. For anytwo functionals f, g on $C^\infty(\Sigma) \otimes \mathbb{C}[\eta^a] \otimes \mathbb{C}[\mathcal{P}_I] \otimes \mathbb{C}[\mathcal{P}_a]$, it is defined as follows:[8]

$$(f,g) := \left(\frac{\partial^R f}{\partial \phi^I} \frac{\partial^L g}{\partial \mathcal{P}_I} - \frac{\partial^R f}{\partial \mathcal{P}_I} \frac{\partial^L g}{\partial \phi^I} \right) + \left(\frac{\partial^R f}{\partial \eta^a} \frac{\partial^L g}{\partial \mathcal{P}_a} - \frac{\partial^R f}{\partial \mathcal{P}_a} \frac{\partial^L g}{\partial \eta^a} \right), \tag{54.97}$$

where the index a denotes ghost (anti)fields of arbitrary degree.

[8] Where the partial derivatives should be read as functional derivatives.

Property 54.3.10 (Odd symplectic structure). As before, if one considers the extended state space with the fields (and ghosts, ...) as 'coordinates' and the antifields as 'momenta', the antibracket gives the structure of an odd symplectic manifold and, in particular, that of a *BV manifold* as defined below:

$$(f,g) = \frac{\partial^R f}{\partial z^\mu}\omega^{\mu\nu}\frac{\partial^L g}{z^\nu}\,, \tag{54.98}$$

where $(z^\mu) \equiv (\phi^I, \eta^a, \mathcal{P}_I, \mathcal{P}_a)$ and

$$(\omega^{\mu\nu}) := \begin{pmatrix} 0 & \delta^I_J + \delta^a_b \\ -\delta^I_J - \delta^a_b & 0 \end{pmatrix}. \tag{54.99}$$

The BV-antibracket, accordingly, has the following algebraic properties:

- It is BRST-odd: $\mathsf{gh}\big((f,g)\big) = \mathsf{gh}(f) + \mathsf{gh}(g) + 1$.
- It induces the structure of a Gerstenhaber algebra (Definition 4.9.4), where the degree is the Grassmann parity.

Formula 54.3.11 (Classical master equation). As in the Hamiltonian framework, the BRST differential can be generated by a canonical transformation. The BRST generator S satisfies the classical master equation (44.8), i.e. it is nilpotent (cf. Eq. (54.52)):

$$(S,S) = 0\,. \tag{54.100}$$

This again allows to recursively solve for S and obtain the following expression for the lowest-order terms:

$$S = S_0 + \mathcal{P}_I R^I_{a_0}\eta^{a_0} + \mathcal{P}_{a_0} Z^{a_0}_{a_1}\eta^{a_1} + \frac{1}{2}C^{a_0}_{b_0c_0}\eta^{b_0}\eta^{c_0} + \text{higher antighost terms}\,. \tag{54.101}$$

Note that the lowest-order term is exactly the initial (gauge-invariant) action.

Remark 54.3.12. In the Hamiltonian setting, the Poisson bracket descended to BRST cohomology (Property 54.2.18) and showed that physical observables also formed a Poisson algebra. Here, the antibracket is of ghost number 1 and, hence, would move outside of $H^0(s)$. Moreover, when the gauge algebra is a proper Lie algebra, (f,g) is always BRST exact, so the antibracket always gives cohomologically trivial results in this case.

Property 54.3.13 (Gauge invariance). Let there be $n \in \mathbb{N}$ fields (including ghosts, ghost-of-ghosts, etc.). The BRST generator S admits exactly n independent gauge symmetries, given by

$$\delta z^\mu = \omega^{\mu\nu}\frac{\partial^L\partial^R}{\partial z^\nu \partial z^\rho}\varepsilon^\rho\,. \tag{54.102}$$

The $2n$ symmetries one would naively expect (due to presence of the antifields) are reduced in number to n because the transformation matrix

$$R^{\mu}_{\rho} := \omega^{\mu\nu} \frac{\partial^L \partial^R}{\partial z^\nu \partial z^\rho} \tag{54.103}$$

is on-shell nilpotent, i.e. the symmetries are on-shell reducible. Note that every gauge symmetry of the BRST action S can be obtained by differentiating it, with one symmetry corresponding to every field-antifield pair!

The above structure can be generalized as follows.

Definition 54.3.14 (BV manifold). A **Batalin–Vilkovisky manifold** is a triple (M, ω, S) where M is a graded manifold, ω is a degree-1 symplectic form and S is a degree-0 function such that the classical master equation (44.8) is satisfied:

$$(S, S) = 0\,, \tag{54.104}$$

where $\{\cdot,\cdot\}$ is the Poisson bracket induced by ω (Definition 44.1.9).

The most straightforward example of a BV manifold is, of course, the BV-BRST complex associated to a field theory as introduced above. For this reason, the Poisson bracket is more generally called the **antibracket**, while the grading is generally called the **ghost number** and denoted by gh.

Property 54.3.15 (Multivector fields). By Corollary 44.1.11, BV manifolds can be represented as spaces of multivector fields with the Poisson bracket mapping to the Schouten–Nijenhuis bracket. For the BV-BRST complex, the vector fields are given by the antifields. This correspondence can be seen by noting that, when the fields transform as

$$\phi^I \longrightarrow \widetilde{\phi}^J(\phi)\,, \tag{54.105}$$

the antifields have to transform as

$$P_I \longrightarrow \widetilde{\mathcal{P}}_J(\phi, \widetilde{\phi}) = \widetilde{\mathcal{P}}_I \frac{\delta^R \phi^I}{\delta \widetilde{\phi}_J} \tag{54.106}$$

to preserve the canonical structure.

Example 54.3.16 (AKSZ model). The Alexandrov–Kontsevich–Schwarz–Zabronsky model is the (noninear) σ-model on a dg-manifold (M, Q) with target space a dg-symplectic manifold (N, ω, X_H), where X_H is Hamiltonian.

For any graded manifold Σ, one can construct the source manifold by taking $M := \Pi T\Sigma$ and $Q := \mathsf{d}$. A symplectic form on $C^\infty(M, N)$ is then given by

$$\Omega := \int_{\Pi T\Sigma} \omega_{\mu\nu} \delta\Phi^\mu \delta\Phi^\nu \ \mathsf{Vol}\ . \tag{54.107}$$

The BV action is defined as follows:

$$S := \int_{\Pi T\Sigma} \left(\alpha_\mu d\Phi^\mu + \Theta\right) \mathsf{Vol}\,, \tag{54.108}$$

where α is a symplectic potential for ω, which necessarily exists globally by Property 44.1.10 if $\mathsf{gh}(\omega) \neq 0$.

In general, the symplectic form on $C^\infty(M,N)$ is induced from that on N by a pull-push operation. First, one pulls back this form along the evaluation map $\mathsf{ev} : C^\infty(M,N)\times M \to N$ and, then, one pushes it forward along the projection on the first factor (cf. fibre integration (Definition 34.9.24)).

@@ COMPLETE (check S for example) @@

Chapter 55

Electromagnetism

Chapter Contents

55.1 Electricity **1033**
55.1.1 Conductors and resistors 1034
55.1.2 Ohm's law 1036
55.1.3 Electric dipoles 1036
55.2 Magnetism **1038**
55.2.1 Lorentz force 1039
55.2.2 Energy and momentum 1039
55.2.3 Electric charges in a magnetic field 1040
55.3 Maxwell equations **1040**
55.4 Potentials **1041**
55.4.1 Decomposition in potentials 1041
55.4.2 Gauge theory 1041
55.4.3 Differential-geometric perspective 1042

55.1 Electricity

Definition 55.1.1 (Coulomb). The SI unit of charge is the coulomb (symbol: C). One coulomb equals the amount of charge that passes through a conductor when the electric current equals one ampère (1A).

Property 55.1.2 (Conservation of charge). Charge is a conserved quantity.

Formula 55.1.3 (Coulomb's law). The force between two point charges q_1, q_2 is given by

$$F = \frac{q_1 q_2}{4\pi\varepsilon_0 r^2}, \tag{55.1}$$

where r is the distance between the charges and ε_0 is the permittivity of the vacuum ($\varepsilon_0 \approx 8.854\text{e}{-12}\ \frac{\text{F}}{\text{m}}$).

Definition 55.1.4 (Electric potential). The electric field gives the force experienced by a (positive) unit charge. The electric potential is then defined as the work done by this charge when moving through the field:

$$V(\vec{\mathbf{a}}, \vec{\mathbf{b}}) := \int_{\vec{\mathbf{a}}}^{\vec{\mathbf{b}}} \vec{\mathbf{E}} \cdot d\vec{\mathbf{l}}. \tag{55.2}$$

By fixing an endpoint (some fixed reference location[1]), the electric field and electric potential are related in the same way as an ordinary potential and its associated (conservative) force (cf. Eq. (52.4)):

$$E = -\nabla V. \tag{55.3}$$

Definition 55.1.5 (Capacitance). The capacitance is a (geometrical) property that reflects the amount of charge an object can store:

$$C := \frac{Q}{V}. \tag{55.4}$$

55.1.1 Conductors and resistors

Definition 55.1.6 (Current). The amount of charge that passes through a given point in one second:

$$I := \frac{\mathrm{d}Q}{\mathrm{d}t}. \tag{55.5}$$

Formula 55.1.7 (Power). Given that the electric potential determines the work done by a positive charge, the electric power of a current I is given by

$$P = VI. \tag{55.6}$$

Property 55.1.8 (Electric field of conductor). The electric field inside a conductor vanishes at equilibrium: all free charges are located at the surface. Moreover, the electric field at the surface points along the normal vector.

Definition 55.1.9 (Drift velocity). The average velocity of independent charge carriers is the drift velocity $\vec{\mathbf{v}}_d$. It is important to note that v_d is not equal to the propagation speed of the electric signal, which is the result of particle collisions. In fact, it is several orders of magnitude smaller!

[1] As usual, the potential is only determined up to some constant.

Definition 55.1.10 (Mobility).

$$\mu := \frac{v_d}{E}\,, \tag{55.7}$$

where E is the external electric field.

Definition 55.1.11 (Conductivity).

$$\sigma := nq\mu\,, \tag{55.8}$$

where $n \in \mathbb{N}$ is the number of charge carriers.

Definition 55.1.12 (Resistivity).

$$\rho := \frac{1}{\sigma} \tag{55.9}$$

Formula 55.1.13 (Pouillet's law). The resistance of a wire is given by the following formula:

$$R = \frac{\rho l}{A}\,, \tag{55.10}$$

where

- ρ is the resistivity of the material,
- l is the length of the resistor, and
- A is the cross-sectional area of the resistor.

Property 55.1.14 (Resistor circuits). Consider an electrical circuit of the form

The potential difference over both resistors is equal:

$$R_1 I_1 = R_2 I_2\,. \tag{55.11}$$

Theorem 55.1.15 (Kirchhoff's first law). *The oriented (algebraic)*[2] *sum of currents meeting at any point is zero.*

Theorem 55.1.16 (Kirchhoff's second law). *The oriented sum of potential differences in a closed circuit is zero.*

[2]Currents out of a point are counted with a minus sign.

55.1.2 Ohm's law

Formula 55.1.17 (Free current density). The current density generated by free charges is given by

$$\vec{\mathbf{J}} = nq\vec{\mathbf{v}}_d\,. \tag{55.12}$$

Formula 55.1.18 (Ohm's law).

$$\vec{\mathbf{J}} = \sigma \cdot \vec{\mathbf{E}}\,, \tag{55.13}$$

where σ is the conductivity tensor. (Compare this to Definition 55.1.10 and Definition 55.1.11 for 1D systems.)

Formula 55.1.19 (Ohm's law in wires). The following formula can be obtained by combining Definition 55.1.12 and Formulas 55.1.13 and 55.1.18, and by assuming that the conductivity tensor is a scalar (this follows from the isotropic behaviour of most resistors):

$$V = RI\,. \tag{55.14}$$

55.1.3 Electric dipoles

Formula 55.1.20 (Electric dipole).

$$\vec{\mathbf{p}} := q\vec{\mathbf{l}}\,, \tag{55.15}$$

where

- q is the positive charge, and
- $\vec{\mathbf{l}}$ is the vector pointing from the negative charge to the positive charge.

Assume that the positive charge is located at the origin and that the vector $\vec{\mathbf{l}}$ points in the direction of the x-axis, i.e. $\vec{\mathbf{l}} = l\hat{\mathbf{e}}_x$ for some $l > 0$ (the negative charge is, hence, located to the left of the origin). The electric potential of the dipole is given by:

$$V(\vec{\mathbf{r}}) = \frac{1}{4\pi\varepsilon_0}\left(\frac{q}{r} - \frac{q}{r + l\cos(\theta)}\right) = \frac{q}{4\pi\varepsilon_0}\frac{l\cos(\theta)}{r\big(r + l\cos(\theta)\big)}\,, \tag{55.16}$$

where θ is the angle between $\hat{\mathbf{e}}_x$ and $\vec{\mathbf{r}} + \vec{\mathbf{l}}$ (see Fig. 55.1). In the far-field regime, where $l\cos(\theta) \ll r$, the potential is given by

$$V(\vec{\mathbf{r}}) \approx \frac{q}{4\pi\varepsilon_0}\frac{l\cos(\theta)}{r^2}\,, \tag{55.17}$$

i.e. it decays quadratically.

Figure 55.1: Electric dipole configuration.

Formula 55.1.21 (Energy). If an electric dipole is placed in an external electric field, its potential energy is given by

$$V = -\vec{\mathbf{p}} \cdot \vec{\mathbf{E}}\,. \tag{55.18}$$

Formula 55.1.22 (Torque). If an electric dipole is placed in an external electric field, the torque on this system is given by

$$\vec{\boldsymbol{\tau}} = \vec{\mathbf{p}} \times \vec{\mathbf{E}}\,. \tag{55.19}$$

Formula 55.1.23 (Multipole expansion). Consider a general charge distribution $\rho : \mathbb{R}^3 \to \mathbb{R}$ with support in a bounded region Γ and assume that $\|\vec{\mathbf{r}}\| \gg \mathsf{diam}(\Gamma)$. If $\vec{\mathbf{r}}$ is oriented along the z-axis, one can write the (squared) distance to an infinitesimal volume element as

$$R^2 = r^2 + (r')^2 - 2rr'\cos(\theta) \tag{55.20}$$

or

$$\frac{1}{R} = \frac{1}{r}\frac{1}{\sqrt{1 - \left(\frac{r'}{r}\right)^2 - 2\left(\frac{r'}{r}\right)\cos(\theta)}}\,, \tag{55.21}$$

where $\vec{\mathbf{r}}'$ is the position of the volume element and θ is the polar angle of this position vector. In the general integral

$$V(\vec{\mathbf{r}}) = \frac{1}{4\pi\varepsilon_0}\int \frac{\rho(\vec{\mathbf{r}}')}{\|\vec{\mathbf{r}} - \vec{\mathbf{r}}'\|}\,dV'\,, \tag{55.22}$$

one can now expand the denominator in powers of $1/r$,

$$V(\vec{\mathbf{r}}) = \frac{1}{4\pi\varepsilon_0}\left(\frac{K_0}{r} + \frac{K_1}{r^2} + \frac{K_2}{r^3} + \cdots\right). \tag{55.23}$$

Since the second factor in Eq. (55.21) is the generating function of Legendre polynomials, one obtains:

- $K_0 := \displaystyle\int \rho(\vec{\mathbf{r}})\,dV$

- $K_1 := \int r\cos(\theta)\rho(\vec{\mathbf{r}})\,dV$
- $K_2 := \int \frac{3\cos^2(\theta)-1}{2} r^2\rho(\vec{\mathbf{r}})\,dV$
- $\cdots$
- $K_n := \int P_n\big(\cos(\theta)\big) r^n \rho(\vec{\mathbf{r}})\,dV,$

with P_n the nth *Legendre polynomial*. The first term gives an approximation where all charge is concentrated at a single point (this is called the **monopole approximation**). At very large distances, where the charge distribution appears as a single point, this gives a good approximation.

When the position vector $\vec{\mathbf{r}}$ is not assumed to be oriented along the z-axis, the Legendre polynomials have to be replaced by spherical harmonics.

55.2 Magnetism

Definition 55.2.1 (Tesla). The SI unit of magnetic field strength/magnetic induction is the tesla (symbol: T). It is defined such that a charge of one coulomb (1C) moving at one meter per second experiences a force of one newton (1N) when the magnetic induction equals one tesla: $\mathrm{T} = \frac{\mathrm{N}}{\mathrm{A}}$.

Definition 55.2.2 (Magnetic field). The **magnetizing field** $\vec{\mathbf{H}}$ is the field generated by all external sources. When applying an external (magnetic) field, some materials will try to oppose this external influence. Similar to polarization in the case of electricity, one can define the **magnetization** (in linear media):

$$\vec{\mathbf{M}} := \chi\vec{\mathbf{H}}, \tag{55.24}$$

where χ is the **magnetic susceptibility**.

@@ ADD polarization @@

The **magnetic induction** $\vec{\mathbf{B}}$ is the field generated by both the external sources and the internal magnetization. This is the total magnetic field and, hence, the only field that can be measured. In vacuum, the following relation between the magnetic induction, the magnetizing field and the magnetization exists:

$$\vec{\mathbf{B}} = \mu_0\Big(\vec{\mathbf{H}} + \vec{\mathbf{M}}\Big). \tag{55.25}$$

By combining the previous two formulas, one obtains (again only in linear media)

$$\vec{\mathbf{B}} = \mu_0(1+\chi)\vec{\mathbf{H}}. \tag{55.26}$$

The proportionality constant in this formula is called the **magnetic permeability**:

$$\mu := \mu_0(1+\chi)\,, \tag{55.27}$$

where μ_0 is the magnetic permeability of the vacuum. The factor $1+\chi$ is called the **relative permeability** and it is often denoted by μ_r.

Remark 55.2.3 (Tensorial formulation). In anisotropic materials, one has to use a tensorial formulation:

$$B_i = \sum_j \mu_{ij} H_j\,, \tag{55.28}$$

$$M_i = \sum_j \chi_{ij} H_j\,. \tag{55.29}$$

In this case, both μ and χ are $(1,1)$-tensors, i.e. linear maps.

@@ CHECK (linear/tensorial) @@

55.2.1 Lorentz force

The relation between the force acting on a charged particle and the strength of the electric and magnetic fields is given by the following law.

Formula 55.2.4 (Lorentz force).

$$\vec{\mathbf{F}} = q\left(\vec{\mathbf{E}} + \vec{\mathbf{v}} \times \vec{\mathbf{B}}\right) \tag{55.30}$$

The charge of a particle acts as the constant of proportionality, just as the mass in the case of gravity.

Formula 55.2.5 (Lorentz force density).

$$\vec{\mathbf{f}} = \rho\vec{\mathbf{E}} + \vec{\mathbf{J}} \times \vec{\mathbf{B}} \tag{55.31}$$

55.2.2 Energy and momentum

Definition 55.2.6 (Poynting vector).

$$\vec{\mathbf{S}} := \vec{\mathbf{E}} \times \vec{\mathbf{H}} \tag{55.32}$$

Formula 55.2.7 (Energy density).

$$W = \frac{1}{2}\left(\vec{\mathbf{E}} \cdot \vec{\mathbf{D}} + \vec{\mathbf{B}} \cdot \vec{\mathbf{H}}\right) \tag{55.33}$$

55.2.3 Electric charges in a magnetic field

Solving the Lorentz force law in a uniform magnetic field gives the following solution.

Formula 55.2.8 (Gyroradius). Consider a charge q moving with velocity $\vec{\mathbf{v}}$ in a magnetic field B such that $\vec{\mathbf{v}} \perp \vec{\mathbf{B}}$. The charge will follow a circular path with radius

$$r = \frac{mv_\perp}{|q|B}\,. \tag{55.34}$$

Formula 55.2.9 (Gyrofrequency[3]). The circular velocity is given by

$$\omega = \frac{|q|B}{m}\,. \tag{55.35}$$

55.3 Maxwell equations

Formula 55.3.1 (Gauss's law for electricity).

$$\nabla \cdot \vec{\mathbf{E}} = \frac{\rho}{\varepsilon} \tag{55.36}$$

Formula 55.3.2 (Gauss's law for magnetism).

$$\nabla \cdot \vec{\mathbf{B}} = 0 \tag{55.37}$$

If magnetic monopoles would exist, the right-hand side would contain a (magnetic) charge density as in the case of electricity.

Formula 55.3.3 (Faraday's law).

$$\nabla \times \vec{\mathbf{E}} = -\frac{\partial \vec{\mathbf{B}}}{\partial t} \tag{55.38}$$

Formula 55.3.4 (Maxwell's law[4]).

$$\nabla \times \vec{\mathbf{B}} = \varepsilon\mu \frac{\partial \vec{\mathbf{E}}}{\partial t} + \mu \vec{\mathbf{J}} \tag{55.39}$$

Property 55.3.5 (Electric-magnetic duality). Note that in vacuum, where $\rho = 0$ and $\vec{\mathbf{J}} = 0$, there exist a duality between the electric and magnetic fields:

$$\vec{\mathbf{E}} \longrightarrow \vec{\mathbf{B}} \qquad\qquad \vec{\mathbf{B}} \longrightarrow -\vec{\mathbf{E}}\,. \tag{55.40}$$

(To obtain this duality with matter, one would need to include magentic charges, but magnetic monopoles have not been observed.)

In vacuum, one can consider the unified field $\vec{\mathcal{E}} := \vec{\mathbf{E}} + i\vec{\mathbf{B}}$ and duality then boils down to

$$\vec{\mathcal{E}} \longrightarrow i\vec{\mathcal{E}}\,. \tag{55.41}$$

[3] Also called the **Larmor** or **cyclotron frequency**.
[4] Also called the **Maxwell–Ampère law**.

@@ EXPAND ON EM-duality @@

Property 55.3.6 (Continuity equation). By applying Property 23.1.12, Maxwell's equations also readily give rise to the continuity equation (Eq. (52.64)):

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \vec{\mathbf{J}} = 0\,. \tag{55.42}$$

55.4 Potentials

55.4.1 Decomposition in potentials

The Helmholtz decomposition 23.1.14 together with Gauss's law 55.3.2 imply the following general form for $\vec{\mathbf{B}}$:

$$\vec{\mathbf{B}} = \nabla \times \vec{\mathbf{A}}\,, \tag{55.43}$$

where $\vec{\mathbf{A}}$ is called the **magnetic potential**.

Combining Eq. (55.43) with Faraday's law 55.3.3 (and doing some rewriting), gives the following general form for $\vec{\mathbf{E}}$:

$$\vec{\mathbf{E}} = -\nabla V - \frac{\partial \vec{\mathbf{A}}}{\partial t}\,, \tag{55.44}$$

where V is called the **electrostatic potential**.

Property 55.4.1. Substituting the previous two expressions into Formula 55.3.1 and Formula 55.3.4, gives the following two (coupled) conditions for the electromagnetic potentials:

$$\Delta \vec{\mathbf{A}} - \varepsilon\mu \frac{\partial^2 \vec{\mathbf{A}}}{\partial t^2} = \nabla\left(\nabla \cdot \vec{\mathbf{A}} + \varepsilon\mu \frac{\partial V}{\partial t}\right) - \mu \vec{\mathbf{J}}\,, \tag{55.45}$$

$$\Delta V - \varepsilon\mu \frac{\partial^2 V}{\partial t^2} = -\frac{\partial}{\partial t}\left(\nabla \cdot \vec{\mathbf{A}} + \varepsilon\mu \frac{\partial V}{\partial t}\right) - \frac{\rho}{\varepsilon}\,. \tag{55.46}$$

55.4.2 Gauge theory

Looking at Eq. (55.43), it is clear that a transformation

$$\vec{\mathbf{A}} \longrightarrow \vec{\mathbf{A}} + \nabla\chi \tag{55.47}$$

has no effect on $\vec{\mathbf{B}}$ due to Eq. (23.12). To compensate for this transformation in Eq. (55.44), one also has to perform a transformation

$$V \longrightarrow V - \frac{\partial \chi}{\partial t}\,. \tag{55.48}$$

These transformations are called **gauge transformations** (compare to the previous chapter).

Definition 55.4.2 (Gauge fixing conditions). Conditions that fix a certain gauge (or class of gauge transformations). These select one of many physically equivalent configurations. Mathematically this corresponds to picking a representative for any orbit of the gauge transformations. (See Chapter 69 for more information.)

Remark 55.4.3. It often happens that a gauge condition does not completely fix the gauge, i.e. the condition does not pick a unique representative but only fixes a subset of the gauge orbit that still has some remaining symmetry or freedom.

Example 55.4.4 (Lorenz gauge). A first example of a gauge fixing condition is the Lorenz gauge[5]:

$$\nabla \cdot \vec{\mathbf{A}} + \varepsilon\mu \frac{\partial V}{\partial t} = 0. \tag{55.49}$$

When using this gauge fixing condition, Eqs. (55.45) and (55.46) uncouple. This allows to rewrite them as

$$\Box \vec{\mathbf{A}} = -\mu \vec{\mathbf{J}}\,, \tag{55.50}$$

$$\Box V = -\frac{\rho}{\varepsilon}\,, \tag{55.51}$$

which, in vacuum, are simply wave equations!

To see which gauge functions χ satisfy the Lorenz condition, perform a transformation as explained above:

$$\vec{\mathbf{A}}' = \vec{\mathbf{A}} + \nabla\chi \qquad \text{and} \qquad V' = V - \frac{\partial \chi}{\partial t}\,. \tag{55.52}$$

Substituting these transformations in Eq. (55.49) and using the fact that both sets of potentials $(\vec{\mathbf{A}}, V)$ and $(\vec{\mathbf{A}}', V')$ satisfy the Lorenz gauge (55.49), gives the following condition for the gauge function χ:

$$\Box\chi = 0\,. \tag{55.53}$$

Example 55.4.5 (Coulomb gauge).

$$\nabla \cdot \vec{\mathbf{A}} = 0 \tag{55.54}$$

55.4.3 Differential-geometric perspective

Using the tools introduced from Chapter 34 (e.g. differential forms), one can rewrite all of the above formulas in a more elegant form. This will also allow to generalize them to higher dimensions and to more general settings. See, for example, Sontz (2015a) for a complete derivation and interpretation.

[5]Named after *L. Lorenz*, not to be confused with *H. Lorentz*.

Definition 55.4.6 (Field strength). Define the electric and magnetic field forms as follows:

$$\mathbf{E} = E_1 \mathrm{d}x^1 + E_2 \mathrm{d}x^2 + E_3 \mathrm{d}x^3 \tag{55.55}$$

and

$$\mathbf{B} = B_1 \mathrm{d}x^2 \wedge \mathrm{d}x^3 + B_2 \mathrm{d}x^3 \wedge \mathrm{d}x^1 + B_3 \mathrm{d}x^1 \wedge \mathrm{d}x^2 \,. \tag{55.56}$$

Using these differential forms, one can write the field strength as follows (cf. Eq. (55.41)):

$$\mathbf{F} = \mathbf{B} - \mathrm{d}t \wedge \mathbf{E} \,. \tag{55.57}$$

Formula 55.4.7 (Maxwell's equations). Define the electric 4-current as

$$\mathbf{J} = \rho \mathrm{d}t - J_1 \mathrm{d}x^1 - J_2 \mathrm{d}x^2 - J_3 \mathrm{d}x^3 . \tag{55.58}$$

Maxwell's equations (Section 55.3) can now be rewritten as follows:

$$\mathrm{d}\mathbf{F} = 0 \,, \tag{55.59}$$

$$*\mathrm{d}*\,\mathbf{F} = \mathbf{J} \,, \tag{55.60}$$

where $*$ is the Hodge operator 23.4.27. (Note that the second line is essentially the codifferential from Definition 36.1.4.) Moreover, electric-magnetic duality corresponds to

$$\mathbf{F} \longrightarrow *\mathbf{F} \tag{55.61}$$

in this notation.

Definition 55.4.8 (Potential). The Bianchi identity (55.59) and Poincaré's lemma 34.9.8 imply that there exists a differential one-form $\mathbf{A}$ such that

$$\mathbf{F} = \mathrm{d}\mathbf{A} \,. \tag{55.62}$$

This one-form is called the potential or **gauge field**. It can be related to the ordinary scalar potential V and vector potential $\vec{\mathbf{A}}$ as follows:

$$\mathbf{A} = -V \mathrm{d}t + A_1 \mathrm{d}x^1 + A_2 \mathrm{d}x^2 + A_3 \mathrm{d}x^3 \,. \tag{55.63}$$

Property 55.4.9 (Gauge transformation). Because $\mathrm{d}^2 \equiv 0$, the above equation is invariant under a transformation $\mathbf{A} \longrightarrow \mathbf{A} + \mathrm{d}f$ for any $f \in C^\infty(\mathbb{R}^3)$. When written out in local coordinates, this gives exactly the gauge transformations from Eq. (55.52).

@@ COMPLETE (perhaps move to chapter on gauge theory) @@

Chapter 56

Optics

Chapter Contents

56.1 General **1044**

56.1.1 Conservation of energy 1044

56.2 Wave mechanics **1045**

56.3 Electromagnetic radiation **1046**

56.4 Propagation in media **1048**

56.4.1 Reflection 1048

56.4.2 Refraction 1048

56.4.3 Absorption 1049

56.4.4 Interference & Diffraction 1050

56.1 General

56.1.1 Conservation of energy

From the law of conservation of energy, one can derive the following formula:

$$T + R + A = 1\,, \tag{56.1}$$

where:

- T is the transmission coefficient,
- R is the reflection coefficient, and
- A is the absorption coefficient.

56.2 Wave mechanics

Formula 56.2.1 (d'Alemberts formula). Consider the wave equation (Formula 21.1.3):

$$\Box f = 0\,. \tag{56.2}$$

By applying the method from Section 21.4, it is clear that the characteristics are given by

$$\xi = x + ct \qquad \text{and} \qquad \eta = x - ct\,. \tag{56.3}$$

Furthermore, it follows that the wave equation is a hyperbolic equation and, accordingly, can be rewritten in the canonical form

$$\frac{\partial^2 u}{\partial \xi \partial \eta}(\xi, \eta) = 0\,. \tag{56.4}$$

Integration with respect to ξ and η, and rewriting the solution in terms of x and t gives

$$u(x,t) = f(x+ct) + g(x-ct)\,, \tag{56.5}$$

where $f, g \in C^\infty(\mathbb{R})$ are arbitrary. This solution represents a superposition of a left-moving and a right-moving wave.

Now, consider the wave equation subject to the general boundary conditions

$$u(x,0) = v(x) \qquad \text{and} \qquad \frac{\partial u}{\partial t}(x,0) = q(x)\,. \tag{56.6}$$

By inserting these conditions in the solution (56.5), it can be shown that the general solution, subject to the given boundary conditions, is given by

$$u(x,t) := \frac{1}{2}\big(v(x+ct) + v(x-ct)\big) + \frac{1}{2c}\int_{x-ct}^{x+ct} q(z)\,dz\,. \tag{56.7}$$

Definition 56.2.2 (Harmonic wave). The wave form obtained by setting

$$v(x \pm vt) = A \sin\left(\frac{2\pi}{\lambda}(x \pm vt)\right) \tag{56.8}$$

in d'Alembert's formula, where the constants A, λ are called the **amplitude** and **wave length**, respectively. The wave length determines the distance between consequent minima/maxima.

This form can be expressed in various other forms:

$$\begin{aligned}
\xi(x,t) &= A \sin\left(\frac{2\pi}{\lambda}(x \pm vt)\right) \\
&= A \sin\left(2\pi\left(\frac{x}{\lambda} \pm \nu t\right)\right) \\
&= A \sin\left(2\pi\left(\frac{x}{\lambda} \pm \frac{t}{\tau}\right)\right) \\
&= A \sin\big(kx \pm \omega t\big)\,,
\end{aligned}$$

with

$$k := \frac{2\pi}{\lambda} \qquad \text{and} \qquad \omega := 2\pi\nu\,. \tag{56.9}$$

The quantities ω, ν, τ and k are, respectively, called the **circular frequency, frequency, period** and **wave number**. The speed at which such a wave moves is called the **phase velocity**:

$$v := \lambda\nu\,. \tag{56.10}$$

One can also introduce an additional constant such that the wave form is shifted in space/time:

$$\xi(x,t) = A\sin\big(\omega t \pm kx + \varphi\big)\,. \tag{56.11}$$

This constant φ is called a **phase** (shift).

Definition 56.2.3 (Group velocity). Consider a superposition of two harmonic waves with equal amplitudes:

$$\begin{aligned} u(x,t) &= A\cos(kx-\omega t) + A\cos(k'x-\omega' t) \\ &= 2A\cos\left(\frac{k'-k}{2}x - \frac{\omega'-\omega}{2}t\right)\cos\left(\frac{k+k'}{2}x - \frac{\omega+\omega'}{2}t\right). \end{aligned}$$

When $\omega \neq \omega'$ or $k \neq k'$, the resulting wave is not purely harmonic. The result is a combination of a **carrier wave** with phase velocity $\frac{\omega+\omega'}{k+k'}$ and an **envelope** with phase velocity $\frac{\omega-\omega'}{k-k'}$. The phase velocity of the envelope is called the group velocity. This is the velocity with which the information contained in the modulated signal is transmitted.

When the wave numbers are infinitesimally separated ($k \approx k'$), the group velocity can be written as a derivative:

$$v_g := \frac{\mathrm{d}\omega}{\mathrm{d}k}\,. \tag{56.12}$$

This is the definition of the group velocity of an arbitrary wave packet

$$u(x,t) = \int_{-\infty}^{+\infty} A(k)e^{i(kx-\omega t)}\,dk\,. \tag{56.13}$$

Method 56.2.4 (Huygens–Fresnel principle). Every point of the wave front acts as a source for 'secondary' spherical waves. The wave front at later points is obtained by taking the boundary of the convex hull of secondary waves emitted at the same time.

56.3 Electromagnetic radiation

Consider Maxwell's equations from Section 55.3. By inserting Maxwell's law in the rotor of Faraday's law, the following equation for the electric field is obtained:

$$\Delta\vec{\mathbf{E}} = \varepsilon_0\mu_0\frac{\partial^2\vec{\mathbf{E}}}{\partial t^2}\,. \tag{56.14}$$

This has the form of the wave equation (21.1.3) with speed $c^2 = \frac{1}{\varepsilon_0 \mu_0}$. A similar derivation gives rise to a wave equation of the magnetic field. Moreover, through the Maxwell equations, the solutions of these wave equations are coupled. For example, if one considers harmonic solutions:

$$\vec{\mathbf{E}}(x,t) = \vec{\mathbf{E}}_0 \sin(kx - \omega t)\,,$$
$$\vec{\mathbf{B}}(x,t) = \vec{\mathbf{B}}_0 \sin(kx - \omega t)\,,$$

Faraday's law gives

$$\vec{\mathbf{E}}_0 k \cos(kx - \omega t) = \vec{\mathbf{B}}_0 \omega \cos(kx - \omega t)\,. \tag{56.15}$$

This in turns gives rise to the following important identity:

$$\frac{\|\vec{\mathbf{E}}_0\|}{\|\vec{\mathbf{B}}_0\|} = \frac{\omega}{k} = c\,, \tag{56.16}$$

i.e. the pahe velocity of electromagnetic waves is a constant: the speed of light!

Formula 56.3.1 (Energy).

$$E = h\nu = \hbar\omega = \frac{hc}{\lambda} \tag{56.17}$$

Formula 56.3.2 (Momentum).

$$p = \frac{h}{\lambda} = \hbar k \tag{56.18}$$

Definition 56.3.3 (Polarization). The polarization of an electromagnetic wave is defined as the direction of the $\vec{\mathbf{E}}$-component.

Consider a superposition of two orthogonally polarized plane waves:

$$\vec{\mathbf{E}}_1(x,t) = E_1 \sin(kx - \omega t)\hat{\mathbf{e}}_y\,,$$
$$\vec{\mathbf{E}}_2(x,t) = E_2 \sin(kx - \omega t + \varphi)\hat{\mathbf{e}}_z\,.$$

If the phase shift φ vanishes, the superposition is **linearly polarized**, because the resulting polarization is fixed. If the phase shift is $\varphi = \pm\frac{\pi}{2}$, i.e. the waves are perfectly out of phase, the resulting wave is said to be **elliptically polarized**. The resulting polarization vector rotates in the yz-plane. When the amplitudes coincide, elliptically polarized light is also said to be **circularly polarized**, since the polarization vector draws a circular in the yz-plane.

Consider Maxwell's equations 55.4.7. In 4D, the electromagnetic field strength is a 2-form and, hence, one can decompose it in a selfdual and antiselfdual component:

$$\mathbf{F} = \mathbf{F}_+ + \mathbf{F}_-\,, \tag{56.19}$$

where

$$\mathbf{F}_+ := \frac{1}{2}(\mathbf{F} + *\mathbf{F}) \qquad \text{and} \qquad \mathbf{F}_- := \frac{1}{2}(F - *\mathbf{F})\,. \tag{56.20}$$

If $\mathbf{F}$ is taken to be a plane wave, then $\mathbf{F}_\pm$ corresponds to *left-circularly polarized* and *right-circularly polarized* waves, respectively.

56.4 Propagation in media

56.4.1 Reflection

@@ COMPLETE @@

56.4.2 Refraction

Formula 56.4.1 (Refractive index). The relative refractive index between two materials is defined as follows:

$$n_{12} := \frac{v_1}{v_2}\,, \tag{56.21}$$

where v_1, v_2 are the speeds of light in the first and second material, respectively. If the first material is chosen to be the vacuum, i.e. $v_1 = c$, the **absolute refractive index** n is obtained.

Formula 56.4.2 (Dielectric function). In the case of nonmagnetic materials ($\mu_r \approx 1$), one can write the dielectric function as follows:

$$\epsilon = \epsilon_r + i\epsilon_i =: \tilde{n}^2 = (n + ik)^2\,, \tag{56.22}$$

where $\tilde{n}$ is called the **complex refractive index** and k the **extinction coefficient**.

Theorem 56.4.3 (Malus–Dupin). *Light rays are always perpendicular to the wave fronts.*

Property 56.4.4 (Reflection and refraction laws). For plane waves, reflection and refraction are characterized by the following three rules:

1. The normals to the interface and the incoming, reflected and refracted rays all lie in one plane.
2. The incoming and reflected angles coincide.
3. The incoming and refracted angles satisfy **Snell's law**:

$$n_1 \sin(\theta_i) = n_2 \sin(\theta_2)\,. \tag{56.23}$$

The second and third item can be derived from both Huygens' principle and Malus's law or Fermat's principle.

56.4.3 Absorption

Theorem 56.4.5 (Law of Lambert–Beer). *Let $I(0)$ be the intensity of a light beam incident on a material. After penetrating a distance x in the material, the intensity is given by*

$$I(x) = I(0)\exp\left(-\frac{4\pi\nu k}{c}x\right); . \tag{56.24}$$

Proof. *By Formula 56.4.2, the complex refractive index can be written as*

$$\tilde{n} = n + ik\,,$$

where k is called the ***extinction coefficient****. From classical optics it is also known that in a material the speed of light obeys the following relation:*

$$c = \tilde{n}v\,.$$

It readily follows that the wave number (sadly also denoted by the letter k) can be written as

$$k = \frac{\omega}{v} = \tilde{n}\frac{\omega}{c}\,.$$

In electromagnetis, a plane wave can be written as

$$E(x,t) = \mathrm{Re}\left\{A\exp\left[i(kx - \omega t + \phi)\right]\right\}.$$

So, after putting everything together, one obtains

$$E(x,t) = \mathrm{Re}\left\{A\exp\left[i\left((n+ik)\frac{\omega}{c}x - \omega t + \phi\right)\right]\right\}$$

or

$$E(x,t) = \mathrm{Re}\left\{A\exp\left(in\frac{\omega}{c}x\right)\cdot\exp\left(-k\frac{\omega}{c}x\right)\cdot\exp(-i\omega t)\cdot\exp\left(i\phi\right)\right\}.$$

It is also known that the intensity is given by the following relation:

$$I(x) = |E(x)|^2\,.$$

This implies that only the second exponential factor will remain. Dividing the result by its value at $x = 0$ gives

$$\frac{I(x)}{I(0)} = \frac{|E(x)|^2}{|E(0)|^2} = \exp\left(-\frac{2k\omega}{c}x\right) = \exp(-\alpha x)\,,$$

where α is the absorption coefficient as defined in Definition 56.4.6. □

Definition 56.4.6 (Absorption coefficient). The scale factor

$$\alpha = \frac{4\pi\nu k}{c} \tag{56.25}$$

in the Lambert–Beer law is called the absorption coefficient.

Definition 56.4.7 (Scattering). When scattering light off matter, two situations can occur:

- **Rayleigh scattering**: Photons are elastically scattered off the bound electrons. The incoming and outgoing frequencies are equal.
- **Raman scattering**: Photons are inelastically scattered.

56.4.4 Interference & Diffraction

Formula 56.4.8 (Young experiment). Consider a standard double-slit experiment with a monochromatic beam. If the slit size $a \in \mathbb{R}$ is of the order of the wavelength $\lambda \in \mathbb{R}$, the slits will act as secondary (coherent) sources due to the Huygens–Fresnel principle.

Consider Fig. 56.1. Constructive interference will occur when

$$a \sin(\alpha) = n\lambda \,, \tag{56.26}$$

where $n \in \mathbb{Z}$. On the other hand, destructive interference occurs when

$$a \sin(\alpha) = (2n+1)\frac{\lambda}{2} \,. \tag{56.27}$$

Figure 56.1: Young's two-slit experiment.

Formula 56.4.9 (One-slit diffraction). In the case of one slit, one can consider equally spaced secondary sources spanning the slit (by the Huygens–Fresnel principle). In the **Fraunhofer regime**, where

$$L \gg \frac{a^2}{\lambda} \,, \tag{56.28}$$

one can then take two sources and consider the formulas from the two-split experiment (e.g. the upper source and the middle source). This leads to he following condition for

minima:

$$a \sin(\alpha) = n\lambda \,, \tag{56.29}$$

where $n \in \mathbb{Z}_0$ and $a \in \mathbb{R}^+$ is now the width of the slit. Note that this is the same formula as for constructive interference in the two-split experiment!

Part IX

Quantum Theory

Table of Contents

57 Quantum Mechanics **1055**

57.1 Introduction 1056

57.1.1 Dirac–von Neumann postulates 1056

57.2 Schrödinger picture 1058

57.3 Heisenberg–Born–Jordan picture 1060

57.3.1 Hydrogen atom 1061

57.3.2 Molecular dynamics 1062

57.4 Mathematical formalism 1062

57.4.1 Weyl systems 1062

57.4.2 Dirac–von Neumann postulates: revisited 1064

57.4.3 Symmetries 1064

57.4.4 Symmetric states 1065

57.5 Foundations ♣ 1066

57.5.1 Measurement problem 1066

57.5.2 Copenhagen interpretation 1066

57.5.3 Many-worlds interpretation 1066

57.5.4 Relational quantum mechanics 1067

57.6 Angular Momentum 1070

57.6.1 Angular momentum operator 1070

57.6.2 Rotations 1071

57.6.3 Spinor representation 1071

57.6.4 Coupling of angular momenta 1072

57.7 Electromagnetism 1073

57.7.1 Pauli theory 1073

57.7.2 Aharonov–Bohm effect 1073

57.8 Approximation methods 1074

57.8.1 WKB approximation 1074

57.9 Curved backgrounds ♣ 1076

57.9.1 Extending quantum mechanics 1076

57.9.2 WKB approximation 1077

58 Quantum Information Theory **1079**
58.1 Entanglement 1079
58.1.1 Introduction 1079
58.1.2 Bell states 1080
58.1.3 SRE states 1081
58.2 Density operators 1081
58.3 Channels 1082
58.4 Quantum logic 1083
58.4.1 Birkhoff–von Neumann logic 1083
58.4.2 Linear logic 1084
58.5 Topos theory ♣ 1090
58.5.1 Bohr topos 1090
58.5.2 Internal logic 1091

59 Quantization ♣ **1093**
59.1 Introduction 1093
59.2 Deformation quantization 1094
59.3 Geometric Quantization 1095
59.3.1 Prequantization 1095
59.3.2 Coadjoint orbits 1100
59.4 Toeplitz quantization 1100
59.5 Constrained systems 1102
59.5.1 Dirac procedure 1102
59.5.2 BRST quantization 1103
59.5.3 BFV quantization 1103

60 Perturbation Theory **1105**
60.1 Interaction picture 1105
60.2 Rayleigh–Schrödinger theory 1106
60.3 Adiabatic approximation 1107
60.3.1 Berry phase 1107

61 Scattering Theory **1109**
61.1 Cross sections 1109
61.2 Lippman–Schwinger equations 1109

Chapter 57

Quantum Mechanics

The main reference for this chapter is Bransden and Joachain (2000). In the first two sections, the two basic formalisms of quantum mechanics are introduced: wave and matrix mechanics. The main reference for the mathematically rigorous treatment of quantum mechanics, in particular in the infinite-dimensional setting, is Moretti (2016). The main reference for the generalization to curved backgrounds is Schuller (2016). The section on the WKB approximation is based on Bates and Weinstein (1997). Relevant chapters in this compendium are, amongst others, 19, 25 and 26.

Chapter Contents

57.1 Introduction **1056**

57.1.1 Dirac–von Neumann postulates 1056

57.2 Schrödinger picture **1058**

57.3 Heisenberg–Born–Jordan picture **1060**

57.3.1 Hydrogen atom 1061

57.3.2 Molecular dynamics 1062

57.4 Mathematical formalism **1062**

57.4.1 Weyl systems 1062

57.4.2 Dirac–von Neumann postulates: revisited 1064

57.4.3 Symmetries 1064

57.4.4 Symmetric states 1065

57.5 Foundations ♣ **1066**

57.5.1 Measurement problem 1066

57.5.2 Copenhagen interpretation 1066

57.5.3 Many-worlds interpretation 1066

57.5.4 Relational quantum mechanics 1067

57.6 Angular Momentum **1070**
57.6.1 Angular momentum operator 1070
57.6.2 Rotations 1071
57.6.3 Spinor representation 1071
57.6.4 Coupling of angular momenta 1072
57.7 Electromagnetism **1073**
57.7.1 Pauli theory 1073
57.7.2 Aharonov–Bohm effect 1073
57.8 Approximation methods **1074**
57.8.1 WKB approximation 1074
57.9 Curved backgrounds ♣ **1076**
57.9.1 Extending quantum mechanics 1076
57.9.2 WKB approximation 1077

57.1 Introduction

This section will give both an introduction and formal treatment of the objects and notions used in quantum mechanics.

57.1.1 Dirac–von Neumann postulates

Axiom 57.1 (States). The states of a (closed) system are represented by vectors in a (complex) Hilbert space $\mathcal{H}$. In the infinite-dimensional setting, one often further restricts to separable spaces, i.e. the spaces are required to admit a countable Hilbert basis.

Notation 57.1.1 (Dirac notation). State vectors $|\psi\rangle$ are called **ket**'s and their duals $\langle\psi|$ are called **bra**'s. The inner product of a state $|\phi\rangle$ and a state $|\psi\rangle$ is denoted by $\langle\phi\,|\,\psi\rangle$. This notation is often called the **braket notation** (or Dirac notation).

Axiom 57.2 (Observables). Every physical property is represented by a bounded, self-adjoint operator. In the finite-dimensional case, this is equivalent to an operator that admits a complete set of eigenfunctions.

Definition 57.1.2 (Compatible observables). Two observables are said to be compatible if they share a complete set of eigenvectors.

Formula 57.1.3 (Closure relation). For a complete set of eigenvectors, the closure relation (also called the **resolution of the identity**) is given by (see also Corollary 42.1.3)

$$\sum_n |\psi_n\rangle\langle\psi_n| + \int_X |x\rangle\langle x|\, dx = \mathbb{1}\,, \tag{57.1}$$

where the sum ranges over the discrete spectrum and the integral over the continuous spectrum. For simplicity, the summation will also be used for the continuous part.

Axiom 57.3 (Born rule). Let $\mathcal{H}$ be the Hilbert space of a physical system and consider an observable $\widehat{O}$. If $|\psi\rangle$ is a state vector and $\widehat{P}_\phi$ is the projection onto an eigenvector $|\phi\rangle$ of $\widehat{O}$, the probability of observing the state $|\phi\rangle$ is given by:

$$\frac{\langle\psi\,|\,\widehat{P}_\phi\,|\,\psi\rangle}{\langle\psi\,|\,\psi\rangle} = \frac{|\langle\psi\,|\,\phi\rangle|^2}{\langle\psi\,|\,\psi\rangle}\,. \tag{57.2}$$

Property 57.1.4 (Projectivization). In light of the Born rule, the dynamics of a system does not depend on the global phase or normalization, i.e. states are represented by rays in a projective Hilbert space $\mathcal{H}\mathbb{P}$ (Definition 10.1.29).

Combining Born's rule with Definition 45.3.17, gives the following definition.

Definition 57.1.5 (Expectation value). The expectation value of an observable $\widehat{A}$ in a (normalized) state $|\psi\rangle$ is defined as follows:

$$\langle\widehat{A}\rangle_\psi := \langle\psi|\widehat{A}|\psi\rangle\,. \tag{57.3}$$

The subscript ψ is often left implicit. As in ordinary statistics (Formula 48.1.17), the uncertainty or variance is defined as follows:

$$\Delta A := \langle\widehat{A}^2\rangle - \langle\widehat{A}\rangle^2\,. \tag{57.4}$$

Formula 57.1.6 (Uncertainty relation). Let $\widehat{A}, \widehat{B}$ be two observables and let $\Delta A, \Delta B$ be the corresponding uncertainties. The (**Robertson**) uncertainty relation reads as follows:

$$\Delta A \Delta B \geq \frac{1}{4}\left|\left\langle\left[\widehat{A}, \widehat{B}\right]\right\rangle\right|^2\,. \tag{57.5}$$

Axiom 57.4 (Projection[1]). Let $\mathcal{H}$ be the Hilbert space of a physical system and consider an observable $\widehat{O}$ with eigenvalues $\{o_i\}_{i\in I}$. After measuring the observable $\widehat{O}$ in the state $|\psi\rangle$, the outcome will be one of the eigenvalues o_i and system will 'collapse' to, i.e. get projected onto, the eigenstate $\widehat{P}_{o_i}|\psi\rangle \equiv |o_i\rangle$.

Axiom 57.5 (Unitary evolution). The evolution of a closed system is unitary, i.e. there exists a unitary operator $\widehat{U}(t,t') \in \mathrm{Aut}(\mathcal{H})$, for all times $t \leq t'$, such that[2]

$$|\psi(t)\rangle = \widehat{U}(t,t')|\psi(t')\rangle\,. \tag{57.6}$$

[1] Also called the **measurement postulate**.

[2] Note that some authors use a different convention whereby the two arguments are interchanged.

57.2 Schrödinger picture

Since the energy is of paramount importance in physics, the associated eigenvalue equation deserves its own name.

Formula 57.2.1 (Time-independent Schrödinger equation).

$$\widehat{H}|\psi\rangle = E|\psi\rangle \tag{57.7}$$

The operator $\widehat{H}$ is called the **Hamiltonian** of the system. The wave function ψ is an element of the vector space $L^2(\mathbb{R},\mathbb{C})\otimes\mathcal{H}$ with $\mathcal{H}$ the internal Hilbert space (describing, for example, the spin or charge of a particle). This is an eigenvalue equation for the energy levels of the system.

@@ INTRODUCE POSITION/CONFIGURATION REPRESENTATION @@

The time evolution of a wave function was governed by Axiom 57.5. By passing to generators, the following equation is obtained.

Formula 57.2.2 (Time-dependent Schrödinger equation).

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = \widehat{H}|\psi(t)\rangle\,. \tag{57.8}$$

In case $\widehat{H}$ is time independent, the TISE can be obtained from this equation by separation of variables (see below).

Proof (Derivation of TISE from TDSE). Starting from the one-dimensional TDSE in position space with a time-independent Hamiltonian, one can perform a separation of variables and assert a solution of the form $\psi(x,t) = X(x)T(t)$. Inserting this in the previous equation gives

$$i\hbar X(x)T'(t) = \big(\widehat{H}X(x)\big)T(t)\,.$$

Dividing both sides by $X(x)T(t)$ and rearranging the terms gives

$$i\hbar\frac{T'(t)}{T(t)} = \frac{\widehat{H}X(x)}{X(x)}\,.$$

Because the left side only depends on t and the right side only depends on x, one can conclude that they both have to equal a constant $E \in \mathbb{C}$. This leads to the following system of differential equations:

$$\begin{cases} i\hbar T'(t) = ET(t)\,, \\ \widehat{H}X(x) = EX(x)\,. \end{cases}$$

The first equation immediately gives a solution for T:

$$T(t) = C \exp\left(-\frac{iE}{\hbar}t\right). \tag{57.9}$$

The second equation is exactly the TISE (Formula 57.2.1). □

Example 57.2.3 (Massive particle in a stationary potential).

$$i\hbar\frac{\partial}{\partial t}\psi(x,t) = \left(-\frac{\hbar^2}{2m}\partial_x^{\,2} + V(x)\right)\psi(x,t) \tag{57.10}$$

In this case, the TISE reads as follows:

$$\psi''(x) = -\frac{2m}{\hbar^2}\big(E - V(x)\big)\psi(x). \tag{57.11}$$

Formula 57.2.4 (General solution). A general solution of the TDSE (for time-independent Hamiltonians) is given by the following formula (cf. Formula 20.3.2):

$$\psi(x,t) = \sum_E c_E \psi_E(x) e^{-\frac{i}{\hbar}Et}, \tag{57.12}$$

where the functions $\psi_E(x)$ are the eigenfunctions of the TISE. The coefficients c_E can be found using the orthogonality relations

$$c_E = \left(\int_{\mathbb{R}} \overline{\psi_E}(x)\psi(x,t_0)\,dx\right) e^{\frac{i}{\hbar}Et_0}. \tag{57.13}$$

Formula 57.2.5 (Dyson series). Inserting the evolution operator from Axiom 57.5 into the TDSE 57.2.2, even generalising to time-dependent Hamiltonians, the following operator equation is obtained:

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t}\widehat{U}(t,t') = \widehat{H}(t)\widehat{U}(t,t'). \tag{57.14}$$

With the initial condition $\widehat{U}(t,t) = 1$ for all $t \in \mathbb{R}$, this can be formally solved as follows (for $t \geq t'$):

$$\widehat{U}(t,t') = \mathbb{1} - \frac{i}{\hbar}\int_{t'}^{t} \widehat{H}(\tau)\widehat{U}(\tau,t')\,d\tau. \tag{57.15}$$

This solution can be iterated to obtain a series expansion of the evolution operator:

$$\widehat{U}(t,t') = 1 - \frac{i}{\hbar}\int_{t'}^{t}\widehat{H}(\tau)\,d\tau + \left(-\frac{i}{\hbar}\right)^2\int_{t'}^{t}\int_{t'}^{\tau}\widehat{H}(\tau)\widehat{H}(\tau')\,d\tau'\,d\tau + \cdots. \tag{57.16}$$

It is clear that the integrands are time-ordered. By explicitly introducing the **time-ordering operator**

$$\mathcal{T}\left(\widehat{H}(t_1)\widehat{H}(t_2)\right) = \begin{cases} \widehat{H}(t_1)\widehat{H}(t_2) & \text{if } t_1 \geq t_2, \\ \widehat{H}(t_2)\widehat{H}(t_1) & \text{if } t_2 > t_1, \end{cases} \tag{57.17}$$

the integrals can be rewritten in a more symmetric form:

$$\widehat{U}(t,t') = 1 - \frac{i}{\hbar}\int_{t'}^{t}\widehat{H}(\tau)\,d\tau + \frac{1}{2!}\left(-\frac{i}{\hbar}\right)\int_{t'}^{t}\int_{t'}^{t}\mathcal{T}\left(\widehat{H}(\tau)\widehat{H}(\tau')\right)d\tau'\,d\tau + \cdots\,. \tag{57.18}$$

By comparing this expression to the series expansion for exponential functions, the following concise formula is obtained:

$$\widehat{U}(t,t') = \mathcal{T}\,\exp\left(-\frac{i}{\hbar}\int_{t'}^{t}\widehat{H}(\tau)\,d\tau\right). \tag{57.19}$$

The expression on the right-hand side is called the **Dyson series**.

57.3 Heisenberg–Born–Jordan picture

In the previous section, the central object was the wave function. It was this object that evolved in time and the operators acting on the Hilbert space of physical states were assumed to be fixed. However, it is also possible to transfer this dependence on time to the operators.

Formula 57.3.1 (Time-dependent observables).

$$\widehat{O}_H(t) := e^{\frac{i}{\hbar}\widehat{H}t}\widehat{O}_S(t)e^{-\frac{i}{\hbar}\widehat{H}t} \tag{57.20}$$

The equivalence between the Schödinger and Heisenberg pictures essentially come from the fact that the time-evolving expectation values of operators are given by the following formula:

$$\langle\widehat{O}(t)\rangle = \langle\psi|e^{\frac{i}{\hbar}\widehat{H}t}\widehat{O}(t)e^{-\frac{i}{\hbar}\widehat{H}t}|\psi\rangle\,. \tag{57.21}$$

The difference between the pictures is simply the choice of whether to include the evolution operator in the states or in the operators.

Using the above transformation, the Schödinger equation (Formula 57.2.2) can also be reexpressed.

Formula 57.3.2 (Time-dependent Schrödinger equation).

$$\frac{\partial\widehat{O}_H}{\partial t}(t) = \frac{i}{\hbar}\left[\widehat{H}_H(t),\widehat{O}_H(t)\right] + \left(\frac{\partial\widehat{O}}{\partial t}(t)\right)_H \tag{57.22}$$

Taking this expression for the Schrödinger equation and taking expectation values (using the linearity of the equation), gives the following (interaction-independent) result.

Theorem 57.3.3 (Ehrenfest). *Let $\widehat{H}$ be the Hamiltonian and consider an observable $\widehat{O}$. The expectation value of this operator evolves as follows:*

$$\frac{\mathrm{d}\langle\widehat{O}\rangle}{\mathrm{d}t} = \frac{1}{i\hbar}\langle[\widehat{O},\widehat{H}]\rangle + \left\langle\frac{\partial\widehat{O}}{\partial t}\right\rangle. \tag{57.23}$$

Remark 57.3.4 (Equivalence). It is important to note that the Schrödinger equation could be replaced by Ehrenfest's theorem. They are entirely equivalent.

But, given the abstract state vectors $|\psi\rangle$ from Section 57.1.1, how does one recover the position (configuration) representation $\psi(x)$? This is simply the projection of the state vector $|\psi\rangle$ on the 'basis function' $\delta(x)$, i.e. $\psi(x)$ represents an expansion coefficient in terms of a 'basis' for the physical Hilbert space. In the same way, one can obtain the momentum representation $\psi(p)$ by projecting onto the plane waves e^{ipx}.

Remark 57.3.5. It should be noted that neither the 'basis states' $\delta(x)$, nor the plane waves e^{ipx} are square integrable and, hence, they are not elements of the Hilbert space $L^2(\mathbb{R},\mathbb{C})$. This issue can be resolved through the concept of *rigged Hilbert spaces*.

@@ COMPLETE @@

57.3.1 Hydrogen atom

Consider the hydrogen atom, i.e. a single proton (the nucleus) orbited by a single electron with only the electrostatic Coulomb force acting between them (gravity can safely be neglected):

$$\widehat{H} := \frac{\widehat{p}_p^2}{2m_p} + \frac{\widehat{p}_e^2}{2m_e} - \frac{e^2}{4\pi\varepsilon r^2}. \tag{57.24}$$

It is not hard to see that this is the quantum mechanical version of the Kepler problem (Section 52.2.3). The special property of the Kepler problem was that it contained a 'hidden' symmetry that gave rise to the conserved Laplace–Runge–Lenz vector (Definition 52.2.16). As is the case for all conserved charges in quantum mechanics, this symmetry induces a degeneracy of the energy eigenvalues. Degeneracy of the magnetic quantum number $m \in \mathbb{N}$ follows from rotational symmetry, but the energy levels of the hydrogen atom only depend on the principal quantum number $n \in \mathbb{N}$. It is the degeneracy of the total angular quantum number $l \in \mathbb{N}$ that is due to this 'hidden' SO(4)-symmetry. It is often called an 'accidental degeneracy' for this reason.

@@ COMPLETE @@

57.3.2 Molecular dynamics

Consider the Hamiltonian of two interacting atoms:

$$\widehat{H} = \frac{\widehat{P}_1^2}{2M_1} + \frac{\widehat{P}_2^2}{2M_2} + \frac{\hat{q}_1\hat{q}_2}{4\pi\varepsilon R^2} + \sum_i \frac{\hat{p}_i^2}{2m} - \frac{e\hat{q}_1}{4\pi\varepsilon r_{i1}^2} - \frac{e\hat{q}_2}{4\pi\varepsilon r_{i2}^2} + \sum_{i\neq j} \frac{e^2}{4\pi\varepsilon r_{ij}^2}\,, \tag{57.25}$$

where the indices i, j indicate the electrons and uppercase symbols denote operators associated to the nuclei.

Except for the most simple situations, solving the Schrödinger equation for this Hamiltonian becomes intractable (both analytically and numerically). However, in general, one can approximate the situation. The masses of nuclei are much larger than those of the electrons and this influences their motion, they move much slower than the electrons. In essence, the nuclei and electrons live on different time scales and this allows to decouple their dynamics:

$$\widehat{H}_{\text{nucl}} = \frac{\widehat{P}_1^2}{2M_1} + \frac{\widehat{P}_2^2}{2M_2} + \frac{Q_1Q_2}{4\pi\varepsilon R^2} + V_{\text{eff}}(R_1, R_2). \tag{57.26}$$

The electrons generate an effective potential for the nuclei and the Schrödinger equation decouples as follows:

$$\begin{aligned} \widehat{H}_{\text{nucl}}(R)\psi(R) &= E\psi(R)\,, \\ \widehat{H}_{\text{el}}(r,R)\phi(r,R) &= E_{\text{el}}\phi(r,R)\,. \end{aligned} \tag{57.27}$$

This is the so-called **Born–Oppenheimer approximation**. From a more modern physical perspective, this approximation can also be seen to be a specific instance of renormalization theory, where the short time-scale (or, equivalently, the high energy-scale) degrees of freedom are integrated out of the theory.

57.4 Mathematical formalism

57.4.1 Weyl systems

Definition 57.4.1 (Canonical commutation relations). Two observables $\widehat{A}, \widehat{B}$ are said to obey a canonical commutation relation (CCR) if they satisfy (up to a constant factor $\hbar$)

$$[\widehat{A}, \widehat{B}] = i\,. \tag{57.28}$$

The prime examples are the position and momentum operators $\hat{x}, \hat{p}$. Through functional calculus, one can also define the exponential operators $e^{is\widehat{A}}$ and $e^{it\widehat{B}}$. The above relation then induces the so-called **Weyl form** of the CCR:

$$e^{is\widehat{A}}e^{it\widehat{B}} = e^{ist}e^{it\widehat{B}}e^{is\widehat{A}}\,. \tag{57.29}$$

Theorem 57.4.2 (Stone–von Neumann). *All pairs of irreducible, unitary one-parameter subgroups satisfying the Weyl form of the CCRs are unitarily equivalent.*

Corollary 57.4.3. The Schrödinger and Heisenberg pictures are unitarily equivalent.

In fact, one can generalize the Weyl form of the CCRs.

Definition 57.4.4 (Weyl system). Let (L, ω) be a symplectic vector space and let K be a complex vector space. Consider a map W from L to the space of unitary operators on K. The pair (K, W) is called a Weyl system over (L, ω) if it satisfies

$$W(z)W(z') = e^{i/2\omega(z,z')}W(z+z') \tag{57.30}$$

for all $z, z' \in L$, i.e. W is a projective representation of the Abelian group L and ω is, up to rescaling, the group cocycle inducing it (Section 24.4.1). The relation itself is called a **Weyl relation**.

Definition 57.4.5 (Heisenberg system). Let W be a Weyl system. The selfadjoint generators $\phi(z)$, which exist by Stone's theorem 26.1.26, of the maps $t \mapsto W(tz)$ are said to form a Heisenberg system. These operators satisfy the following properties:

1. **Positive homogeneity**: $\lambda\phi(z) = \phi(\lambda z)$ for all $\lambda > 0$,
2. **Commutator**: $[\phi(z), \phi(z')] = -i\omega(z, z')$, and
3. **Weak additivity**: $\phi(z + z')$ is the closure (Definition 26.1.18) of $\phi(z) + \phi(z')$.

Remark 57.4.6. It should be noted that the Weyl relations are more fundamental than their infinitesimal counterparts. Only the Weyl relations are well defined on more general spaces and when passing to a relativistic setting.

Recall Section 42.1, where the framework of measure theory and distributions was generalized to the noncommutative context.

Property 57.4.7 (Schrödinger representation). Consider a distribution d on a (real) TVS V. There exists a unique unitary representation U of the additive group V^* on $L^2(V, d)$ such that

$$U(\lambda)f = e^{id(\lambda)}f \tag{57.31}$$

for all bounded tame functions f and such that 1 is cyclic for U in $L^2(V, d)$. Moreover, this representation is continuous with respect to the finest locally convex topology on V (the one generated by all seminorms on V)[3].

@@ EXPLAIN RELEVANCE e.g. Baez, Segal, and Zhou (2014) @@

[3]This topology is also known as the **algebraic topology**

57.4.2 Dirac–von Neumann postulates: revisited

Section 57.1.1 presented the axioms of quantum mechanics in terms of Hilbert spaces and the operators thereon. However, the incredible insight of *von Neumann* was that one can do away with the Hilbert space. By Example 26.2.8, the observables of a quantum-mechanical system form a C^*-algebra. Consequently, the idea is to rephrase the axioms in purely C^*-algebraic terms (Section 26.2). By Theorem 26.2.42, these two approaches are equivalent.

Axiom 57.6 (Observables). A physical system is characterized by a C^*-algebra, with the observables corresponding to the self-adjoint elements.

Axiom 57.7 (States). A state of a quantum-mechanical system is given by a state of the associated C^*-algebra (Definition 26.2.16).

Axiom 57.8 (Born rule). The expectation value of an observable a in a state ω is given by the evaluation $\omega(a)$.

Remark 57.4.8. Section 58.2 will link this axiom to traces and operator theory through Property 26.2.27 and Definition 26.2.29.

Axiom 57.9 (Projection).

Axiom 57.10 (Unitary evolution).

@@ CORRECT ALL AXIOMS @@

57.4.3 Symmetries

Property 57.4.9 (States). By the postulates of quantum mechanics, states are represented by rays in the projective Hilbert space $\mathcal{H}\mathbb{P}$. The probabilities, given by the Born rule (Axiom 57.3), can be expressed in terms of the *Fubini–Study metric* on $\mathcal{H}\mathbb{P}$ as follows:

$$\mathcal{P}(\psi, \phi) := \cos^2\big(d_{\mathrm{FS}}(\psi, \phi)\big) = \frac{|\langle \psi \,|\, \phi \rangle|^2}{\langle \psi \,|\, \psi \rangle \langle \phi \,|\, \phi \rangle}, \tag{57.32}$$

where $|\psi\rangle, |\phi\rangle$ are representatives of the states ψ, ϕ in $\mathcal{H}\mathbb{P}$.

Definition 57.4.10 (Symmetry). A quantum symmetry (or **quantum automorphism**) is an isometric automorphism of $\mathcal{H}\mathbb{P}$. The group of these symmetries is denoted by $\mathsf{Aut}_{\mathrm{QM}}(\mathcal{H}\mathbb{P})$.

The following theorem due to *Wigner* gives a (linear) characterization of quantum symmetries.[4]

Theorem 57.4.11 (Wigner). *Every quantum automorphism of $\mathcal{H}\mathbb{P}$ is induced by a unitary or anti-unitary operator on $\mathcal{H}$.*

[4] It is a particular case of a more general theorem in projective geometry.

This is equivalent to saying that the group morphism

$$\pi : \mathsf{Aut}(\mathcal{H}, \mathcal{P}) := \mathsf{U}(\mathcal{H}) \times \mathsf{AU}(\mathcal{H}) \to \mathsf{Aut}_{\mathrm{QM}}(\mathcal{H}\mathbb{P}) \tag{57.33}$$

is surjective. Together with the kernel U(1), given by phase shifts, this forms a short exact sequence:

$$1 \longrightarrow \mathsf{U}(1) \longrightarrow \mathsf{Aut}(\mathcal{H}, \mathcal{P}) \longrightarrow \mathsf{Aut}_{\mathrm{QM}}(\mathcal{H}\mathbb{P}) \longrightarrow 1\,. \tag{57.34}$$

In the case of symmetry breaking (e.g. lattice systems), the full symmetry group is reduced to a subgroup $G \subset \mathsf{Aut}_{\mathrm{QM}}(\mathcal{H}\mathbb{P})$. The group of operators acting on $\mathcal{H}$ is then given by the pullback $\widetilde{G}$ of the diagram

$$\mathsf{Aut}(\mathcal{H}, \mathcal{P}) \longrightarrow \mathsf{Aut}_{\mathrm{QM}}(\mathcal{H}\mathbb{P}) \longleftarrow G\,. \tag{57.35}$$

It should also be noted that the kernel of the homomorphism $\widetilde{G} \to G$ is again U(1). This leads to the property that $\widetilde{G}$ is a $\mathbb{Z}_2$-twisted (hence noncentral) U(1)-extension of G, where the twist is induced by the homomorphism $\phi : \mathsf{Aut}(\mathcal{H}, \mathcal{P}) \to \mathbb{Z}_2$ that says whether an operator is implemented unitarily or anti-unitarily.

@@ COMPLETE @@

57.4.4 Symmetric states

Axiom 57.11 (Symmetrization postulate). Let $\mathcal{H}$ be the single-particle Hilbert space. A system of $n \in \mathbb{N}$ identical particles is described by a state $|\Psi\rangle$ belonging to either $S^n\mathcal{H}$ or $\Lambda^n\mathcal{H}$. These **bosonic** and **fermionic** states are, respectively, of the form

$$|\Psi_B\rangle = \sum_{\sigma \in S_n} |\psi_{\sigma(1)}\rangle \cdots |\psi_{\sigma(n)}\rangle \tag{57.36}$$

and

$$|\Psi_F\rangle = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) |\psi_{\sigma(1)}\rangle \cdots |\psi_{\sigma(n)}\rangle\,, \tag{57.37}$$

where the $|\psi_i\rangle$ are single-particle states and S_n is the permutation group on n elements.

Remark 57.4.12. In ordinary quantum mechanics, this is a postulate, but in quantum field theory, this is a consequence of the *spin-statistics theorem*. @@ ADD THIS THEOREM TO [QFT] @@

Definition 57.4.13 (Slater determinant). Let $\{\phi_i(\vec{\mathbf{q}})\}_{i \leq n}$ be a set of wave functions, called **spin orbitals**, describing a system of n identical fermions. The totally antisymmetric wave function of the system is given by

$$\psi(\vec{\mathbf{q}}_1, \dots, \vec{\mathbf{q}}_n) = \frac{1}{\sqrt{n!}} \det \begin{pmatrix} \phi_1(\vec{\mathbf{q}}_1) & \cdots & \phi_n(\vec{\mathbf{q}}_1) \\ \vdots & \ddots & \vdots \\ \phi_1(\vec{\mathbf{q}}_n) & \cdots & \phi_n(\vec{\mathbf{q}}_n) \end{pmatrix}. \tag{57.38}$$

A similar function can be defined for bosonic systems using the concept of *permanents*.

57.5 Foundations ♣

57.5.1 Measurement problem

If one looks at the Schrödinger equation (Formula 57.2.2) or Ehrenfest's theorem (Theorem 57.3.3), it is easy to see that time evolution is entirely linear and deterministic. Superpositions are preserved under Hamiltonian flow (a crucial ingredient of quantum mechanics) and, given an initial state, time evolution will always lead to the same final state. However, the Born rule (Axiom 57.3), which governs 'measurements' is very nonlinear and nondeterministic. It is probabilistic and, once a 'measurement' has been performed, the state has 'collapsed' onto an eigenstate of the observable under consideration.

The issue of what constitutes a 'measurement' — Is it a conscious human doing an experiment? Is it a mouse interfering with an experiment? Is it two particles interacting? ...[5] — and why exactly the Born rule holds and what it entails, i.e. how probabilities arise, is known as the measurement problem. On a historical note, it should be noted that, after an initial surge of interest shortly after the 5th Solvay Conference (1927), where quantum mechanics was formally established, the study of the foundations of quantum mechanics (the measurement problem specifically) became an infamous topic due to the pragmatic mentality of nuclear physics during the 20th century.

@@ ADD (dynamical collapse, epistemic) @@

57.5.2 Copenhagen interpretation

The Copenhagen interpretation[6] takes the foundations of quantum mechanics as presented above very literally.

@@ COMPLETE (e.g. collapse) @@

57.5.3 Many-worlds interpretation

This interpretation, originating with *Everett*, posits a different idea, which does away with the need of the explicit Born rule axiom. In this interpretation, there is a kind of 'universal wave function', which governs both the observer and the experiment. A 'measurement' is then simply an entanglement-inducing interaction between these two subsystems.

[5] This (perhaps artificial) boundary between classical and quantum is sometimes called the **Heisenberg cut**.

[6] This name stems from the fact that its initial proponents were from the group of physicists centered around *Bohr*.

The main implication of such an interpretation is, however, that the universal wave function branches every time such an interaction occurs. More precisely, assume that 'we', the observers, perform a measurement on some system (for simplicity, assume that the measurement has a binary outcome). The measurement process is then described as follows:

$$|\text{in}\rangle_{\text{obs}}|\text{in}\rangle_{\text{exp}} \longrightarrow \lambda_0|0\rangle_{\text{obs}}|0\rangle_{\text{exp}} + \lambda_1|1\rangle_{\text{obs}}|1\rangle_{\text{exp}}\,. \tag{57.39}$$

Taking this superposition as a physical reality, this means that if we had measured the state 0, a copy of us living on the other branch will have measured 1 (and the other way around).

@@ COMPLETE (e.g. origin of probabilities) @@

57.5.4 Relational quantum mechanics

An important notion in classical physics is that of a *reference frame*, i.e. a choice of axes and scales. Usually, this corresponds to choosing an observer, relative to which one expresses the motion of all other objects. In relativity, the relative treatment of physics was the grand breakthrough by Einstein. However, although this notion had been left aside for a long time in the treatment of quantum mechanics and a specific choice of reference frame was silently assumed, this assumption was not as innocuous as it appears. Superposition and complementarity make a definite choice of absolute reference frame impossible.

To understand the relevance of a relational approach to quantum mechanics, consider the following thought experiment.

Definition 57.5.1 (Wigner's friend). Consider two observers, Wigner and his friend, performing an experiment as shown diagrammatically in Fig. 57.1. One envisions Wigner standing outside the laboratory, having no way to observe what happens inside the lab, and his friend who performs an experiment inside the lab. The paradox arises from the two ways one can describe the sequence of the friend performing a measurement and Wigner checking up on the results in the classical (Copenhagen) interpretation.

From the point of view of the friend, at the moment of measurement, the projection/collapse axiom states that the wave function describing friend + experiment 'collapses' to:

$$|\psi\rangle = |\uparrow\rangle_{\text{friend}}|\uparrow\rangle_{\text{exp}}\,. \tag{57.40}$$

However, from the point of view of Wigner, who has not observed the measurement, the state is described by

$$|\psi\rangle' = \alpha|\uparrow\rangle_{\text{friend}}|\uparrow\rangle_{\text{exp}} + \beta|\downarrow\rangle_{\text{friend}}|\downarrow\rangle_{\text{exp}}\,. \tag{57.41}$$

Figure 57.1: Wigner's friend thought experiment.

Whereas in the many-world approach one would simply take the branching approach, which is fully unitary and resolves this issue by avoiding collapse, the relational approach takes collapse at face value, but states that observations are relative, i.e. always with respect to some fixed observer (be it a person, a classical object or another quantum-mechanical system). From this point of view, textbook Copenhagen QM is simply quantum mechanics with respect to some god-given observer, and collapse and unitary evolution do not have to be reconciled.

The more general idea is that information and, hence, the values of observables are a relative notion, i.e. variables only attain their values when considered with respect to a certain observer. As such, RQM is an epistemic interpretation of quantum mechanics in that the wave function only captures 'our' information about the system (or universe) and not the 'true' physical state. When applied to the notion of instantaneity (or velocity), this line of thinking will give rise to (special) relativity as in Chapter 66 (and Chapter 67).

@@ COMPLETE @@

For example, consider three observers: Alice, Bob and Charlie. Assume that each observer has a spin-$\frac{1}{2}$ particle and that, relative to Alice, the joint state is given by

$$|\psi\rangle^A_{ABC} = |\uparrow\rangle^A_A \Big(|\uparrow\rangle^A_B + |\downarrow\rangle^A_B\Big)|\downarrow\rangle^A_C \,. \tag{57.42}$$

Note that this state is separable. Now, what would the state be relative to Bob? If one supposes that changes of reference frame are *coherent* (to be formalized below), the joint state will be

$$|\psi\rangle^B_{ABC} = |\uparrow\rangle^B_B \big(|\uparrow\rangle^B_A|\downarrow\rangle^B_C + |\downarrow\rangle^B_A|\uparrow\rangle^B_C\big) \,. \tag{57.43}$$

A mere change of reference frame, an operation that would classically leave the physics invariant, has transformed a product state into an entangled state.

Axiom 57.12 (Relational physics). Given $n \in \mathbb{N}$ systems[7], any state is described relative to one of these systems. Given a choice of 'observing system', let it be system i, the state of system i is given by a fiducial state $|0\rangle_i^i$.

Axiom 57.13 (Coherent change). Consider a change of reference frame $0 \longrightarrow i$ such that

$$\begin{cases} |\psi\rangle^0 \longrightarrow |\psi\rangle^i \\ |\phi\rangle^0 \longrightarrow |\phi\rangle^i . \end{cases} \tag{57.44}$$

Then

$$\alpha|\psi\rangle^0 + \beta|\phi\rangle^0 \longrightarrow \alpha|\psi\rangle^i + \beta|\phi\rangle^i \tag{57.45}$$

for all $\alpha, \beta \in \mathbb{C}$.

Abstractly, a (classical) reference frame is defined as follows in the spirit of Section 44.4 and Section 44.9.

Definition 57.5.2 (Reference frame). Let X be an object of interest. Whereas a coordinate chart on X, modeled on an object Y, is given by a morphism $Y \to X$, a **coordinate system** on X is given by an isomorphism $Y \cong X$, i.e. a global coordinate chart. A reference frame is coordinate system for which Y corresponds the a physical system.

Let the system of interest X admit a group action that is both free and transitive, turning it into a G-torsor (Definition 2.3.16). At the level of sets, one has $X \cong G$ and a choice of origin, i.e. a specific choice of isomorphism, corresponds to a choice of reference frame (the identity element corresponding to the fiducial state above). A change of reference frames $s^0 \longrightarrow s^i$, frome system 0 to system i, is given by the right regular action of the relative coordinate of i on all relative coordinates:

$$\phi^{0\to i}(e, g_1^0, \dots, g_n^0) \mapsto (g_0^i, g_1^0 g_0^i, \dots, e, \dots, g_n^0 g_0^i) , \tag{57.46}$$

where the relation $g_i^0 = (g_0^i)^{-1}$ was used. It should be noted that this boils down to a *passive transformation*. When passing to the quantization of these systems, one should assume that G is locally compact and comes equipped with the canonical Haar measure (Theorem 19.4.3). In this case, a quantization is given by the space of square-integrable functions $L^2(G)$, where basis states are labeled by group elements.

@@ VERIFY THIS STATEMENT @@

The change-of-reference-frame operator is given as follows:

$$\widehat{U}^{0\to i} := \mathsf{SWAP}_{0,i} \circ \int_G \mathbb{1}_{L^2(G)} \otimes \widehat{U}_R(g_i^0)^{\otimes i-2} \otimes |g_0^i\rangle\langle g_i^0| \otimes \widehat{U}_R(g_i^0)^{\otimes n-i-2} \, dg_i^0 , \tag{57.47}$$

[7] An abstraction of the notion of observer.

where

$$\widehat{U}_R(g) : |x\rangle \mapsto |xg^{-1}\rangle \tag{57.48}$$

is the unitary implementation of the right regular action and dg denotes integration with respect to the Haar measure on G. It can be shown that $\widehat{U}^{0\to i}$ is unitary, its inverse being given by $\widehat{U}^{i\to 0}$ and composition is transitive. It can be shown that this procedure can be extended to any one-particle Hilbert space $\mathcal{H}$ as long as the inclusion $G \to \mathcal{H}$ is injective and maps G to an orthonormal basis of (a subset of) $\mathcal{H}$.

57.6 Angular Momentum

57.6.1 Angular momentum operator

Property 57.6.1 (Lie algebra). The angular momentum operators generate a Lie algebra (Definition 32.2.2). The Lie bracket is defined by the following commutation relation:

$$[\hat{J}_i, \hat{J}_j] = i\hbar\varepsilon_{ijk}\hat{J}_k\,. \tag{57.49}$$

Since rotations correspond to actions of the orthogonal group $\mathsf{SO}(3)$, it should not come as a surprise that the above relation is exactly the defining relation of the Lie algebra $\mathfrak{so}(3)$ from Example 32.2.21.

Property 57.6.2. The mutual eigenbasis of $\hat{J}^2$ and $\hat{J}_z$ is defined by the following two eigenvalue equations:

$$\hat{J}^2|j,m\rangle = j(j+1)\hbar^2\,|j,m\rangle\,, \tag{57.50}$$

$$\hat{J}_z|j,m\rangle = m\hbar\,|j,m\rangle\,. \tag{57.51}$$

Definition 57.6.3 (Ladder operators[8]). The raising and lowering operators $\hat{J}_+$ and $\hat{J}_-$ are defined as follows:

$$\hat{J}_+ := \hat{J}_x + i\hat{J}_y \qquad \text{and} \qquad \hat{J}_- := \hat{J}_x - i\hat{J}_y\,. \tag{57.52}$$

These operators only change the quantum number $m_z \in \mathbb{N}$, not the total angular momentum.

Corollary 57.6.4. From the commutation relations of the angular momentum operators, one can derive the commutation relations of the ladder operators:

$$[\hat{J}_+, \hat{J}_-] = 2\hbar\hat{J}_z\,. \tag{57.53}$$

[8] Also called the **creation** and **annihilation** operators (especially in quantum field theory).

Formula 57.6.5. The total angular momentum operator $\hat{J}^2$ can now be expressed in terms of $\hat{J}_z$ and the ladder operators using the commutation relation (57.49):

$$\hat{J}^2 = \hat{J}_+\hat{J}_- + \hat{J}_z^2 - \hbar\hat{J}_z\,. \tag{57.54}$$

Remark 57.6.6 (Casimir operator). From the definition of $\hat{J}^2$, it follows that this operator is a Casimir invariant (Definition 32.4.59) of $\mathfrak{so}(3)$.

57.6.2 Rotations

Formula 57.6.7. An infinitesimal rotation $\widehat{R}(\delta\vec{\boldsymbol{\varphi}})$ is given by the following formula:

$$\widehat{R}(\delta\vec{\boldsymbol{\varphi}}) = \mathbb{1} - \frac{i}{\hbar}\vec{\mathbf{J}}\cdot\delta\vec{\boldsymbol{\varphi}}\,. \tag{57.55}$$

A finite rotation can be generated by applying this infinitesimal rotation repeatedly:

$$\widehat{R}(\vec{\boldsymbol{\varphi}}) = \left(\mathbb{1} - \frac{i}{\hbar}\vec{\mathbf{J}}\cdot\frac{\vec{\boldsymbol{\varphi}}}{n}\right)^n = \exp\left(-\frac{i}{\hbar}\vec{\mathbf{J}}\cdot\vec{\boldsymbol{\varphi}}\right). \tag{57.56}$$

Formula 57.6.8 (Matrix elements). Applying a rotation over an angle φ about the z-axis to a state $|j,m\rangle$ gives

$$\widehat{R}(\varphi\vec{\mathbf{e}}_z)|j,m\rangle = \exp\left(-\frac{i}{\hbar}\hat{J}_z\varphi\right)|j,m\rangle = \exp\left(-\frac{i}{\hbar}m\varphi\right)|j,m\rangle\,. \tag{57.57}$$

Multiplying these states with a bra $\langle j',m'|$ and using the orthonormality of the eigenstates, gives the matrix elements of the rotation operator:

$$\widehat{R}_{ij}(\varphi\vec{\mathbf{e}}_z) = \exp\left(-\frac{i}{\hbar}m\varphi\right)\delta_{jj'}\delta_{mm'}\,. \tag{57.58}$$

From the expression of the angular momentum operators and the rotation operator, it is clear that a general rotation has no effect on the total angular momentum number $j \in \mathbb{N}$. This means that the rotation matrix will be block diagonal with respect to j. This amounts to the following reduction of the representation of the rotation group:

$$\langle j,m'|\widehat{R}(\varphi\vec{\mathbf{n}})|j,m\rangle = \mathcal{D}^{(j)}_{m,m'}(\widehat{R})\,, \tag{57.59}$$

where the functions $\mathcal{D}^{(j)}_{m,m'}(\widehat{R})$ are called the **Wigner D-functions**. For every value of j, there are $(2j+1)$ values for m. This implies that the matrix $\mathcal{D}^{(j)}(\widehat{R})$ is a $(2j+1)\times(2j+1)$-matrix.

57.6.3 Spinor representation

Definition 57.6.9 (Pauli matrices).

$$\sigma_x := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \sigma_y := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \sigma_z := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{57.60}$$

From this definition, it is clear that the Pauli matrices are Hermitian and unitary. Together with the 2×2 identity matrix, they form a basis for the space of 2×2 Hermitian matrices. For this reason, the identity matrix is often denoted by σ_0 (especially in the context of relativistic QM).

Formula 57.6.10. In the spinor representation ($J = \frac{1}{2}$), the Wigner-D matrix reads as follows:

$$\mathcal{D}^{(1/2)}(\varphi \vec{\mathbf{e}}_z) = \begin{pmatrix} e^{-i/2\varphi} & 0 \\ 0 & e^{i/2\varphi} \end{pmatrix}. \tag{57.61}$$

57.6.4 Coupling of angular momenta

Due to the tensor product structure of a coupled Hilbert space, the angular momentum operator $\hat{J}_i$ should now be interpreted as $\mathbb{1} \otimes \cdots \otimes \hat{J}_i \otimes \cdots \otimes \mathbb{1}$ (cf. Notation 23.3.20). Because the angular momentum operators $\hat{J}_{k \neq i}$ do not act on the space $\mathcal{H}_i$, one can pull these operators through the tensor product:

$$\hat{J}_i |j_1\rangle \otimes \cdots \otimes |j_n\rangle = |j_1\rangle \otimes \cdots \otimes \hat{J}_i |j_i\rangle \otimes \cdots \otimes |j_n\rangle . \tag{57.62}$$

The basis used above is called the **uncoupled basis**.

For simplicity, the total Hilbert space is, from here on, assumed to be that of a two-particle system. Let $\hat{J}$ denote the total angular momentum:

$$\hat{J} = \hat{J}_1 + \hat{J}_2 . \tag{57.63}$$

With this operator, one can define a **coupled** state $|J, M\rangle$, where M is the total magnetic quantum number which ranges from $-J$ to J.

Formula 57.6.11 (Clebsch–Gordan coefficients). Because both bases (coupled and uncoupled) span the total Hilbert space $\mathcal{H}$, there exists an invertible transformation between them. The transformation coefficients can be found by using the resolution of the identity:

$$|J, M\rangle = \sum_{m_1=-j_1}^{j_1} \sum_{m_2=-j_2}^{j_2} |j_1, j_2, m_1, m_2\rangle \langle j_1, j_2, m_1, m_2 | J, M\rangle . \tag{57.64}$$

These coefficients are called the Clebsch–Gordan coefficients.

Property 57.6.12. By acting with the operator $\hat{J}_z$ on both sides of Formula 57.6.11, it is possible to prove that the Clebsch–Gordan coefficients are nonzero if and only if $M = m_1 + m_2$.

57.7 Electromagnetism

57.7.1 Pauli theory

Consider a charge q moving though an electromagnetic field with potential $\mathbf{A} \equiv (\varphi, \vec{\mathbf{A}})$. In classical physics, the potential is often considered 'nonphysical', as only the fields themselves are measurable. However, when also including quantum mechanics, this picture changes (see the next subsection).

To account for the electromagnetic field, the ordinary Hamiltonian (57.10) first has to be modified. In this section this will be considered in a nonrelativistic way, i.e. the Dirac equation will not be considered.

Definition 57.7.1 (Minimal coupling). In an electromagnetic potential $(\varphi, \vec{\mathbf{A}})$, one performs the transformation

$$\hat{p} \longrightarrow \hat{p} + q\vec{\mathbf{A}}\,. \tag{57.65}$$

The reason for this transformation is that the canonical momentum $\hat{p} = -i\hbar\nabla$ takes into account the effect of the electromagnetic field, but the kinetic part of the Hamiltonian should only care about the 'kinetic momentum $m\vec{\mathbf{v}}$'. This gives the following (minimally coupled) Hamiltonian:

$$\widehat{H} = \frac{(\hat{p} - q\vec{\mathbf{A}})^2}{2m} + q\varphi\,. \tag{57.66}$$

Note that transformation can be derived from the Lorentz force by writing in the form of Eq. (52.5). Through the Lagrangian formalism, this then gives rise to a canonical momentum, which can be shown to be the minimally coupled momentum in Eq. (57.65) above.

57.7.2 Aharonov–Bohm effect

Under a gauge transformation (Section 55.4.2), the wave function changes as follows:

$$\psi' = \psi e^{-\frac{iq}{\hbar}\chi}\,. \tag{57.67}$$

Now, consider a system where $\vec{\mathbf{B}} = 0$ or, equivalently, $\vec{\mathbf{A}} = \nabla\chi$, i.e. the field is **pure gauge**. As the name implies, this configuration can be gauge transformed to get $\vec{\mathbf{A}} = 0$. This gives

$$\psi' = \psi \exp\left(\frac{iq}{\hbar}\int_{\vec{\mathbf{r}}_0}^{\vec{\mathbf{r}}} \vec{\mathbf{A}} \cdot d\vec{\mathbf{r}}\right). \tag{57.68}$$

Hence, to get the wave function for a pure gauge configuration, one simply has to solve Schrödinger equation for a free particle and multiply by a phase factor (which is determined by the potential)! In particular, if one takes a path around a closed loop, one obtains

$$\frac{q}{\hbar}\oint_\gamma \vec{\mathbf{A}} \cdot d\vec{\mathbf{r}} = 2\pi n \tag{57.69}$$

for $n \in \mathbb{Z}$, since the free particle will return to its initial state without nontrivial phase factors. Now, by the Kelvin–Stokes Theorem 23.2.5, the left-hand side is proportional to the magnetic flux through the loop:

$$\frac{q}{\hbar}\Phi = 2\pi n\,. \tag{57.70}$$

This **flux quantization** can be observed in superconducting loops, where the field outside the solenoid has nonzero potential even though the magnetic field vanishes.

Moreover, if one cuts the loop in two, representing the path of two charges going around the solenoid along different paths, one obtains a phase difference

$$\Delta\phi = \frac{q}{\hbar}\Phi\,. \tag{57.71}$$

It follows that the magnetic flux induces an interference pattern, even though the charges themselves do not move through a nonzero field!

Formula 57.7.2 (Aharonov–Casher effect). A dual (in the sense of electric-magnetic duality) also exists: the Aharonov–Casher effect.

The phase shift along a path γ is given by

$$\Delta\phi_\gamma = \frac{1}{\hbar c^2}\int_\gamma (\vec{\mathbf{E}} \times \vec{\boldsymbol{\mu}}) \cdot d\vec{\mathbf{x}}\,, \tag{57.72}$$

where $\vec{\boldsymbol{\mu}}$ is the magnetic dipole moment.

@@ CHECK if c^2 is required in SI units (vs Gaussian) @@

57.8 Approximation methods

57.8.1 WKB approximation

The Wentzel–Kramers–Brillouin (WKB) approximation[9] starts from the ansatz

$$\psi(\vec{\mathbf{q}}) := \exp\big(iS(\vec{\mathbf{q}})/\hbar\big)\,, \tag{57.73}$$

[9] This approach to solving second-order ODEs was essentially introduced a century earlier by *Green* and *Liouville*.

with $S : \mathbb{R}^n \to \mathbb{R}$ a phase function that is to be determined. Inserting this in the TDSE (in configuration representation) gives:

$$\left[\frac{\|\vec{\nabla}S(\vec{\mathbf{q}})\|^2}{2m} + \big(V(\vec{\mathbf{q}}) - E\big) - \frac{i\hbar\Delta S(\vec{\mathbf{q}})}{2m}\right] \exp\big(iS(\vec{\mathbf{q}})/\hbar\big) = 0\,. \tag{57.74}$$

To first order, i.e. for slowly varying potentials, the last term can be ignored. In this case, the phase function satisfies the Hamilton–Jacobi equation (52.54):

$$H\big(\vec{\mathbf{q}}, S'(\vec{\mathbf{q}})\big) = \frac{\|\vec{\nabla}S(\vec{\mathbf{q}})\|^2}{2m} + \big(V(x) - E\big) = 0\,. \tag{57.75}$$

In physics, the Hamilton–Jacobi equation without time derivative is often called the **eikonal equation**[10]. This leads to the following result.

Property 57.8.1. A function $S : \mathbb{R}^n \to \mathbb{R}$ is a phase function for a first-order solution to the Schrödinger equation if its differential lies in a level set of the classical Hamiltonian $H : T^*\mathbb{R}^n \to \mathbb{R}$. These solutions are said to be **admissible**.

To obtain higher-order approximations, the solution has to be generalized beyond a pure phase function:

$$\psi(\vec{\mathbf{q}}) = a(\vec{\mathbf{q}}) \exp\big(iS(\vec{\mathbf{q}})/\hbar\big)\,. \tag{57.76}$$

Assuming S is admissible, the factor $a : \mathbb{R}^n \to \mathbb{R}$ satisfies the **homogeneous transport equation**:

$$a\Delta S + 2\vec{\nabla}a \cdot \vec{\nabla}S = 0\,. \tag{57.77}$$

If a satisfies this equation, ψ is called a **semiclassical state**. Note that this equation is equivalent to $a^2\vec{\nabla}S$ being divergence free or, equivalently:

$$\mathcal{L}_{\pi_* X^H}(a^2 \operatorname{Vol}) = 0\,. \tag{57.78}$$

Since Lie derivatives pull back under diffeomorphisms (Property 34.4.19) and the image $\operatorname{im}(\mathrm{d}S)$ gives a trivial subbundle of $T^*\mathbb{R}^n$, this is also equivalent to

$$\mathcal{L}_{X^H}(a^2\pi^* \operatorname{Vol}) = 0\,. \tag{57.79}$$

This quadratic behaviour in a leads to the idea that the correct object for representing quantum states is a half-density (see also Method 59.3.7). This leads to the following statement:

> A second-order solution to the Schrödinger equation is given by a pair (S, a), where S is an admissible phase function and $a \in \Omega^{1/2}(\operatorname{im}(\mathrm{d}S))$ is a half-form that is invariant under the (classical) Hamiltonian flow.

The generalization to curved spaces, i.e. replacing $\mathbb{R}^{2n}$ by a symplectic manifold M, will be covered in Section 57.9.2.

[10] This name stems from optics.

57.9 Curved backgrounds ♣

Using the tools of distribution theory and differential geometry (Chapters 19, 33 and onwards), one can introduce quantum mechanics on curved backgrounds (in the sense of 'space', not 'spacetime').

57.9.1 Extending quantum mechanics

Remark 57.9.1 (Rigged Hilbert spaces). A first important remark to be made is that the classical definition of the wave function as an element of $L^2(\mathbb{R}^d, \mathbb{C})$ is not sufficient, even in flat Cartesian space. A complete description requires the introduction of so-called *Gel'fand triples* or *rigged Hilbert spaces*, where the space of square-integrable functions is replaced by the Schwartz space (Definition 19.1.27) of rapidly decreasing functions. The linear functionals on this space are then given by the tempered distributions.

When working on curved spaces or even in non-Cartesian coordinates on flat space, one can encounter problems with the definition of the self-adjoint operators $\hat{q}^i$ and $\hat{p}_i$. The naive definition $\hat{q}^i = q^i, \hat{p}_i = -i\partial_i$ gives rise to extra terms that break the canonical commutation relations and the selfadjointness of the operators (e.g. the angular position operator $\widehat{\varphi}$ on the circle together with its conjugate $\hat{L}$) when calculating inner products.

An elegant solution to this problem is obtained by giving up the definition of the wave function as a well-defined function $\psi : \mathbb{R}^d \to \mathbb{C}$. Assume that the physical space has the structure of a Riemannian manifold (M, g) and that the 'naive' wave functions take values in a vector space V. Then, construct a vector bundle E with typical fibre V over M. By Property 35.1.21, an invariant description of the 'true' wave function is a map $\Psi : F(E) \to V$ or, locally, the pullback $\psi := \varphi^*\Psi$ for some local section $\varphi : U \subseteq M \to F(E)$. The Levi-Civita connection on M also induces a covariant derivative ∇ on E that can be used to define differential operators.

Now, a general inner product can be introduced:

$$\langle \psi, \phi \rangle := \int_M \overline{\psi(x)}\phi(x)\,\mathsf{Vol}_M \,. \tag{57.80}$$

Because the factor $\sqrt{\det(g)}$ transforms in the inverse manner of the measure dx, the integrand is invariant under coordinate transforms (something that is generally required of physical laws). Using this new inner product, one can for example check the selfad-

jointness of the momentum operator $\widehat{P}_i := -i\nabla_i$:

$$\begin{aligned}
\langle \psi, \widehat{P}_i \phi \rangle &= \int_M \overline{\psi(x)}(-i\nabla_i)\phi(x)\sqrt{\det(g)}\,dx \\
&\overset{Eq.\ (35.77)}{=} \int_M \overline{\psi(x)}(-i\partial_i - i\omega_i)\phi(x)\sqrt{\det(g)}\,dx \\
&= \int_M \overline{(-i\partial_i\psi)(x)}\phi(x)\sqrt{\det(g)}\,dx + i\int_M \overline{\psi(x)}\phi(x)\Big(\partial_i\sqrt{\det(g)}\Big)\,dx \\
&\qquad - i\int_M \overline{\psi(x)}\omega_i\phi(x)\sqrt{\det(g)}\,dx \\
&= \langle \widehat{P}_i\psi, \phi\rangle - i\int_M \overline{\psi(x)}\overline{\omega_i}\phi(x)\sqrt{\det(g)}\,dx \\
&\qquad + i\int_M \overline{\psi(x)}\phi(x)\Big(\partial_i\sqrt{\det(g)}\Big)\,dx \\
&\qquad - i\int_M \overline{\psi(x)}\omega_i\phi(x)\sqrt{\det(g)}\,dx\,.
\end{aligned}$$

Selfadjointness then requires that

$$\sqrt{\det(g)}(\omega_i + \overline{\omega_i}) = \partial_i\sqrt{\det(g)} \tag{57.81}$$

or

$$2\mathrm{Re}(\omega_i) = \partial_i \ln\Big(\sqrt{\det(g)}\Big)\,. \tag{57.82}$$

@@ COMPLETE (rewrite in global terms) @@

57.9.2 WKB approximation

Property 57.8.1 is generalized quite trivially after replacing $\mathbb{R}^n$ by a configuration manifold Q. A further step is provided by also generalizing Property 37.3.6.

Property 57.9.2. A Lagrangian submanifold $\iota : L \hookrightarrow T^*Q$ will be called an admissible phase function for a first-order solution to the Schrödinger equation if it satisfies the classical Hamilton–Jacobi equation, i.e. lies in a level set of the classical Hamiltonian $H : T^*Q \to \mathbb{R}$, for a regular value.

To obtain a second-order solution, one also needs prefactor for the semiclassical states. The homogeneous transport equation (57.77) is generalized as follows:

$$a\Delta S + 2\mathcal{L}_{\nabla S}a = 0\,, \tag{57.83}$$

where Δ is the Laplace–Beltrami operator on Q. As before, a general second-order solution, assuming S is admissible, is given by a half-form $a \in \Omega^{1/2}(L)$ satisfying

$$\mathcal{L}_Y a = 0\,, \tag{57.84}$$

where Y is the (nonsingular) vector field on L induced by X^H. This then gives a second-order solution on Q by pulling back along the inverse $(\pi \circ \iota)^{-1}$, which is a diffeomorphism since L is projectable. Moreover, if L is exact (Definition 37.3.12), then S is induced by a primitive of the induced Liouville form $\iota^*\alpha$. If both the exactness and projectability conditions are dropped, the notion of a **geometric solution** are obtained.

To pass to this more general situation, some more structure is needed. If L is not exact, the Liouville form does not admit a global primitive. However, L does admit a (good) cover $\{U_k\}_{k\in I}$ such that on every patch, a second-order solution can be found, and then the problem becomes how to glue these together. The gluing condition is the following integrality condition:

$$\phi_k(x) - \phi_l(x) \in 2\pi\hbar\mathbb{Z}\,, \tag{57.85}$$

where ϕ_k is the phase function on U_k, for all $x \in U_k \cap U_l$. Note that this condition can only be satisfied for all $\hbar \in \mathbb{R}^+$ if $[\alpha] = 0$. However, this is exactly the condition that should be relaxed. Luckily, $\hbar$ should be a fixed value.

Definition 57.9.3 (Quantizable Lagrangian). A projectable Lagrangian submanifold $L \subset T^*M$ is said to be quantizable if there exists an $\hbar \in \mathbb{R}^+$ such that the restriction of the Liouville class to L is $\hbar$-integral, i.e. the integrality condition (57.85) is satisfied. All values $\hbar$ for which the integrality condition is satisfied, are said to be **admissible**.

Remark 57.9.4. Note that the admissible values for $\hbar$ will form a decreasing sequence of the form

$$\hbar_0, \frac{\hbar_0}{2}, \dots\,, \tag{57.86}$$

where $\hbar_0$ is the greatest admissible value.

For the weakening of the projectability condition, see Bates and Weinstein (1997). However, even without weakening that condition, there is still a remaining issue to the quantization of classical solutions. This will involve Maslov indices (Section 37.3.2) and Morse theory (Theorem 37.3.10).

Chapter 58

Quantum Information Theory

The section on (quantum) reference frames is based on De La Hamette and Galley (2020).

Chapter Contents

58.1 Entanglement **1079**
58.1.1 Introduction 1079
58.1.2 Bell states 1080
58.1.3 SRE states 1081
58.2 Density operators **1081**
58.3 Channels **1082**
58.4 Quantum logic **1083**
58.4.1 Birkhoff–von Neumann logic 1083
58.4.2 Linear logic 1084
58.5 Topos theory ♣ **1090**
58.5.1 Bohr topos 1090
58.5.2 Internal logic 1091

58.1 Entanglement

58.1.1 Introduction

Construction 58.1.1 (Schmidt decomposition). Consider a bipartite state $|\psi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$. For any such state, there exist orthonormal sets $\{|e_i\rangle, |f_j\rangle\}_{i,j\leq\kappa}$ such that

$$|\psi\rangle = \sum_{i=1}^{\kappa} \lambda_i |e_i\rangle \otimes |f_i\rangle \,, \tag{58.1}$$

where the coefficients λ_i are nonnegative real numbers. All objects in this expression can be obtained from a singular value decomposition of the coefficient matrix $\mathbf{C}$ of $|\psi\rangle$ in some bases of $\mathcal{H}_1$ and $\mathcal{H}_2$. The number $\kappa \in \mathbb{N}$ is called the **Schmidt rank** of $|\psi\rangle$.

Definition 58.1.2 (Entangled states). Consider a state $|\psi\rangle$ and consider its Schmidt decomposition. If the Schmidt rank is 1, i.e. the state can be written as $|\psi\rangle = |v\rangle \otimes |w\rangle$, the state is said to be **separable**. Otherwise, the state is said to be entangled.

The following theorem follow from the linearity of quantum mechanics.

Theorem 58.1.3 (No-cloning). *There is no unitary operator $\widehat{U}$ on a Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$ such that*

$$\widehat{U}|\psi\rangle_1|\phi\rangle_2 = e^{i\alpha(\psi,\phi)}|\psi\rangle_1|\psi\rangle_2 \tag{58.2}$$

for all (normalized) $|\psi\rangle_1 \in \mathcal{H}_1$ and $|\phi\rangle_1 \in \mathcal{H}_2$.

Theorem 58.1.4 (No-deleting). *Consider a tripartite system $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3$ such that $\mathcal{H}_1 \cong \mathcal{H}_2$. If $\widehat{U}$ is a unitary operator on $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3$ such that*

$$\widehat{U}|\psi\rangle_1|\psi\rangle_2|\phi\rangle_3 = |\psi\rangle_1|0\rangle_2\left|\phi_\psi\right\rangle_3 \tag{58.3}$$

for all $|\psi\rangle_1 \in \mathcal{H}_1$, where the final ancilla state $\left|\phi_\psi\right\rangle_3$ might depend on the initial state $|\psi\rangle_1$, then $\widehat{U}$ is simply a swap, i.e. $|\psi\rangle_1 \mapsto \left|\phi_\psi\right\rangle_3$ is an isometric embedding.

58.1.2 Bell states

Definition 58.1.5 (Bell state). A (binary) Bell state (also called a **cat state** or **Einstein–Podolsky–Rosen pair**) is defined as the following entangled state:

$$|\Phi^+\rangle := \frac{1}{\sqrt{2}}\big(|00\rangle + |11\rangle\big)\,. \tag{58.4}$$

In fact, this state can be extended to a full maximally entangled basis for the 2-qubit Hilbert space:

$$\begin{aligned} |\Phi^-\rangle &:= \frac{1}{\sqrt{2}}\big(|00\rangle - |11\rangle\big)\,, \\ |\Psi^+\rangle &:= \frac{1}{\sqrt{2}}\big(|01\rangle + |10\rangle\big)\,, \\ |\Psi^-\rangle &:= \frac{1}{\sqrt{2}}\big(|01\rangle - |10\rangle\big)\,. \end{aligned} \tag{58.5}$$

Method 58.1.6 (Dense coding[1]). Consider the Bell state $|\Phi^+\rangle$. By acting with one of the (unitary) spin-flip operators $\widehat{X}, \widehat{Y}, \widehat{Z}$, one can obtain any of the other three Bell

[1] Sometimes called **superdense coding**.

states:

$$\begin{aligned}\widehat{X}|\Phi^+\rangle &= |\Phi^-\rangle\,,\\ \widehat{Y}|\Phi^+\rangle &= |\Psi^+\rangle\,,\\ \widehat{Z}|\Phi^+\rangle &= |\Psi^-\rangle\,.\end{aligned} \tag{58.6}$$

In a typical Alice-and-Bob-style experiment, one can ask whether this observation allows to achieve a better-than-classical communication channel. If Alice performs a spin flip on her qubit, although the resulting state has instantly 'changed' (cf. *spooky action at a distance*), Bob still cannot uniquely determine what this state is (since the resulting state is still maximally entangled). However, if Alice sends her qubit to Bob, the latter can perform a measurement on the composite system to find out what the state is and in this way determine which operation Alice performed ($\mathbb{1}, \widehat{X}, \widehat{Y}, \widehat{Z}$). Alice has thus effectively sent 2 classical bits of information through 1 qubit. Note that due to the fact that Alice still has to send her qubit through classical means, no faster-than-light communication is achieved.

Definition 58.1.7 (GHZ state). The Greenberger–Horne–Zeilinger state is defined as the multiparticle qudit ($d, N > 2$) version of the Bell state and is, therefore, also referenced to as a cat state:

$$|\mathsf{GHZ}\rangle = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} |i\rangle^{\otimes N}\,. \tag{58.7}$$

58.1.3 SRE states

@@ ADD @@

58.2 Density operators

Definition 58.2.1 (Density operator). Consider a (finite-dimensional) Hilbert space $\mathcal{H}$. A density operator on $\mathcal{H}$ is a linear operator $\rho \in \mathsf{End}(\mathcal{H})$ satisfying the following properties:

1. **Positivity**: $\langle v \,|\, \rho v\rangle \geq 0$ for all $v \in \mathcal{H}$,
2. **Hermiticity**: $\rho^\dagger = \rho$, and
3. **Unit trace**: $\mathrm{tr}(\rho) = 1$.

More concisely, density operators are the representing objects of normal states (Definition 26.2.29) on $\mathcal{B}(\mathcal{H})$.

Example 58.2.2 (Classical probability). A diagonal density matrix corresponds to a discrete probability distribution.

Formula 58.2.3 (Expectation value). Extending Definition 57.1.5, the expectation of an observable $\widehat{O}$ with respect to a density operator $\hat{\rho}$ is given by

$$\langle \widehat{O} \rangle_{\hat{\rho}} = \mathrm{tr}\big(\hat{\rho}\widehat{O}\big) \,. \tag{58.8}$$

Definition 58.2.4 (Pure state). A state is said to be pure if it is described by an outer product of a state vector or, equivalently, by an idempotent density matrix:

$$\rho = |\psi\rangle\langle\psi| \,. \tag{58.9}$$

A density matrix that is not of this form gives rise to a **mixed state**.

Definition 58.2.5 (Reduced density operator). Let $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ be the state of a bipartite system. The reduced density operator ρ_A of A is defined as follows:

$$\rho_A := \mathrm{tr}_B |\Psi\rangle\langle\Psi| \,. \tag{58.10}$$

Definition 58.2.6 (Purification). Let ρ_A be the density operator of a system A. A purification of ρ_A is a pure state $|\Psi\rangle$ of some composite system $A \otimes B$ such that

$$\rho_A = \mathrm{tr}_B |\Psi\rangle\langle\Psi| \,. \tag{58.11}$$

Property 58.2.7. Any two purifications of the same density operator ρ_A are related by a transformation $\mathbb{1}_A \otimes \widehat{V}$ with $\widehat{V}$ an isometry.

58.3 Channels

The following definition generalizes the content of Section 26.1.6 to the setting of partial information. When generalizing the projections in a PVM (spectral measure), one obtains a POVM.

Definition 58.3.1 (Positive operator-valued measure). First, let $\mathcal{H}$ be a finite-dimensional Hilbert space. A POVM on $\mathcal{H}$ consists of a finite set of positive (semi)definite operators $\{P_i\}_{i \leq n}$ such that

$$\sum_{i=1}^{n} P_i = \mathbb{1}_{\mathcal{H}} \,. \tag{58.12}$$

The probability to obtain state i, given a general state $\hat{\rho}$, is given by $\mathrm{tr}(\hat{\rho}P_i)$. Note that the operators are not necessarily orthogonal projectors, so n can be greater than $\dim(\mathcal{H})$.

Now, consider a measurable space (X, Σ) and a (possibly infinite-dimensional) Hilbert space $\mathcal{H}$. A POVM on X consists of a function $P : \Sigma \to \mathcal{B}(\mathcal{H})$ satisfying the following conditions:

1. P_E is positive and self-adjoint for all $E \in \Sigma$,

2. $P_X = \mathbb{1}_{\mathcal{H}}$, and
3. for all disjoint $(E_n)_{n\in\mathbb{N}} \subset \Sigma$:

$$\sum_{n\in\mathbb{N}} P_{E_n} = P_{\cup_{n\in\mathbb{N}} E_n} . \tag{58.13}$$

The following theorem can be derived from Stinespring's theorem 26.2.43.

Theorem 58.3.2 (Naimark dilation theorem). *Every POVM P on $\mathcal{H}$ can be realized as a PVM Π on a, possibly larger, Hilbert space $\mathcal{K}$, i.e. there exists a bounded operator $V : \mathcal{K} \to \mathcal{H}$ such that*

$$P(\cdot) = V\Pi(\cdot)V^\dagger . \tag{58.14}$$

In the finite-dimensional setting, V can be chosen to be an isometry.

Recall the content of Section 26.2.

Definition 58.3.3 (Completely positive trace-preserving). Consider a map $\Phi : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ between bounded operators on two (finite-dimensional) Hilbert spaces. This map preserves density matrices if it positive (Definition 26.2.14) and if it is trace-preserving (Definition 26.2.25). Furthermore, to ensure that an operation applied to a subsystem does not interfere with the positivity of the complete system, they are also required to be completely positive (Definition 26.2.15).

Completely positive, trace-preserving (CPTP) maps are often called **quantum channels** or **superoperators**.

Theorem 58.3.4 (Choi–Jamiołkowski). *The following map between quantum channels $\Phi : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ and density operators $\rho \in \mathsf{End}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ is an isomorphism:*

$$\Phi \mapsto (\mathbb{1}_{\mathcal{H}_1} \otimes \Phi)|\mathsf{GHZ}\rangle\langle\mathsf{GHZ}| , \tag{58.15}$$

where the GHZ state was introduced in Definition 58.1.7.

58.4 Quantum logic

58.4.1 Birkhoff–von Neumann logic

Consider classical propositional logic. This is governed by the Boolean property, i.e. the set of all propositions admits the structure of a complete Boolean algebra (Property 15.1.12). Now, the question becomes how to treat propositions in quantum mechanics (as needed in, for example, quantum computing). In the sense of *von Neumann*, the propositions should be characterized by $\{0,1\}$-valued observables or, equivalently, by projection operators. As such, the natural lattice to consider logic is that of closed subspaces of

the state space $\mathcal{H}$ (which is a Hilbert space). Birkhoff–von Neumann logic is the study of such lattices.

Note that, in contrast to classical logic, the lattices of closed subspaces are not Boolean. The lattices are merely complete, orthomodular lattices (Definition 1.7.42).[2] Now, although these lattices are themselves very interesting, there relevance for quantum logic are heavily discussed for several reasons (e.g. lack of distributivity, lack of a clear implication operator, lack of an extension to predicate logic). In the next section it will be explained how this issue can be avoided by embedding Birkhoff–von Neumann logic into the more general framework of linear logic.

58.4.2 Linear logic

A basic component of standard logic are the *structural inference rules*. These are inference rules that do not involve any logical operations. The following two inference rules for first-order logic control context extension (for an introduction to sequent calculus, see Section 15.1.3):

- **Contraction**:

$$\frac{\Gamma, p_1 : P, p_2 : P \vdash t_{p_1,p_2} : T}{\Gamma, p : P \vdash t_{p,p} : T\,.} \tag{58.16}$$

  This rule states that, in a valid judgement, premises might be used more than once.

- **Weakening**:

$$\frac{\Gamma \vdash P : \mathsf{Type} \qquad \Gamma \vdash t : T}{\Gamma, P \vdash t : T\,.} \tag{58.17}$$

  This rule states that, any premise can be added to (the premises of) a valid judgement.

In terms of categorical semantics, these two rules correspond (in the independent setting) to the diagonal and projection morphisms in Cartesian categories (Example 4.1.2).

Now, when considering quantum mechanics, two important results are the no-cloning and no-deleting theorem (Theorem 58.1.3 and Theorem 58.1.4). These correspond to the fact that the categories **FinVect** and **Hilb** are monoidal, but not Cartesian monoidal, i.e. the tensor product does not admit diagonal and projection morphisms. The natural type of logic in this setting is than a *substructural* one where the contraction and weakening rules are not valid.

[2] It should be noted that complete, orthomodular lattices are, in general, very different from those originating from Hilbert spaces (cf. *Piron's theorem*).

In linear logic, the following propositions exist:

1. **Variables**: Every propositional variable is a proposition.
2. **Negation**: If P is a proposition, so is $P^{\perp}$.
3. **Connectives**: If P, Q are propositions, then
   - **Additive conjunction**: $P \& Q$ is a proposition. (Read: P with Q.)
   - **Additive disjunction**: $P \oplus Q$ is a proposition. (Read: P plus Q.)
   - **Multiplicative conjunction**: $P \otimes Q$ is a proposition. (Read: P times Q.)
   - **Multiplicative disjunction**: $P \parr Q$ is a proposition. (Read: P par Q.)
4. **Constants**:
   - **Additive truth**: $\top$,
   - **Additive falsity**: $\mathbf{0}$,
   - **Multiplicative truth**: $\mathbf{1}$, and
   - **Multiplicative falsity**: $\perp$.
5. **Exponential connectives**: If P is a proposition, then
   - **Exponential conjunction**: $!P$ is a proposition. (Read: of course P.)
   - **Exponential disjunction**: $?P$ is a proposition. (Read: why not P.)

Given a context, the following inference rules are valid:[3]

1. **Identity***: If P is a propositional variable, then $P \vdash P$.
2. **Exchange***: Sequents remain valid under permutations.
3. **Restricted weakening***: If P is a proposition, then

$$\frac{\Gamma \vdash \Theta}{\Gamma, !P \vdash \Theta} \tag{58.18}$$

   and, dually,

$$\frac{\Gamma \vdash \Theta}{\Gamma \vdash \Theta, ?P}. \tag{58.19}$$

[3]The form of these rules heavily depends on the exchange rule (the second item). Care must be taken if this rule is weakened.

4. **Restricted contraction***: If P is a proposition, then

$$\frac{!P, !P \vdash \Theta}{!P \vdash \Theta} \tag{58.20}$$

and, dually,

$$\frac{\Gamma \vdash ?P, ?P}{\Gamma \vdash ?P}. \tag{58.21}$$

5. **Negation**: If P is a proposition, then

$$\frac{\Gamma \vdash \Theta, P}{\Gamma, P^{\perp} \vdash \Theta} \tag{58.22}$$

and, conversely,

$$\frac{\Gamma, P \vdash \Theta}{\Gamma \vdash \Theta, P^{\perp}}. \tag{58.23}$$

Note that these rules allow to write any sequent in right form, i.e. $\vdash \Gamma^{\perp}, P$.

6. **Additive conjunction**: If P, Q are propositions, then

$$\frac{P \vdash \Theta}{P\&Q \vdash \Theta} \qquad \frac{Q \vdash \Theta}{P\&Q \vdash \Theta} \tag{58.24}$$

and, conversely,

$$\frac{\Gamma \vdash P \qquad \Gamma \vdash Q}{\Gamma \vdash P\&Q}. \tag{58.25}$$

7. **Additive disjunction**: If P, Q are propositions, then

$$\frac{\Gamma \vdash P}{\Gamma \vdash P \oplus Q} \qquad \frac{\Gamma \vdash Q}{\Gamma \vdash P \oplus Q} \tag{58.26}$$

and, conversely,

$$\frac{P \vdash \Theta \qquad Q \vdash \Theta}{P \oplus Q \vdash \Theta}. \tag{58.27}$$

8. **Multiplicative conjunction**: If P, Q are propositions, then

$$\frac{P, Q \vdash \Theta}{P \otimes Q \vdash \Theta} \tag{58.28}$$

and, conversely,

$$\frac{\Gamma \vdash P \qquad \Lambda \vdash Q}{\Gamma, \Lambda \vdash P \otimes Q}. \tag{58.29}$$

9. **Multiplicative disjunction**: If P, Q are propositions, then

$$\frac{\Gamma \vdash P, Q}{\Gamma \vdash P \parr Q} \tag{58.30}$$

and, conversely,

$$\frac{P \vdash \Delta \qquad Q \vdash \Theta}{P \parr Q \vdash \Delta, \Theta}. \tag{58.31}$$

10. **Truth and falsity**:

$$\begin{array}{cc} \Gamma \vdash \top & \mathbf{0} \vdash \Theta \\ \dfrac{\Gamma \vdash \Theta}{\Gamma, \mathbf{1} \vdash \Theta} & \vdash \mathbf{1} \\ \dfrac{\vdash \Theta}{\vdash \Theta, \bot} & \bot \vdash . \end{array} \tag{58.32}$$

11. **Exponential conjunction**: If P is a proposition, then

$$\frac{P \vdash \Theta}{!P \vdash \Theta} \tag{58.33}$$

and, conversely, whenever Γ consists solely of !-propositions and Θ consists solely of ?-propositions,

$$\frac{\Gamma \vdash P}{\Gamma \vdash !P}. \tag{58.34}$$

12. **Exponential disjunction**: If P is a proposition, then

$$\frac{\Gamma \vdash P}{\Gamma \vdash ?P} \tag{58.35}$$

and, conversely, whenever Γ consists solely of !-propositions and Θ consists solely of ?-propositions,

$$\frac{P \vdash \Theta}{?P \vdash \Theta}. \tag{58.36}$$

The inference rules with an asterisk are the structural rules. Note that a *cut-elimination theorem* holds and, hence, the identity and cut rules for general propositions can be derived from the rules above.

Linear implication is characterized as follows:

$$P \vdash Q \iff \vdash P^{\perp} \parr Q \iff \vdash P \multimap Q. \tag{58.37}$$

Remark 58.4.1 (Resource theory). Before passing to the properties that follow from the basic rules and the categorical semantics of linear logic (eventually passing to linear type theory), it is useful to rephrase the connectives and their inference rules in terms of 'resources'.

In this interpretation, an implication $A \implies B$ would mean that the resources A can be used to obtain the resources B. However, in ordinary logic, if A and $A \implies B$ hold, one can derive that B holds, but A also still holds. This is something that does not work with resources. If you can use resources A to construct B, the resources A are (usually[4]) used up.[5] One, hence, needs a more subtle and nuanced framework to capture these notions: linear logic. The implication that will be used, where resources are spent, is denoted by $A \multimap B$ for clarity. The two conjunctives, $\otimes$ and &, mean that two resources are available concurrently and separately, respectively. So, if $A \multimap B$ holds, then, since the contraction rule is not valid, one does not have $A \multimap B \otimes B$. However, one does have $A \otimes A \multimap B \otimes B$.

The interpretation of the connectives is as follows:

- $A \otimes B$: A and B are both available for use at the same time, e.g. one has a warehouse with A and a warehouse with B at the same time.
- $A \& B$: Either A or B are available for use, but not both.
- $A \parr B$:
- $A \oplus B$:

The connectives in linear logic satisfy (or generalize) many of the properties of ordinary logic.

Property 58.4.2 (Distributivity).

$$\begin{aligned} P \otimes (Q \oplus R) &= (P \otimes Q) \oplus (P \otimes R) \\ P \parr (Q \& R) &= (P \parr Q) \& (P \parr R) \\ P \otimes \mathbf{0} &= \mathbf{0} \\ P \parr \top &= \top \end{aligned} \tag{58.38}$$

The exponential connectives can be used to turn additive connectives into multiplicative ones (and the other way around) as with the ordinary exponential function in calculus.

[4] This is not the case with catalysts.

[5] One could also give this a causal flavour (Girard, 1995).

Property 58.4.3.

$$\begin{aligned} !(P\&Q) &= !P\otimes !Q \\ ?(P\oplus Q) &= !P\parr !Q \\ !\top &= \mathbf{1} \\ ?\mathbf{0} &= \bot \end{aligned} \tag{58.39}$$

Moreover, due to the apparent similarity with the operators in (S4) modal logic (Section 15.6), the exponential connectives are sometimes also called **modalities**.

Linear negation can also be defined alternatively.

Property 58.4.4 (Negation). Linear negation admits the following recursive definition:

- $P^{\perp\perp} = P$,
- $(P\&Q)^\perp = P^\perp \oplus Q^\perp$,
- $(P\otimes Q)^\perp = P^\perp \parr Q^\perp$,
- $\top^\perp = \mathbf{0}$,
- $\mathbf{1}^\perp = \bot$, and
- $(!P)^\perp = ?P^\perp$.

Property 58.4.5 (Categorical semantics). Whereas standard Boolean logic is the internal logic of Cartesian closed categories — where conjunction, disjunction and implication correspond, respectively to products, coproducts and internal homs — linear logic is the internal logic of (a subclass of) $*$-autonomous categories (Definition 4.4.8).

The multiplicative conjunction $\otimes$ corresponds to the tensor product, hence the notation. Similar to ordinary logic, the linear implication $\multimap$ corresponds to taking internal homs. The important part, now, is that negation comes as a separate entity, in this case given by taking duals: $x^\perp \equiv x^*$. The multiplicative disjunction $\parr$ is then constructed through Property 58.4.4 (which corresponds to de Morgan duality as in Property 4.4.9).

For the additive connectives, one needs the existence of finite products. The $*$-autonomy then also implies the existence of finite coproducts (again through de Morgan duality).

For the exponential connectives, some more structure is needed. As with modal logic, the structure is given by the existence of a suitable (co)monad.

58.5 Topos theory ♣

58.5.1 Bohr topos

Definition 58.5.1 (Bohr topos). Consider a C^*-algebra A and denote by $\mathsf{ComSub}(A)$ the poset (Definition 1.7.1) of commutative C^*-subalgebras. This set can be equipped with the **Alexandrov topology**[6], i.e. the topology for which the open sets are the upward closed subsets. The topological space $(\mathsf{ComSub}(A), \tau_{\text{Alex}})$ is called the Bohr site of A.

The sheaf topos over the Bohr site is called the Bohr topos $\mathbf{Bohr}(A)$. It can be turned into a ringed topos, where the ring object (which is even an internal commutative C^*-algebra) is given by the tautological functor

$$\underline{A} : \mathsf{ComSub}(A) \to \mathbf{Set} : C \mapsto C\,. \tag{58.40}$$

Property 58.5.2. A morphism in $\mathbf{C^*Alg}$ is commutativity reflecting if and only if the induced morphism on posets admits a right adjoint. Moreover, there exists a bijection between the following two classes of morphisms:

- Geometric morphisms $f : \mathbf{Bohr}(B) \to \mathbf{Bohr}(A)$ admitting a right adjoint together with epimorphisms of internal algebras $\underline{A} \to f^*\underline{B}$.
- Commutativity-reflecting functions $f : A \to B$ that restrict to algebra morphisms on all commutative subalgebras.

Definition 58.5.3 (Spectral presheaf). The presheaf Σ on a Bohr site assigning to every commutative subalgebra its Gel'fand spectrum.

The idea behind the Bohr topos is that, given a general C^*-algebra A, the Bohr topos $\mathbf{Bohr}(A)$ is interpreted as its quantum phase space. This is similar to Section 44.4, where smooth spaces are also reinterpreted in terms of sheaf topoi.

Theorem 58.5.4 (Kochen–Specker). *For $A = \mathcal{B}(\mathcal{H})$, the spectral presheaf has no global elements if* $\dim(\mathcal{H}) > 2$.

Property 58.5.5 (Gleason's theorem). There exists a natural bijection between the quantum states of a C^*-algebra A and the classical states of $\underline{A}$ internal to $\mathbf{Bohr}(A)$.

Definition 58.5.6 (Bohrification). Consider a C^*-algebra A together with its Bohr topos $\mathbf{Bohr}(A)$. To its internal C^*-algebra $\underline{A}$, one can assign an internal locale $\underline{\Sigma}_A$ by (internal) Gel'fand duality (Definition 26.2.33). Under the equivalence 12.4.38, one then obtains a locale Σ_A. The functor

$$\Sigma : \mathbf{C^*Alg} \to \mathbf{Loc} : A \mapsto \Sigma_A \tag{58.41}$$

[6]There exist an equivalences $\mathbf{Pre} \cong \mathbf{AlexTop}$ and $\mathbf{Pos} \cong \mathbf{AlexTop}_{T_0}$.

is called Bohrification. This locale can be constructed as the disjoint union

$$\Sigma_A = \bigsqcup_{C \in \mathsf{ComSub}(A)} \Phi_C \,, \tag{58.42}$$

the étale locale corresponding to the spectral presheaf (i.e. the spectral presheaf is the internal Gel'fand spectrum of $\underline{A}$). Its open sets are given by those subsets whose restrictions to commutative subalgebras are open in such a way that these restrictions are compatible with subalgebra inclusions.

Example 58.5.7 (Gel'fand spectrum). If A is a commutative C^*-algebra, its Bohrification is not isomorphic to its ordinary Gel'fand spectrum Φ_A. However, after replacing the topology on $\mathbf{Bohr}(A)$ by the double negation topology (Example 12.4.41) and repeating the above construction, one obtains

$$\Phi_A \cong \Sigma_A^{\neg\neg} \,. \tag{58.43}$$

This locale can also be obtained in another way. Double negation $\neg\neg$ defines an (internal) **nucleus** on the (internal) locale Σ_A, i.e. a left-exact monad. $\Sigma_A^{\neg\neg}$ is then given by the fixed points of $\neg\neg : \Sigma_A \to \Sigma_A$.

By Property 58.5.2 above, the following relation is obtained.

Property 58.5.8 (Observables). Morphisms $\mathbf{Bohr}(A) \to \mathbf{Bohr}(C(\mathbb{R})_0)$ admitting a right adjoint together with an epimorphism $\underline{C_0(\mathbb{R})} \to f^*\underline{A}$ correspond to observables on A.

The topological bundle $\Sigma_A \to \mathsf{Alex}(\mathsf{ComSub}(A))$ also admits a topos-theoretic incarnation. There exists a (canonical) morphism of ringed topoi

$$\pi : \mathbf{Bohr}(A) \to \big(\mathbf{Sh}\big(\mathsf{Alex}(\mathsf{ComSub}(A))\big), \underline{\mathbb{R}}\big) \,, \tag{58.44}$$

whose underlying geometric morphism is simply the identity.

Property 58.5.9 (States). A positive and normalized section of the morphism $\pi : \mathbf{Bohr}(A) \to \big(\mathbf{Sh}\big(\mathsf{Alex}(\mathsf{ComSub}(A))\big), \underline{\mathbb{R}}\big)$ in the category of $\underline{\mathbb{R}}$-*module topoi*.

58.5.2 Internal logic

Whereas Section 58.4 covers quantum logic from the external point of view, the spectral presheaf (Definition 58.5.3) allows to treat it internally. By functional calculus (Section 26.1.6), every proposition $A \in \Delta$ about a self-adjoint operator on a Hilbert space $\mathcal{H}$, i.e. every measurable subset of the spectrum $\sigma(A)$, corresponds to a projection $P_{A,\Delta}$. Now, the goal will be to assign to every such projection a 'classical interpretation' in the classical **context** given by a commutative subalgebra of A. This is achieved by the subobject of the spectral presheaf Σ constructed as follows:

$$\delta_P(V) := \{\omega \in \Sigma_v \mid \omega(P|_V) = 1\} \,,$$

where $P|_V$ is the smallest projection in V such that $\mathsf{im}(P) \subseteq \mathsf{im}(P|_V)$. The morphism

$$\delta : \mathcal{P}(\mathcal{H}) \to \mathsf{Sub}(\Sigma)$$

is called the **daseinization** map. Note that this map extends to all self-adjoint operators through functional calculus.

A 'pure state' ψ in classical mechanics corresponds to a map from subsets of phase space to the subobject classifier (saying whether the point lies in the given subset or not):

$$T^{\psi} : \mathsf{Sub}(P) \to \{0, 1\} .$$

By analogy, a pure state $|\psi\rangle$ in quantum mechanics will correspond to a morphism from subobjects of the spectral presheaf:

$$T^{|\psi\rangle} : \mathsf{Sub}(\Sigma) \to \mathsf{Hom}_{\mathbf{Psh}(\mathsf{ComSub}(A))}(1, \Omega) ,$$

where Ω is the subobject classifier from Property 12.4.4.

Chapter 59

Quantization ♣

The content of Chapters Vector Bundles, Principal Bundles and Symplectic Geometry are prerequisites for the section on geometric quantization. References for the section on geometric quantization are Bates and Weinstein (1997); Brylinski (1993); Camosso (2021); Tuynman (2016). The section on Toeplitz quantization is mainly based on Hawkins (2000). A general reference comparing deformation and geometric quantization is Hawkins (1998).

Chapter Contents

59.1 Introduction **1093**
59.2 Deformation quantization **1094**
59.3 Geometric Quantization **1095**
59.3.1 Prequantization 1095
59.3.2 Coadjoint orbits 1100
59.4 Toeplitz quantization **1100**
59.5 Constrained systems **1102**
59.5.1 Dirac procedure 1102
59.5.2 BRST quantization 1103
59.5.3 BFV quantization 1103

59.1 Introduction

Given the content of Section 57.4, the general quantization procedure for a dynamical system (M, ω, H) can be axiomatized as follows.

Method 59.1.1 (Abstract quantization). A quantization of a symplectic manifold (M, ω) is a pair $(\mathcal{H}, \mathcal{O})$ where:

- $\mathcal{H}$ is a (complex) separable Hilbert space,
- $\mathcal{O}$ takes real functions $C^\infty(M)$ to self-adjoint operators,
- $\mathcal{O}$ is $\mathbb{C}$-linear,
- $\mathcal{O}(1) = \mathbb{1}_{\mathcal{H}}$, and
- **Dirac correspondence**: $[\mathcal{O}(f), \mathcal{O}(g)] = i\hbar\mathcal{O}(\{f, g\}) + O(\hbar^2)$.

Remark 59.1.2 (C^*-bundles). Hawkins (2000) defines a general quantization as a *continuous bundle* $\pi : A \to [0, +\infty[$ *of* C^**-algebras* such that $\mathrm{ev}_0(\Gamma(A)) = C^\infty(M)$. See Section 59.4 for more information.

Because a Hilbert space forms an irreducible representation of any complete set of observables, it makes sense to add an additional irreducibility axiom.

Axiom 59.1 (Irreducibility postulate). Let (M, ω) be a $2n$-dimensional symplectic manifold. If the observables $\{f_i\}_{i\leq n}$ form a complete set, i.e. any function that Poisson commutes with all f_i is necessarily constant, the quantum state space $\mathcal{H}$ is required to be irreducible with respect to the action of $\{\mathcal{O}(f_i)\}_{i\leq n}$. Equivalently, if a group G acts transitively on M, the state space is required to be an irreducible representation of a $\mathrm{U}(1)$-central extension of G.

The simplest example is again $T^*\mathbb{R}^n$. Here, the usual choice of observables are the coordinate and momentum functions $\{q^i, p_i\}_{i\leq n}$ and the group G is given by the *Heisenberg group*, a $\mathrm{U}(1)$-central extension of the translation group $\mathbb{R}^{2n}$. Irreducible representations are characterized by the Stone–von Neumann theorem 57.4.2.

Although the above procedure sounds reasonable, it is known to be inconsistent. *Groenewold* and *Van Hove* showed that there does not exist a map $\mathcal{O}$, satisyfing the axioms 59.1.1, that takes the entire (Lie) algebra of classical observables to the (Lie) algebra of corresponding quantum observables. As such, a different approach should be followed.

59.2 Deformation quantization

Definition 59.2.1 (Formal deformation quantization). Let A be a Poisson algebra. A **star product** on A is an associative, $\mathbb{R}$-bilinear product on $A[[\hbar]]$ such that for all $a, b \in A$:

$$a * b = \sum_{i=0}^{+\infty} B_i(a, b)\hbar^i \tag{59.1}$$

with $B_0(a, b) = ab$. A formal deformation quantization of A is a star product such that

$$\{a, b\} = B_1(a, b) - B_1(b, a)\,. \tag{59.2}$$

Definition 59.2.2 (Moyal deformation quantization). Let V be a Poisson vector space[1], i.e. a vector space equipped with a Poisson bivector $\pi \in V \wedge V$. The ring $C^\infty(V)[[\hbar]]$ can be equipped with the following product:

$$f * g := \mu \circ \exp(\hbar\pi)(f \otimes g)\,, \tag{59.3}$$

where π is viewed as an endomorphism on $C^\infty(V) \otimes C^\infty(V)$ and μ is the ordinary product map.

Definition 59.2.3 (Fedosov deformation quantization). Consider a symplectic manifold (M, ω). Every tangent space T_pM carries the structure of a Poisson vector space and, hence, admits a Moyal quantization $(T_pM, *_\omega)$. These spaces can be turned into an algebra bundle $\mathcal{A}$ over M. It can, moreover, be shown that any such bundle admits at least one flat connection that respects the algebra structure and such that there exists an isomorphism as follows:

$$\ker(\nabla^{\text{Fed}}) \cong C^\infty(M)[[\hbar]]\,. \tag{59.4}$$

This connection is called a **Fedosov connection**. The restriction of the star product on $\Gamma(\mathcal{A})$ to this subbundle gives a formal deformation quantization of $C^\infty(M)$.

Remark 59.2.4 (Kontsevich quantization). *Kontsevich* had previously proven that every finite-dimensional Poisson manifold admits a deformation quantization of its algebra of smooth functions. However, a slight variation of *Fedosov*'s approach can be used even for infinite-dimensional algebras such as those occuring in field theory (where Kontsevich's approach is invalid).

59.3 Geometric Quantization

In this section, the natural constant $\hbar$ has been set to 1.

59.3.1 Prequantization

Definition 59.3.1 (Prequantum line bundle). Consider a symplectic manifold (M, ω). A prequantum line bundle on M is a Hermitian line bundle equipped with a connection ∇ such that $\omega = F_\nabla$, where F_∇ denotes the curvature of ∇.

Property 59.3.2. Complex line bundles are classified by the (first) Chern class $c_1 \in H^2(M; \mathbb{Z})$, which is proportional to the curvature form through Chern–Weil theory. Therefore, a prequantum line bundle exists if and only if the symplectic form is integral (up to a factor of 2π). For simply connected manifolds, this is equivalent to

$$\forall S \in H_2(M) : \int_S \omega \in 2\pi\mathbb{Z}\,. \tag{59.5}$$

[1] Also called a **linear Poisson manifold**.

This condition resembles the 'old' *Bohr–Sommerfeld condition*[2] and is, in general, known as the **Weil integrality condition**.

There are two contributions to the moduli space of prequantum line bundles or, essentially, U(1)-principal bundles with connection. The latter are known to be classified by differential U(1)-cohomology as shown in Section 35.7. Since the curvature is fixed by the sympletic form, prequantum line bundles are classified by lifts from curvature forms to differential cohomology along the curvature projection. Differential cohomology consists of two parts, roughly corresponding to the following two aspects: there can exist topologically inequivalent bundles and there can exist inequivalent connections on the same bundle differing by a flat connection, where two flat connections are in turn inequivalent if they differ by a closed one-form that is neither integral nor exact. The former are classified by integral curvature forms $H^2_{\mathrm{dR}}(M)$, while the latter are classified by the Čech cohomology group $H^1(M;\mathsf{U}(1))$ or, equivalently by isomorphism (34.146), $H^2(M;\mathbb{Z})$.

Corollary 59.3.3 (Dirac quantization condition). Weil integrality implies that, The minimal coupling procedure, where ∇ is coupled to an external gauge field, is given by the following substitution: $\omega \longrightarrow \omega + eF$. Weil integrality then implies that e is an integer.

Classically, the Dirac quantization condition is obtained by assuming the existence of a magnetic charge g. To obtain a well-defined holonomy when transporting an electric charge q around g, one needs

$$qg \in 2\pi\mathbb{Z}\,. \tag{59.6}$$

As soon as one magnetic monopole is present in the universe, the quantization of electric charges would be evident!

Definition 59.3.4 (Prequantum Hilbert space). Consider a symplectic manifold (M,ω) together with a prequantum line bundle L. The prequantum Hilbert space $\mathcal{H}_L$ is defined as (the L^2-completion of) the space of square-integrable sections of L with respect to the metric on L and the Liouville volume form on M.

To every smooth function $f \in C^\infty(M,\mathbb{C})$, one can associate a (**Segal–Kostant–Souriau**) prequantum operator $\hat{f} : \Gamma(L) \to \Gamma(L)$ by the following formula:

$$\hat{f} : \psi \mapsto f \cdot \psi - i\nabla_{X_f}\psi\,, \tag{59.7}$$

where X_f is the Hamiltonian vector field associated to f and, locally,

$$\nabla = \mathsf{d} - i\theta \tag{59.8}$$

with θ the symplectic potential. This operator can also be interpreted in terms of a Hamiltonian flow. The Hamiltonian flow of X_f can be lifted (up to a phase) to an

[2] In fact, it is closer to Einstein's quantization rule (Stone, 2005).

automorphism $\psi_t[f]$ on L that preserves both the metric and the connection. The prequantum operator $\hat{f}$ is then simply given by

$$\hat{f}s = -i\frac{\mathrm{d}}{\mathrm{d}t}(\psi_t[f]s)\bigg|_{t=0}. \tag{59.9}$$

Example 59.3.5 (Spinning particle). Consider as phase space the 2-sphere $S^2(r)$ with radius $r \in \mathbb{R}$. In this case, the symplectic form can be written as $\omega = r^2 \sin\theta \mathrm{d}\theta \wedge \mathrm{d}\varphi$. This form is only integral for a discrete set of values of r, namely for $r \in \mathbb{Z}/2$. Up to a factor $\hbar$, this is exactly the quantization rule for the angular momentum. The reason is that S^2 is a homogeneous space for $\mathsf{SU}(2)$, the group characterizing spinning particles. In fact, using the theory of coadjoint orbits (see further below), one can show that this quantization procedure coincides with the *KKS quantization* of the coadjoint orbit $S^2 = \mathsf{SU}(2)/\mathsf{U}(1)$.

At this point, it can easily be seen that there is a problem with the dimension of the prequantum state space. For the cotangent bundle $T^*\mathbb{R}^n \cong \mathbb{R}^n \times \mathbb{R}^n$, the resulting state space would be $L^2(\mathbb{R}^{2n}, \mathbb{C})$. However, from ordinary quantum theory it is well-known that the right Hilbert space is $L^2(\mathbb{R}^n, \mathbb{C})$. In general, the above procedure would give wave functions that depend on $2n$ variables instead of the n coordinates of configuration space that are normally found in quantum mechanics. A solution is obtained by making a choice of 'configuration space' or, in terms of ordinary quantum mechanics, to choose a 'representation' of the system.

Construction 59.3.6 (Geometric quantization). Let (M, ω) be a symplectic manifold. A (geometric) quantization of (M, ω) is given by a prequantum line bundle L together with a polarization $\mathcal{P}$ of M. The 'naive' modification of the quantum state space is given by the subspace of $\mathcal{H}_L$ of those sections that are covariantly constant along $\mathcal{P}$, i.e. those sections $s \in \Gamma(L)$ that satisfy $\nabla_X s = 0$ for all $X \in \mathcal{P}$. These sections are also called **polarized sections**.

The fact that a polarization is required, and not merely an n-dimensional involutive distribution, follows from an additional consistency condition imposed by the condition $\nabla_X s = 0$ for all $X \in \mathcal{P}$. Because ω also represents the curvature of the connection on L, one obtains

$$\omega(X, Y)s = [\nabla_X, \nabla_Y]s - \nabla_{[X,Y]}s = 0 \tag{59.10}$$

for all $X, Y \in \mathcal{P}$. This implies that $\mathcal{P}$ defines an isotropic submanifold. For a completely integrable system, a natural choice would be given by the distribution spanned by the Hamiltonian vector fields.

Now, if the prequantum operators ought to represent genuine operators on the quantum state space, one should have

$$\nabla_X s = 0 \implies \nabla_X(\hat{f}s) = 0 \tag{59.11}$$

for all sections $s \in \Gamma(L)$ and $X \in \mathcal{P}$. Using the general formula for iterated covariant derivatives and the fact that the leaves of P are Lagrangian, one finds the following condition:

$$[X, X_f] = 0 \tag{59.12}$$

for all $X \in \mathcal{P}$. So, in general, one should restrict to the subspace of $C^\infty(M)$ on those functions whose Hamiltonian flow preserves $\mathcal{P}$.

There is, however, a problem with this construction. Nothing ensures that $\mathcal{H}_L$ contains any polarized sections. As an example, consider a cotangent bundle with its vertical polarization. In this case, the polarized sections are given by functions that only depend on the base coordinates q^i and not on the fibre (momentum) coordinates p_i. However, because the fibres are noncompact, the integral of such a section with respect to the Liouville measure will always diverge. This particular issue can be resolved by integrating over the leaf space $M \cong T^*M/D$, where D is the isotropic foliation associated to $\mathcal{P}$.

However, even though a possible divergence coming from noncompact fibres is resolved, another problem arises. To be able to integrate over a manifold one needs a volume form, but there is not always a canonical choice available. A solution is given by working with densities.

Method 59.3.7 (Half-form quantization). For this method, the polarization is assumed to be real and have simply connected leaves. Furthermore, the manifold M is assumed to admit a metaplectic structure or, by virtue of Property 39.1.26, $\mathcal{P}$ is assumed to admit a metalinear structure. (A choice of such a metaplectic structure is called a **metaplectic correction**.) Using this structure, one can define the half-form bundle $\delta^{1/2}$. Given a prequantum line bundle L, one defines the twisted bundle of L-valued half-forms $L \otimes \delta^{1/2}$. A **wave function** is defined as a section $\psi \in \Gamma(L \otimes \delta^{1/2})$ such that, locally, $\psi = \lambda \otimes \mu$ with $\nabla_X \lambda = 0$ and $\mathcal{L}_X \mu = 0$ for all $X \in \overline{\mathcal{P}}$. By pairing two wave functions, one obtains a 1-density on M that can be integrated. The quantum state space is then defined as the L^2-completion of the space of wave functions.

To extend the definition of operators to half-form quantized manifolds, one simply needs to extend the definition to density bundles. Because $\hat{f}$ represents the Hamiltonian flow, a natural choice is the Lie derivative:

$$\hat{f}(s \otimes \mu) := (\hat{f}s) \otimes \mu - is \otimes \mathcal{L}_{X_f} \mu \,. \tag{59.13}$$

Example 59.3.8 (Kähler quantization). For this method, the polarization is assumed to be positive Kähler, i.e. P is the antiholomorphic tangent bundle of a Kähler manifold. By taking the (local) symplectic potential to be the holomorphic derivative of the Kähler potential, one obtains the space of holomorphic sections as the prequantum

state space (a different choice of potential results in a phase transformation). It can be shown that a natural inner product is given by

$$\langle \psi_1 \mid \psi_2 \rangle = \int_{\mathbb{C}^n} \overline{\psi_1}(z)\psi_2(z) \exp(-|z|^2/2)\, dz^n \,. \tag{59.14}$$

This is often called the **Bargmann**, **Segal–Bargmann** or **Bargmann–Fock** representation. The coordinates $z, \bar{z}$ are represented by the operators z and ∂_z. These correspond to creation and annihilation operators.

Remark 59.3.9. If the positivity assumption would be dropped, the Kähler potential K, which was $-|z|^2/2$ above, would be indefinite. This would, in turn, imply that the integral diverges.

Remark 59.3.10. Recall Section 39.4.2. If M is compact Kähler, the existence of a metaplectic correction is equivalent to the existence of a spin structure. This leads to the L-twisted Dolbeault–Dirac operator:

$$\bar{\partial}_\nabla + \bar{\partial}_\nabla^\dagger : \Omega^{0,\bullet}(M) \otimes L \otimes \delta^{1/2} \to \Omega^{0,\bullet}(M) \otimes L \otimes \delta^{1/2} \,. \tag{59.15}$$

From the perspective of the Dolbeault complex, the differential is twisted by a metaplectically corrected prequantum line bundle, but from the perspective of the Dirac operator, one only twists by the line bundle.

Method 59.3.11 (Bohr–Sommerfeld quantization). Here, the polarization $\mathcal{P}$ is again assumed to be real, but the leaves are not required to be simply connected. The connection ∇ along $\mathcal{P}$ is flat when restricted to a single leaf $\Lambda_m \subset M$. When Λ_m is not simply connected, the holonomy group can be nontrivial. However, the defining condition of $\mathcal{H}$ is that sections should be covariantly constant. This implies that either the section is zero or that the holonomy around any loop vanishes. The support of all sections in $\mathcal{H}$ is, therefore, given by the union S of all leaves on which ∇ is trivial. This space is called the **Bohr–Sommerfeld variety**.

Vanishing holonomy implies

$$\exp\left(i \oint_\gamma \theta \right) = 1 \tag{59.16}$$

for all loops γ, where θ is the symplectic potential. In terms of Darboux coordinates, this gives (up to a factor 2π)

$$\oint_\gamma p_i \, \mathrm{d}q^i \in \mathbb{Z} \,. \tag{59.17}$$

This is exactly the old Bohr–Sommerfeld quantization condition. When using half-density quantization, an additional contribution coming from the covariant derivative

on densities would have to be added to the right-hand side:

$$\oint_\gamma p_i \, \mathrm{d}q^i = 2\pi(k_\gamma + d_\gamma)\,, \tag{59.18}$$

where $k_\gamma \in \mathbb{Z}$.

Definition 59.3.12 (Bohr–Sommerfeld leaf). More generally, a Lagrangian submanifold L of a prequantized symplectic manifold (M, ω) is said to be a **Bohr–Sommerfeld leaf** if $[\nabla_L]$ vanishes in $H^2_{\mathrm{conn}}(L)$, the differential cohomology of L (Section 35.7). On simply connected manifolds, this is equivalent to having trivial holonomy.

This can be restated in terms of lifts to differential cohomology. The prequantization of a symplectic manifold is a lift as shown in the following diagram:

$$\begin{tikzcd} & \mathbf{B}\mathrm{U}(1)_{\mathrm{conn}} \arrow[d, "F"] \\ M \arrow[ur, dashed, "\nabla"] \arrow[r, "\omega"'] & \Omega^2_{\mathrm{cl}}(M) \end{tikzcd} \tag{59.19}$$

where $\mathbf{B}\mathrm{U}(1)_{\mathrm{conn}}$ is the Deligne complex (Definition 35.7.12). A submanifold $L \subset M$ is isotropic (Lagrangian if it is maximal) if the lift

$$\begin{tikzcd} & & \mathbf{B}\mathrm{U}(1)_{\mathrm{conn}} \arrow[d, "F"] \\ L \arrow[r, hook] & M \arrow[ur, dashed, "\nabla"] \arrow[r, "\omega"'] & \Omega^2_{\mathrm{cl}}(M) \end{tikzcd} \tag{59.20}$$

is trivial in $H^2_{\mathrm{conn}}(L)$.

59.3.2 Coadjoint orbits

@@ COMPLETE (SEE ALSO Section 32.3.3) @@

59.4 Toeplitz quantization

Definition 59.4.1 (C^*-algebra bundle). A (continuous) bundle of C^*-algebras over a locally compact Hausdorff space X is a bundle $\pi : A \to X$ such that:

1. $\pi^{-1}(x)$ is a C^*-algebra for all $x \in X$.
2. $\Gamma(A)$ is a C^*-algebra.
3. $\mathrm{ev}_x(\Gamma(A))$ is dense in $\pi^{-1}(x)$ for all $x \in X$.

4. ev_x is a $*$-epimorphism for all $x \in X$.

5. For all $\nu \in \Gamma(A)$,

$$\|\nu\| = \sup_{x \in X} \|\nu(x)\| \,. \tag{59.21}$$

6. For all $f \in C_0(X)$ and $\nu \in \Gamma(A)$, there exists a $\nu_f \in \Gamma(A)$ such that $\nu_f(x) = f(x)\nu(x)$ for all $x \in X$.

7. **Continuity**: For all $\nu \in \Gamma(A)$, the map $x \mapsto \|\nu(x)\|$ is in $C_0(X)$.

@@ CHECK DEFINITION (difference between field and bundle of algebras) @@

In the remainder of this section, M will be assumed to be a compact, connected Kähler manifold. Moreover, consider a prequantization L and a Hermitian line bundle L_0. For all $n \in \mathbb{N}$, consider the line bundle $L \otimes L_0^n$, assumed to be positive (Definition 39.3.3) with its Hilbert space of holomorphic sections $\mathcal{H}_n$ (which is finite dimensional by Property 39.3.2).

Definition 59.4.2 (Toeplitz quantization). The map $\mathfrak{T}_n : C^\infty(M) \to \mathsf{End}(\mathcal{H}_n)$ given by[3]

$$\mathfrak{T}_n(f) := \pi_n \circ f \,, \tag{59.22}$$

where $\pi_n : \Gamma(L \otimes L_0) \to \mathcal{H}_n$ is the canonical projection onto holomorphic sections. It can be shown that $\mathfrak{T}_n$ is unital and completely positive.

To generalize this procedure to vector bundles, note that the Kodaira vanishing theorem (Theorem 39.6.5) implies that the kernel of the $L \otimes L_0^n$-twisted Dolbeault operator $\overline{\partial}$ is concentrated in degree 0 and, hence, that

$$\ker(\overline{\partial}) = \mathcal{H}_n \,. \tag{59.23}$$

The appropriate generalization to vector bundles will be the $E^* \otimes L \otimes L_0^n$-twisted Dolbeault–Dirac operator D_E from Eq. (39.42):

$$\mathfrak{T}_n : \Gamma(E) \to \mathsf{Hom}\big(\ker(D_E), \mathcal{H}_n\big) : \sigma \mapsto \pi_n \circ \sigma \,, \tag{59.24}$$

where multiplication by σ is defined through the natural pairing of E and E^*.

Now, the Toeplitz maps $\mathfrak{T}_n$ piece together to give an operator $\mathfrak{T} : C^\infty(M) \to \prod_{n \in \mathbb{N}} \mathsf{End}(\mathcal{H}_n)$. This operator induces an isomorphism

$$\mathcal{T} : C^\infty(M) \cong \mathbb{A}/\mathbb{A}_0 \,, \tag{59.25}$$

[3]This particular form, first multiplying by a smooth function and then projecting onto holomorphic sections, is very similar to the definition of *Toeplitz operators*, where the projection is onto the *Hardy space*, hence the name.

where $\mathbb{A}_0 := \bigoplus_{n\in\mathbb{N}} \mathsf{End}(\mathcal{H}_n)$ and $\mathbb{A} := \mathbb{A}_0 \oplus \mathsf{im}(\mathfrak{T})$.

A general quantization, in the sense of Remark 59.1.2, is now given by the fact that there exists a bundle $A \to \widehat{\mathbb{N}}$, with $\widehat{\mathbb{N}}$ the one-point compactification$\{0, \dots, +\infty\}$, such that $\Gamma(A) = \mathbb{A}$, $A_n = \mathsf{End}(\mathcal{H}_n)$ for all $n \in \mathbb{N}$, and $A_{+\infty} = \mathcal{T}^{-1} \circ \pi_{\mathbb{A}_0}(\mathbb{A})$.

Remark 59.4.3. An equivalent quantization can be obtained from the operators

$$\mathfrak{Q}_n(f) := \mathfrak{T}_n\left(f + \frac{\Delta f}{2n}\right). \tag{59.26}$$

@@ COMPLETE (Hawkins (2000)) @@

59.5 Constrained systems

In this section, classical systems with constraints, i.e. dynamical systems (M, ω, H) with an algebra of first-class constraints $\{\phi_a\}_{a\in I}$, are considered as in Chapter 54.

59.5.1 Dirac procedure

The first approach to the quantization of constrained systems is due to *Dirac*. Instead of trying to pass to the reduced phase space or introducing additional gauge fixing conditions, *Dirac* simply worked with all variables and represented these as operators acting on an enlarged Hilbert space. The constraints, represented by operators $\widehat{G}_a$, satisfy

$$\widehat{G}_a|\psi\rangle = 0 \tag{59.27}$$

for all physical states $|\psi\rangle$. The constraint algebra $\{\phi_a, \phi_b\} = C^c_{ab}\phi_c$ gives rise to a quantum algebra

$$[\widehat{G}_a, \widehat{G}_b] = i\hbar\, C^c_{ab}\widehat{G}_c + \hbar^2\widehat{D}_{ab}\,, \tag{59.28}$$

where $\widehat{D}_{ab}$ represents a **quantum anomaly**, i.e. a correction term resulting from the quantization of the classical algebra (e.g. operator ordering). The issue here is that both the left-hand side and the first term on the right-hand side vanish exactly on physical states, so $\widehat{D}_{ab}$ should also vanish on these states for all $a, b \in I$. However, if this would be true, the physical Hilbert space would be heavily restricted (in certain cases, it even becomes trivial). The conclusion is that the quantum anomaly breaks the first-class structure of the constraints and, therefore, the constraints do not generate gauge transformations anymore (this is why the anomy is sometimes called a **gauge anomaly**). A similar issue appears when quantizing the classical evolution equation

$$\{H, \phi_a\} = V^b_a\phi_b\,. \tag{59.29}$$

@@ COMPLETE @@

59.5.2 BRST quantization

For this approach, one starts from the classical BRST construction from Section 54.2 and tries to find a Hilbert space representation of the extended algebra containing the classical functions, the ghosts and the ghost momenta. The BRST charge becomes a self-adjoint operator satisfying

$$[\widehat{\Omega}, \widehat{\Omega}] = 2\widehat{\Omega}^2 = 0\,. \tag{59.30}$$

However, in stark contrast to the classical situation, where a BRST charge always exists, the quantum case does not necessarily admit such a construction.

Property 59.5.1 (Ghost states). If the state space can be decomposed according to ghost number, the following statements hold:

- The ghost number of a homogeneous state is of the form

$$g = g_0 + z \in \mathbb{Z} \tag{59.31}$$

  for g_0 either 0 or $\frac{1}{2}$. Fractional ghost numbers occur when the number of constraints is odd.

- The inner product of two homogeneous states with ghost numbers g, g' vanishes if $g + g' \neq 0$. This implies that states with nonzero ghost number are null states.

Definition 59.5.2 (Physical state space). The physical states are defined similarly to the gauge-invariant functions in the classical setting, i.e. states are deemed physical if they are BRST-closed:

$$\widehat{\Omega}|\psi\rangle = 0\,. \tag{59.32}$$

This operation is linear and, hence, filters out a linear subspace as required. Furthermore, BRST-closed operators (the **physical observables**) preserve this subspace and the BRST-exact operators give vanishing transition elements (and are, therefore, not physically observable as desired). This also implies that acting with a BRST-exact operator on a state, leaves the physical state unchanged. It follows that the true physical state space is given by $H^0(\widehat{\Omega})$.

59.5.3 BFV quantization

The extended phase space M_{ext} is defined by introducing dynamical Lagrange multipliers and their momenta

$$\{\lambda^m, \pi_n\} := \delta^m_n \tag{59.33}$$

together with two collections of (homologically) odd-degree ghosts and their momenta:

$$\begin{aligned} \{C^i, \overline{P}_j\} &:= \delta^i_j\,, \\ \{P^i, \overline{C}_j\} &:= \delta^i_j\,. \end{aligned} \tag{59.34}$$

Note that the Poisson bracket for the ghosts is a graded Poisson bracket. The subalgebra on the variables $(q, p, C, \overline{P})$ is called the **minimal subalgebra**, while the subalgebra on the variables $(\lambda, \pi, P, \overline{C})$ is called the **auxiliary subalgebra**.

A quantized algebra of observables is obtained through the (graded) Dirac correspondence. This algebra is naturally graded with respect to the ghost number defined by the self-adjoint operator

$$\mathcal{G} := \frac{1}{2}\Big(C^i\overline{P}_i - \overline{P}^i C_i + P^i\overline{C}_i - \overline{C}_i P^i\Big)\,. \tag{59.35}$$

This operator acts on observables as follows:

$$[\mathcal{G}, A] = i\hbar\,\mathsf{gh}(A)A\,. \tag{59.36}$$

BFV quantization is obtained by defining a nilpotent BRST operator $\widehat{\Omega} = \widehat{\Omega}_{\text{min}} + \widehat{\Omega}_{\text{aux}}$.

Chapter 60

Perturbation Theory

Chapter Contents

60.1 Interaction picture **1105**

60.2 Rayleigh–Schrödinger theory **1106**

60.3 Adiabatic approximation **1107**

60.3.1 Berry phase 1107

60.1 Interaction picture

Let $\widehat{H}(t) = \widehat{H}_0 + \widehat{V}(t)$ be the total Hamiltonian of a system, where $\widehat{V}(t)$ denotes the (possibly time-dependent) interaction potential. Let $|\psi(t)\rangle$ and $\widehat{O}$ denote a state and operator in the Schrödinger picture, respectively.

Definition 60.1.1 (Interaction representation). In the interaction picture, the state vector is defined as follows:

$$|\psi(t)\rangle_I := e^{\frac{i}{\hbar}\widehat{H}_0 t}|\psi(t)\rangle\,. \tag{60.1}$$

It follows, that the operators in the interaction picture are given by

$$\widehat{O}_I(t) = e^{\frac{i}{\hbar}\widehat{H}_0 t}\,\widehat{O}\,e^{-\frac{i}{\hbar}\widehat{H}_0 t}\,. \tag{60.2}$$

Formula 60.1.2 (Schrödinger equation). Using the previous formulas, the Schrödinger equation can be rewritten as follows:

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle_I = \widehat{V}_I(t)|\psi(t)\rangle_I\,. \tag{60.3}$$

The time evolution of operators in the interaction picture is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\widehat{O}_I(t) = \frac{i}{\hbar}\left[\widehat{H}_0, \widehat{O}_I(t)\right]. \tag{60.4}$$

States evolve solely based on the interaction terms and operators evolve according to the free (time-independent) Hamiltonian.

Formula 60.1.3 (Tomonaga–Schwinger equation). As in Formula 57.2.5, the evolution operator $\widehat{U}(t,t')$ satisfies the following Schrödinger-type equation in the interaction picture:

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t}\widehat{U}_I(t,t') = \widehat{V}_I(t)\widehat{U}_I(t,t')\,. \tag{60.5}$$

Although this equation is often called the Tomonaga–Schwinger equation in the literature, this is actually not correct. The Tomonaga–Schwinger equation generalizes the Schrödinger equation to allow for evolutions between arbitrary *Cauchy surfaces* (see Definition 67.1.7).

Definition 60.1.4 (Møller operator). The operators relating the (free) asymptotic states to the interacting state:

$$|\psi(0)\rangle = \widehat{\Omega}_\pm|\mathrm{in/out}\rangle\,. \tag{60.6}$$

If, for $t \longrightarrow \pm\infty$, the interaction term vanishes, there should exist a wave function $|\phi(t)\rangle$ evolving under the free Hamiltonian such that

$$\|\psi(t) - \phi(t)\| \xrightarrow{t\to\pm\infty} 0\,. \tag{60.7}$$

With $|\mathrm{in/out}\rangle := \lim_{t\to\pm\infty}|\phi(t)\rangle$, this gives

$$\widehat{\Omega}_\pm = \lim_{t\to\mp\infty}\exp\left(\tfrac{i}{\hbar}\widehat{H}(t)t\right)\exp\left(-\tfrac{i}{\hbar}\widehat{H}_0 t\right) \tag{60.8}$$

with respect to the strong topology.

Remark 60.1.5. The limit in the definition of the Møller operators will only result in a well-defined operator for infinite-dimensional spaces.

60.2 Rayleigh–Schrödinger theory

The basic assumptions of Rayleigh–Schrödinger perturbation theory are that the perturbing Hamiltonian is time independent and that the eigenfunctions of the unperturbed Hamiltonian $\widehat{H}_0$ form a complete set for the perturbed Hamiltonian.

Formula 60.2.1. The perturbed eigenfunctions and eigenvalues can be expanded in the following way, where $\lambda \in \mathbb{R}$ is assumed to be a small (perturbation) parameter:

$$\begin{aligned} |\psi_n\rangle &= \sum_{i=0}^{+\infty}\lambda^i|\psi_n^{(i)}\rangle\,, \\ E_n &= \sum_{i=0}^{+\infty}\lambda^i E_n^{(i)}\,, \end{aligned} \tag{60.9}$$

where $i \in \mathbb{N}$ is called the **order** of the perturbation.

@@ EXPAND @@

60.3 Adiabatic approximation

60.3.1 Berry phase

Theorem 60.3.1 (Adiabatic theorem). *If a perturbation is acting slowly enough such that the system can adapt its configuration at every single moment, the system will remain in the same eigenstate.*

Remark. The original formulation by *Born* and *Fock* also required the system to be *gapped*. Later, this was shown to be inessential.

Consider a system for which the adiabatic approximation is valid and assume that the state starts of as an eigenstate. The wave function is then of the form

$$|\psi(t)\rangle = C_a(t) \exp\left(-\frac{i}{\hbar} \int_{t_0}^{t} E_a(t')\, dt' \right) |\psi_a(t)\rangle \,, \tag{60.10}$$

where the $|\psi_a(t)\rangle$ is an instantaneous eigenstate of $\widehat{H}(t)$. It follows from the orthonormality of the eigenstates $|\psi_a(t)\rangle$ that the coefficient $C_a(t)$ is just a phase factor, so it can be written as

$$C_a(t) = e^{i\gamma_a(t)} \,. \tag{60.11}$$

By substituting this ansatz in the wave function, the Schödinger equation gives a differential equation for the phase factors $\gamma_a(t)$. Integration gives

$$\gamma_a(t) = i \int_{t_0}^{t} \left\langle \psi_a(t') \middle| \frac{\partial \psi_a(t')}{\partial t'} \right\rangle dt' \,. \tag{60.12}$$

Due to time evolution, the wave function accumulates a phase through the coefficient $C_a(t)$. This phase is called the **Berry phase**.

Now, try to apply a phase transformation to remove the Berry phase:

$$|\psi_a'(t)\rangle := e^{i\eta(t)} |\psi_a(t)\rangle \,. \tag{60.13}$$

Inserting this in Eq. (60.12), gives

$$\gamma_a'(t) = \gamma_a(t) - \eta(t) + \eta(t_0) \,. \tag{60.14}$$

If the system is cyclic with period $\tau \in \mathbb{R}$, then $\psi_a(t_0) = \psi_a(t_0 + \tau)$. Combining this with the equations above, gives

$$\eta(t_0 + \tau) - \eta(t_0) = 2k\pi \tag{60.15}$$

for some $k \in \mathbb{N}$. This implies that the Berry phase cannot be eliminated through a basis transformation and, hence, this phase is physically observable!

Definition 60.3.2 (Berry connection). The quantity

$$\mathbf{A}(\vec{\mathbf{x}}) := i\langle \psi_a(\vec{\mathbf{x}}) \,|\, \nabla_{\vec{\mathbf{x}}} \psi_a(\vec{\mathbf{x}}) \rangle \,, \tag{60.16}$$

where $\nabla_{\vec{\mathbf{x}}}$ denotes the gradient in phase space, is called the Berry connection (or **Berry gauge potential**). Applying Stokes' theorem to (60.12) gives

$$\gamma_a(t_0 + \tau) = \int \mathcal{B} \cdot d\vec{\mathbf{S}} \,, \tag{60.17}$$

where $\mathcal{B} = \nabla_{\vec{\mathbf{x}}} \times \mathbf{A}(\vec{\mathbf{x}})$ is called the **Berry curvature**. Although the Berry connection is gauge-dependent, the Berry curvature is gauge-invariant.

Remark 60.3.3. Using the language of differential geometry (Chapter 35), one immediately finds that the accumulated phase γ_a is simply the holonomy (Definition 35.3.36) associated with the Berry connection along the considered trajectory through phase space.

Chapter 61

Scattering Theory

Chapter Contents

61.1 Cross sections 1109

61.2 Lippman–Schwinger equations 1109

61.1 Cross sections

Formula 61.1.1 (Differential cross section).

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \frac{N(\theta,\varphi)}{F}\,, \tag{61.1}$$

where F is the incoming flux and N the detected flow rate. Because N is not defined as a flux but as a rate, the differential cross section has the dimension of area.

Formula 61.1.2 (Fermi's golden rule). The transition probability from an initial state to a final state is given by

$$\Gamma_{i\to f} = \frac{2\pi}{\hbar}|\langle f|\widehat{H}|i\rangle|^2\frac{\mathrm{d}n}{\mathrm{d}E_f}\,. \tag{61.2}$$

61.2 Lippman–Schwinger equations

In this section Hamiltonians of the form $\widehat{H} = \widehat{H}_0 + \widehat{V}$ are considered, where $\widehat{H}_0$ is the free Hamiltonian and $\widehat{V}$ the scattering potential. It will also be assumed that both the total Hamiltonian and the free Hamiltonian have the same eigenvalues (as in the previous chapter).

Formula 61.2.1 (Lippman–Schwinger equation).

$$|\psi^{(\pm)}\rangle = |\varphi\rangle + \frac{1}{E-\widehat{H}_0 \pm i\varepsilon}\widehat{V}|\psi^{(\pm)}\rangle\,, \tag{61.3}$$

where $|\varphi\rangle$ is an eigenstate of the free Hamiltonian with the same energy as $|\psi\rangle$, i.e. $\widehat{H}_0|\varphi\rangle = E|\varphi\rangle$.

Remark 61.2.2. The term $\pm i\varepsilon$ is added to the denominator because otherwise it would be singular. The term has no real physical meaning.

Formula 61.2.3 (Born series). If the Lippman–Schwinger equation is rewritten as

$$|\psi\rangle = |\varphi\rangle + \widehat{G}_0\widehat{V}|\psi\rangle\,, \tag{61.4}$$

where $\widehat{G}_0$ is the Green's operator (Definition 21.6.7) associated to $\widehat{H}_0$, one can derive the following series expansion by iterating the above expression:

$$|\psi\rangle = |\varphi\rangle + \widehat{G}_0\widehat{V}|\varphi\rangle + \left(\widehat{G}_0\widehat{V}\right)^2|\varphi\rangle + \cdots\,. \tag{61.5}$$

Convergence issues of this series can be resolved through a Borel resummation procedure (Definition 16.6.6).

Formula 61.2.4 (Born approximation). If the Born series is truncated at the first-order term in $\widehat{V}$, the Born approximation is obtained:

$$|\psi\rangle = |\varphi\rangle + \widehat{G}_0\widehat{V}|\varphi\rangle\,. \tag{61.6}$$

Part X

Statistical Mechanics & Condensed Matter Physics

Table of Contents

62 Thermodynamics **1114**
62.1 General definitions 1114
62.2 Postulates 1115
62.3 Equations of state 1116
62.4 Gases 1116

63 Statistical Mechanics **1117**
63.1 Axioms 1117
63.2 Temperature 1118
63.3 Canonical ensemble 1118
63.4 Grand canonical ensemble 1119
63.5 Energy 1120
63.6 Lattice systems 1120
63.6.1 Ising model 1120
63.7 Black-body radiation 1122

64 Material Physics **1123**
64.1 Crystals 1124
64.1.1 Analytic representation 1124
64.2 Symmetries 1124
64.3 Crystal lattice 1124
64.3.1 Bravais lattice 1125
64.3.2 Reciprocal lattice 1125
64.4 Diffraction 1125
64.4.1 Constructive interference 1125
64.4.2 Intensity of diffracted beams 1127
64.5 Alloys 1127
64.6 Lattice defects 1128
64.7 Electrical properties 1129
64.7.1 Charge carriers 1129
64.7.2 Band structure 1129

64.7.3 Ferroelectricity 1130
64.8 Magnetic properties 1131
64.8.1 Paramagnetism 1131
64.8.2 Ferromagnetism 1132
64.8.3 Antiferromagnetism 1132
64.8.4 Ferrimagnetism 1133

65 Tensor Networks **1134**
65.1 Matrix Product States 1134
65.1.1 Finite-dimensional lattices 1134
65.1.2 Injectivity 1136
65.1.3 Gauge freedom and canonical forms 1136
65.1.4 Translation-invariant states 1137
65.2 Matrix product operators 1137
65.2.1 MPO-injectivity 1138
65.2.2 MPO-symmetries for SPT phases 1139

Chapter 62

Thermodynamics

Chapter Contents

62.1 General definitions 1114

62.2 Postulates 1115

62.3 Equations of state 1116

62.4 Gases 1116

62.1 General definitions

Definition 62.1.1 (System). The part of space that is of interest.

Definition 62.1.2 (Environment). The complement of the system in space. More specifically, this is the part of space that has a potential influence on the system.

Definition 62.1.3 (Thermodynamic coordinate). Macroscopical variable that describes the system. These are also called **state variables**.

Definition 62.1.4 (Intensive coordinate). Coordinate that does not depend on the system size. The opposite notion is an **extensive coordinate**.

Definition 62.1.5 (Thermodynamic equilibrium). A system is said to be in thermodynamic equilibrium if it is simultaneously in thermal, mechanical and chemical equilibrium. In this case, it is fully described by a set of constant thermodynamic coordinates.

Property 62.1.6 (Uniformity). During thermodynamic equilibrium, all intensive coordinates are uniform throughout the system.

Definition 62.1.7 (Isolated system). A system that cannot interact with its environment (e.g. due to the presence of impenetrable walls).

Definition 62.1.8 (Diathermic wall). A wall that only allows heat transfer. This should be distinguished from an **adiabatic wall**, i.e. a wall that does not allow any transfer of heat.

Definition 62.1.9 (Heat bath). A heat bath or **thermal reservoir** is a thermodynamic system (often part of the environment) for which the temperature remains constant during the exchange of heat, i.e. it has a virtually infinite heat capacity.

Definition 62.1.10 (Open system). A system that is allowed to interact with its environment.

Definition 62.1.11 (Quasistatic process). A sequence of equilibrium states separated by infinitesimal changes.

Definition 62.1.12 (Path). The sequence of equilibrium states in a thermodynamic process.

62.2 Postulates

Axiom 62.1 (Zeroth law). If two objects are in thermal equilibrium with a third object, they are also in thermal equilibrium with each other.

Axiom 62.2 (First law). The change in internal energy is given by

$$\Delta U = Q + W\,, \tag{62.1}$$

or, infinitesimally, by

$$\mathrm{d}U = \delta Q + \delta W\,, \tag{62.2}$$

where W denotes the work done on the system and Q denotes the heat that was extracted from the environment.

Remark. The operator δ in the heat and work differentials implies that these are 'inexact' differentials, i.e. they are not the differential of functions of the thermodynamic coordinates alone. See Section 34.4 for more information on differential forms.

Axiom 62.3 (Kelvin–Planck formulation of the second law). No machine can completely transfer an amount of heat into work.

Axiom 62.4 (Clausius formulation of the second law). Heat cannot be passed from a cooler object to a warmer object without performing work.

Formula 62.2.1 (Clausius's inequality). In differential form, the inequality reads as

$$\frac{\delta Q}{T} \geq 0\,. \tag{62.3}$$

@@ COMPLETE THIS STATEMENT (WHAT INEQUALITY?) @@

Axiom 62.5 (Third law). No process can reach absolute zero through a finite sequence of operations.

62.3 Equations of state

Property 62.3.1 (PV-system). Consider a system described by pressure P and volume V. The first law becomes

$$\delta Q = \mathrm{d}U + P\mathrm{d}V\,. \tag{62.4}$$

This formula defines a Pfaffian system (Property 34.6.8). Since the state space is only two-dimensional, it is not too hard to see that it is integrable, i.e. there exists an integrating factor T and a one-form $\mathrm{d}S$ such that

$$T\mathrm{d}S = \mathrm{d}U + P\mathrm{d}V\,. \tag{62.5}$$

T and S represent the temperature and entropy of the system, respectively. If the heat exchange is zero, the entropy is conserved.

62.4 Gases

Formula 62.4.1 (Ideal gas law).

$$PV = nRT\,, \tag{62.6}$$

where R is the **ideal gas constant** $R \approx 8.314\frac{\mathrm{J}}{\mathrm{K\,mol}}$.

Chapter 63

Statistical Mechanics

Chapter Contents

63.1 Axioms 1117
63.2 Temperature 1118
63.3 Canonical ensemble 1118
63.4 Grand canonical ensemble 1119
63.5 Energy 1120
63.6 Lattice systems 1120
63.6.1 Ising model 1120
63.7 Black-body radiation 1122

63.1 Axioms

Axiom 63.1 (Ergodic principle). All microstates corresponding to the same macrostate are equiprobable.

Axiom 63.2 (Boltzmann formula). The entropy of a system is given by

$$S = k \ln \Omega(E, V, N, \alpha), \tag{63.1}$$

where Ω denotes the number of microstates corresponding to the system with energy E and any other state variables (these are denoted by α). In general, S will be the Shannon entropy (Definition 45.5.2).

63.2 Temperature

Definition 63.2.1 (Temperature). The temperature of a system in contact with a heat bath is defined as follows:

$$T := \left(\frac{\partial E}{\partial S}\right)_V . \tag{63.2}$$

63.3 Canonical ensemble

Formula 63.3.1 (Partition function). The partition function for discrete systems is defined as

$$Z(T) := \sum_i g_i e^{-\beta E_i} , \tag{63.3}$$

where the g_i are the **degeneracies** of the states. The analogue for continuous systems is

$$Z(T) := \int_{\mathbb{R}^+} \Omega(E, V, N) e^{-\beta E} \, dE . \tag{63.4}$$

Formula 63.3.2. Consider a system of N indistinguishable, noninteracting particles. Let ε_i be the energy associated with the ith energy level and let g_i be its degeneracy. The probability p_i of finding a particle in the ith energy level is given by

$$p_i = \frac{g_i e^{-\beta E_i}}{Z} . \tag{63.5}$$

Definition 63.3.3 (Helmholtz free energy). The Helmholtz free energy is defined as follows:

$$F := -k_B T \ln Z . \tag{63.6}$$

For the canonical ensemble, it can be shown that this is equal to a Legendre transformation of the energy:

$$F = E - TS . \tag{63.7}$$

One can also obtain the Helmholtz free energy as a different Legendre transform using the ideas of information theory (Chapter 47). There, it was shown that the convex potentials associated to exponential families were related to the free energy. If Eq. (47.26) is compared to the above one, it can be seen that

$$\psi = -\beta F . \tag{63.8}$$

This quantity is sometimes called the **(Helmholtz) free entropy** or **Massieu potential** to distinguish it from the (Helmholtz) free energy. It was also shown that the associated dual coordinates are the expectation values, in this case the internal energy (up to

a sign), and the dual potential was equal to the (negative) Shannon entropy. Putting this together gives:

$$\begin{aligned} \eta &= \beta \frac{\partial \psi}{\partial \beta} - \psi \\ \overset{\text{def.}}{\Longleftrightarrow} \quad -S &= -\beta U + \beta F \\ \Longleftrightarrow \quad F &= U - TS_B \,, \end{aligned} \tag{63.9}$$

where the relation between the Boltzmann entropy S_B and the Shannon entropy S was used.

63.4 Grand canonical ensemble

Formula 63.4.1 (Grand canonical partition function). The partition function of the ith energy level is given by

$$\mathcal{Z}_i := \sum_{n_k} e^{\beta n_k (\mu - \varepsilon_i)} \,, \tag{63.10}$$

where $n_k \in \mathbb{N}$ denotes the number of particles in the ith state and μ is the *chemical potential*, i.e. the amount of energy ot costs to create a new particle. The grand canonical partition function is given by

$$\mathcal{Z} := \prod_i \mathcal{Z}_i = \sum_{n_k, \varepsilon_i} e^{\beta n_k (\mu - \varepsilon_i)} \,. \tag{63.11}$$

Remark 63.4.2. In the case of fermions, i.e. $n_k \in \{0, 1\}$, this formula reduces to

$$\mathcal{Z} = e^{\beta \mu N} Z \,, \tag{63.12}$$

with $N \in \mathbb{N}$ the total particle number.

Definition 63.4.3 (Fugacity).

$$z := e^{\mu N} \tag{63.13}$$

Formula 63.4.4 (Quantum mechanics). For quantum-mechanical systems, one can rewrite the partition function as follows:

$$\mathcal{Z} = \operatorname{tr} \exp\left(-\beta\left(\widehat{H} - \mu \widehat{N}\right)\right) . \tag{63.14}$$

This reduces to the above expressions when working in the single-particle eigenbasis (this is only possible for free theories).

63.5 Energy

Theorem 63.5.1 (Virial theorem).

$$\langle E_{\text{kin}} \rangle = -\frac{1}{2} \sum_k \langle \vec{\mathbf{r}}_k \cdot \vec{\mathbf{F}}_k \rangle \tag{63.15}$$

Corollary 63.5.2. For potentials of the form $V(r) = ar^{-n}$, this becomes

$$2\langle E_{\text{kin}} \rangle = -n\langle V \rangle \,. \tag{63.16}$$

Theorem 63.5.3 (Equipartition theorem). *Let q^i be generalized coordinates.*

$$\left\langle q^k \frac{\partial H}{\partial q^l} \right\rangle = k_B T \delta_{kl} \tag{63.17}$$

Corollary 63.5.4. For quadratic Hamiltonians, this can be rewritten using Euler's theorem for homogeneous functions (Theorem 16.5.14):

$$\langle T \rangle = \frac{1}{2} k_B T \,. \tag{63.18}$$

63.6 Lattice systems

63.6.1 Ising model

The most well-known and widespread lattice system in statistical mechanics is the Ising model. It not only characterizes (ferro)magnets and spin glasses (approximately), but it can also be used to study, e.g., brain signals, social networks or forest growth.

Formula 63.6.1 (Ising Hamiltonian). Let Λ be a lattice in $\mathbb{R}^n$ and consider a collection of spins $\{S_i\}_{i\in\Lambda}$ on this lattice. If the external magnetic field is denoted by $h \in \mathbb{R}$, the Ising Hamiltonian has the following form:

$$H := -\sum_{\langle i,j \rangle \in \Lambda} J_{ij} S_i S_j - h \sum_{i \in \Lambda} S_i \,, \tag{63.19}$$

where $\langle i, j \rangle$ denotes a pair of neighbouring indices. The constants J_{ij} give the interaction strength between the neighbouring spins.[1]

Two general cases can be distinguished. If all $J_{ij} > 0$, the system is said to be **ferromagnetic**, while if all $J_{ij} > 0$, the system is said to be **antiferromagnetic**. In the case of ferromagnetism and a vanishing external field, i.e. $h = 0$, the ground state of the Ising model is degenerate. All spins are alligned, i.e. $S_i S_j = 1$, but the orientation has no influence on the energy. However, when an external field is applied (even a very

[1] One can generalize this Hamiltonian to also allow for site-dependent external fields. However, this will not be of relevance here.

small one), the second term in the Hamiltonian induces a preferential orientation and the degeneracy is lifted.

When the temperature is nonzero, thermal fluctuations will start to randomly flip spins and the equilibrium state will deviate from the perfectly ordered state. The question then becomes whether there exists a critical temperature T_c at which the system changes between a disordered state and an ordered state. For a single spatial dimension, the solution was found by *Ising*.

Property 63.6.2 ($d = 1$)**.** For a one-dimensional lattice, the Ising model does not undergo a phase transition at a nonzero temperature. However, at $T = 0$, a second-order phase transition occurs.

The easiest way to prove this is through the transfer operator method, where the partition function is written as a product of transfer operators. Here, this is shown for a periodic chain with uniform interaction strength $J_{ij} = J$ for all $i, j \in \Lambda$.

Proof.

$$\begin{aligned} Z &= \sum_{\{S_1,\dots,S_L\}} \exp\left(-\sum_{i=1}^{L-1} JS_iS_{i+1} - hS_i\right) \\ &= \prod_{i=1}^{L-1} \sum_{S_i,S_{i+1}} \exp\left(-JS_iS_{i+1} - hS_i\right) \\ &= \prod_{i=1}^{L-1} P_{i,i+1} \\ &= \operatorname{tr}(P^L)\,. \end{aligned}$$

The matrix $P_{i,i+1}$ is defined as

$$P_{i,i+1} := \begin{pmatrix} e^{\beta(J+h)} & e^{-\beta J} \\ e^{-\beta J} & e^{\beta(J-h)} \end{pmatrix}$$

for all $i \in \{1, \dots, L\}$. The result then becomes $Z = \lambda_1^L + \lambda_2^L$, where λ_1 and λ_2 are the eigenvalues of P. In the thermodynamic limit, i.e. $L \gg 1$, only the largest eigenvalue will contribute:

$$\lambda_1 = e^{\beta J}\cosh(\beta h) + \sqrt{e^{2\beta J}\sinh^2(\beta h) + e^{-2\beta J}}\,.$$

So, for $L \gg 1$, the free energy density becomes

$$f(T,J,h) = -\frac{k_BT}{L}\ln(Z) = -k_BT\left(e^{\beta J}\cosh(\beta h) + \sqrt{e^{2\beta J}\sinh^2(\beta h) + e^{-2\beta J}}\right).$$

@@ COMPLETE @@ □

For higher-dimensional lattices, a phase-transition is found at finite temperature. This was first proven by *Onsager et al.* and required considerably more work than in the one-dimensional setting.

63.7 Black-body radiation

Formula 63.7.1 (Planck's law).

$$B_\nu(\nu, T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{\beta h\nu} - 1} \tag{63.20}$$

Formula 63.7.2 (Wien's displacement law).

$$\lambda_{\max} T = b\,, \tag{63.21}$$

where the constant $b = 2.8977729(17)\mathrm{e}{-3}$ Km is called **Wien's displacement constant**.

Chapter 64

Material Physics

Chapter Contents

64.1 Crystals **1124**
64.1.1 Analytic representation 1124
64.2 Symmetries **1124**
64.3 Crystal lattice **1124**
64.3.1 Bravais lattice 1125
64.3.2 Reciprocal lattice 1125
64.4 Diffraction **1125**
64.4.1 Constructive interference 1125
64.4.2 Intensity of diffracted beams 1127
64.5 Alloys **1127**
64.6 Lattice defects **1128**
64.7 Electrical properties **1129**
64.7.1 Charge carriers 1129
64.7.2 Band structure 1129
64.7.3 Ferroelectricity 1130
64.8 Magnetic properties **1131**
64.8.1 Paramagnetism 1131
64.8.2 Ferromagnetism 1132
64.8.3 Antiferromagnetism 1132
64.8.4 Ferrimagnetism 1133

64.1 Crystals

Theorem 64.1.1 (Steno's law). *The angles between crystal faces of the same type are constant and do not depend on the total shape of the crystal.*

Definition 64.1.2 (Zone). The collection of faces parallel to a given axis. The axis itself is called the **zone axis**.

64.1.1 Analytic representation

Definition 64.1.3 (Miller indices). Let a, b, c be the lengths of the (not necessarily orthogonal) basis vectors of the crystal lattice. The lattice plane intersecting the axes at $\left(\frac{a}{h}, \frac{b}{k}, \frac{c}{k}\right)$ is denoted by the Miller indices $(h\ k\ l)$.

Notation 64.1.4. Negative numbers are often written as $\bar{a}$ instead of $-a$.

Formula 64.1.5 (Axes). Let a, b, c denote the lengths of the basis vectors. The axis formed by the intersection of the planes $(h_1\ k_1\ l_1)$ and $(h_2\ k_2\ l_2)$ is denoted by $[u\ v\ w]$. Its direction is determined by the point (au, bv, cw), where

$$u = \begin{vmatrix} k_1 & l_1 \\ k_2 & l_2 \end{vmatrix} \qquad v = \begin{vmatrix} l_1 & h_1 \\ l_2 & h_2 \end{vmatrix} \qquad w = \begin{vmatrix} h_1 & k_1 \\ h_2 & k_2 \end{vmatrix}. \tag{64.1}$$

Theorem 64.1.6 (Hauy's law of rational indices). *The Miller indices of every natural face of a crystal will always have rational proportions.*

64.2 Symmetries

Definition 64.2.1 (Equivalent planes/axes). When applying crystal symmetries, it often happens that a set of equivalent planes and axes is obtained. These equivalence classes are denoted by $\{h\ k\ l\}$ and $\langle h\ k\ l\rangle$, respectively.

Property 64.2.2 (Rotational symmetry). Only $1, 2, 3, 4$ and 6-fold rotational symmetries can occur.

64.3 Crystal lattice

Formula 64.3.1. For an orthogonal crystal lattice, the distance between planes of the family $(h\ k\ l)$ is given by

$$d_{hkl} = \frac{1}{\sqrt{\left(\frac{h}{a}\right)^2 + \left(\frac{k}{b}\right)^2 + \left(\frac{l}{c}\right)^2}}. \tag{64.2}$$

64.3.1 Bravais lattice

Definition 64.3.2 (Bravais lattice). A crystal lattice generated by a point group symmetry. There are 14 different Bravais lattices in 3 dimensions. These are the only possible ways to place (infinitely) many points in 3D space by applying symmetry operations consistent with the given point group.

Definition 64.3.3 (Wigner–Seitz cell). The part of space consisting of all points closer to a given lattice point than to any other.

Theorem 64.3.4 (Neumann's principle). *The symmetry elements of the physical properties of a crystal should at least contain those of the point group of the crystal.*

64.3.2 Reciprocal lattice

Formula 64.3.5 (Reciprocal basis vectors). The reciprocal lattice corresponding to a Bravais lattice with primitive basis $\{\vec{\mathbf{a}}, \vec{\mathbf{b}}, \vec{\mathbf{c}}\}$ is defined by the following reciprocal basis vectors:

$$\vec{\mathbf{a}}^* := 2\pi \frac{\vec{\mathbf{b}} \times \vec{\mathbf{c}}}{\vec{\mathbf{a}} \cdot (\vec{\mathbf{b}} \times \vec{\mathbf{c}})} . \tag{64.3}$$

The vectors $\vec{\mathbf{b}}^*$ and $\vec{\mathbf{c}}^*$ are obtained by cyclic permutation of (a, b, c). These vectors satisfy the relations

$$\begin{aligned} \vec{\mathbf{a}} \cdot \vec{\mathbf{a}}^* &= 2\pi , \\ \vec{\mathbf{b}} \cdot \vec{\mathbf{b}}^* &= 2\pi , \\ \vec{\mathbf{c}} \cdot \vec{\mathbf{c}}^* &= 2\pi . \end{aligned} \tag{64.4}$$

Notation 64.3.6 (Reciprocal lattice vector). The reciprocal lattice vector $\vec{\mathbf{r}}^*_{hkl}$ is defined as follows:

$$\vec{\mathbf{r}}^*_{hkl} := h\vec{\mathbf{a}}^* + k\vec{\mathbf{b}}^* + l\vec{\mathbf{c}}^* . \tag{64.5}$$

Property 64.3.7. The reciprocal lattice vector $\vec{\mathbf{r}}^*_{hkl}$ has the following properties:

- $\vec{\mathbf{r}}^*_{hkl}$ is perpendicular to the family of planes $(h\ k\ l)$ of the direct lattice, and
- $\|\vec{\mathbf{r}}^*_{hkl}\| = \frac{2\pi n}{d_{hkl}}$.

64.4 Diffraction

64.4.1 Constructive interference

Formula 64.4.1 (Laue conditions). Suppose that an incident beam makes angles α_0, β_0 and γ_0 with the lattice axes. A diffracted beam, making angles α, β and γ with the axes,

will be observed if the following conditions are satisfied:

$$\begin{aligned} a(\cos\alpha - \cos\alpha_0) &= h\lambda\,, \\ b(\cos\beta - \cos\beta_0) &= k\lambda\,, \\ c(\cos\gamma - \cos\gamma_0) &= l\lambda\,. \end{aligned} \tag{64.6}$$

If these conditions have been met, a diffracted beam of order *hkl* will be observed.

Remark 64.4.2. Further conditions can be imposed on the angles, such as the Pythagorean formula for orthogonal axes. This has the consequence that the only two possible ways to obtain a diffraction pattern are:

- a fixed crystal and a polychromatic beam, or
- a rotating crystal and a monochromatic beam.

Formula 64.4.3 (Vectorial Laue conditions). Let $\vec{\mathbf{k}}_0, \vec{\mathbf{k}}$ denote the wave vectors of the incident and diffracted beams respectively. The Laue conditions can be reformulated in the following way:

$$\vec{\mathbf{k}} - \vec{\mathbf{k}}_0 = \vec{\mathbf{r}}^*_{hkl}\,. \tag{64.7}$$

Formula 64.4.4 (Bragg's law). Another equivalent formulation of the Laue conditions is given by the following formula:

$$2d_{hkl}\sin\theta = n\lambda\,, \tag{64.8}$$

where

- λ is the wavelength of the incoming beam,
- θ is the **Bragg angle**, and
- d_{hkl} is the distance between neighbouring planes.

Remark 64.4.5. The angle between the incident and diffracted beams is 2θ.

A simple construction to determine where the maxima of Bragg diffraction will occur is the following.

Construction 64.4.6 (Ewald sphere). Put the origin of the reciprocal lattice at the tip of the incident wave vector $\vec{\mathbf{k}}_i$. Construct a sphere with radius $\frac{2\pi}{\lambda}$ centred on the initial point of $\vec{\mathbf{k}}_i$. All points on the sphere that coincide with a reciprocal lattice point satisfy the vectorial Laue condition 64.4.3. Therefore, Bragg diffraction will occur in the direction of all the intersections of the Ewald sphere and the reciprocal lattice.

64.4.2 Intensity of diffracted beams

Property 64.4.7 (Systematic extinctions). Every particle in the motive emits its own waves. These waves will interfere and some will cancel out. This leads to the absence of certain diffraction spots, the systematic extinctions.

Definition 64.4.8 (Atomic scattering factor). The waves produced by the individual electrons of an atom, which can have a different phase, can be combined into a resulting wave. The amplitude of this wave is called the atomic scattering factor.

Definition 64.4.9 (Structure factor). The waves coming from the individual atoms in the motive can also be combined into a resulting wave (again, taking into account the different phases). The amplitude of this wave is called the structure factor. It is given by

$$F(hkl) = \sum_j f_j \exp\big(2\pi i(hx_j + ky_j + lz_j)\big)\,, \tag{64.9}$$

where f_j is the atomic scattering factor of the jth atom in the motive.

Example 64.4.10. An important example of systematic extinctions is the structure factor of an FCC or BCC lattice. If $h + k + l$ is odd, $F(hkl) = 0$ for a BCC lattice. If h, k and l are not all even or all odd, $F(hkl) = 0$ for an FCC lattice.

Definition 64.4.11 (Laue indices). Higher-order diffractions can be rewritten as a first-order diffraction in the following way:

$$2d_{nhnknl} \sin\theta = \lambda \qquad \text{with} \qquad d_{nhnknl} = \frac{d_{hkl}}{n}\,. \tag{64.10}$$

The interpretation of Bragg's law as diffraction being a reflection in the lattice plane $(h\,k\,l)$ implies that one can introduce the (fictitious) plane with indices $(nh\,nk\,nl)$. These indices are called Laue indices.

Remark. In contrast to Miller indices, which cannot possess common factors, the Laue indices obviously can.

64.5 Alloys

Property 64.5.1 (Hume-Rothery conditions). An element can be dissolved in a metal (forming a solid solution) if the following conditions are met:

1. the difference between the atomic radii is $\leq 15\%$,
2. the crystal structures are the same,
3. the elements have a similar electronegativity, and
4. the valence is the same.

64.6 Lattice defects

Definition 64.6.1 (Interstitial). An atom placed at a position that is not a lattice point.

Definition 64.6.2 (Vacancy). A lattice point where an atom is missing. This is also called a **Schottky defect**.

Formula 64.6.3 (Concentration of Schottky defects). Let $N, n \in \mathbb{N}$ denote the number of lattice points and vacancies, respectively. The following relation gives the temperature dependence of Schottky defects:

$$\frac{n}{n+N} = e^{-E_v/kT}\,, \tag{64.11}$$

where T denotes the temperature and E_v the energy needed to create a vacancy. A similar relation holds for interstitials.

Proof. Let E_v be the energy needed to remove a particle from its lattice point and move it to the surface. All surface effects will be neglected and it will be assumed that the the energy E_v is independent of the distance to the surface.

The total energy of all vacancies is then given by $E = nE_v$. The number of possible microstates is

$$\Omega = \frac{(N+n)!}{N!\,n!}\,, \tag{64.12}$$

where the fact that the removal of n particles creates n more lattice points at the surface was used. Using Boltzmann's entropy formula (63.2) and Stirling's formula (16.6.13) one obtains

$$S(N,n) = k\ln\Omega = k\big[(N+n)\ln(N+n) - n\ln n - N\ln N\big]\,. \tag{64.13}$$

Using Definition 63.2.1 of the temperature as the derivative of the energy gives

$$\frac{1}{T} = \left(\frac{\partial S}{\partial E}\right)_{N,V} = \frac{\mathrm{d}S}{\mathrm{d}n}\frac{\mathrm{d}n}{\mathrm{d}E} = \frac{k}{E_v}\ln\frac{N+n}{n} \tag{64.14}$$

or

$$\frac{n}{N+n} = \exp\left(-\frac{E_v}{kT}\right). \tag{64.15}$$

The density of Frenkel pairs can be derived analogously. □

Definition 64.6.4 (Frenkel pair). An atom displaced from a lattice point to an interstitial location, thereby creating a vacancy-interstitial pair.

Formula 64.6.5 (Concentration of Frenkel pairs). Let $n_i \in \mathbb{N}$ denote the number of atoms displaced from the bulk of the lattice to any of $N_i \in \mathbb{N}$ possible interstitial posi-

tions, thereby creating n_i vacancies. The following relation holds:

$$\frac{n_i}{\sqrt{NN_i}} = e^{-E_{\text{fr}}/2kT}, \tag{64.16}$$

where E_{fr} denotes the energy needed to create a Frenkel pair.

Remark 64.6.6. In compounds, the number of vacancies can be much higher than in mono-atomic lattices.

Remark 64.6.7. The existence of these defects creates the possibility of diffusion.

64.7 Electrical properties

64.7.1 Charge carriers

Formula 64.7.1 (Conductivity). Definition 55.1.11 can be modified to account for both positive and negative charge carriers:

$$\sigma := n_n q_n \mu_n + n_p q_p \mu_p \,. \tag{64.17}$$

Remark. The difference between the concentration of positive and negative charge carriers can differ by orders of magnitude (up to 20) across different materials.

64.7.2 Band structure

Definition 64.7.2 (Valence band). The energy band corresponding to the outermost (partially) filled atomic orbital.

Definition 64.7.3 (Conduction band). The first unfilled energy band.

Definition 64.7.4 (Band gap). The energy difference between the valence and conduction bands (if they do not overlap). It is the energy zone where no electron states can exist.[1]

Definition 64.7.5 (Fermi level). The energy level having a 50% chance of being occupied at thermodynamic equilibrium.

Formula 64.7.6 (Fermi function). The following distribution gives the probability of a state with energy E_i being occupied by an electron:

$$f(E_i) = \frac{1}{e^{\beta(E_i - E_f)} + 1}, \tag{64.18}$$

where E_f is the Fermi level as defined above.

[1] For a basic derivation, see Bransden and Joachain (2000).

Formula 64.7.7. Let $n \in \mathbb{N}$ denote the charge carrier density as before. The following temperature dependence can be found:

$$n \sim e^{-\beta E_g/2}, \tag{64.19}$$

where E_g is the band gap. This formula can be directly derived from the Fermi function by noting that, for intrinsic semiconductors, the Fermi level sits in the middle of the band gap, i.e. $E_c - E_f = E_g/2$, and that, for most semiconductors, $E_g \gg kT$.

Definition 64.7.8 (Doping). Intentionally introducing impurities to modify the (electrical) properties.

Definition 64.7.9 (Acceptor). A group-III element added to create an excess of holes in the valence band. The resulting semiconductor is called a **p-type semiconductor**.

Definition 64.7.10 (Donor). A group-IV element added to create an excess of electrons in the valence band. The resulting semiconductor is called an **n-type semiconductor**.

64.7.3 Ferroelectricity

Some materials can exhibit phase transitions between a paraelectric and ferroelectric state. Paraelectric materials have the property that the polarization $\vec{\mathbf{P}}$ and the electric field $\vec{\mathbf{E}}$ are proportional. Ferroelectric materials have the property that they exhibit permanent polarization, even in the absence of an electric field. This permanent behaviour is the result of symmetry breaking. The ions in the lattice have been displaced from their 'central' positions, which induces a permanent dipole moment.

The temperature at which this phase transition occurs is called the **ferroelectric Curie temperature**. Above this temperature, the material will behave as a paraelectric material.

Remark 64.7.11. Ferroelectricity can only occur in crystals with unit cells that do not have a center of symmetry. This would rule out the possibility of having the asymmetry needed for the dipole moment.

Definition 64.7.12 (Saturation polarization). The maximum polarization obtained by a ferroelectric material. It it obtained when the *domain formation* reaches a maximum.

Definition 64.7.13 (Remanent polarization). The residual polarization of the material when the external electric field is turned off.

Definition 64.7.14 (Coercive field). The electric field needed to cancel out the *remanent* polarization.

Definition 64.7.15 (Piezoelectricity). Materials that obtain a polarization when exposed to mechanical stress are called piezoelectric materials.

Remark 64.7.16. All ferroelectric materials are piezoelectric, but the converse is not true. Moreover, all crystals without a center of symmetry are piezoelectric. This property is, however, only a necessary (and not a sufficient) condition for ferroelectricity.

Example 64.7.17 (Transducer). A device that converts electrical energy to mechanical energy (and vice versa).

64.8 Magnetic properties

Definition 64.8.1 (Diamagnetism). In diamagnetic materials, the magnetization is oriented oppositely to the applied field, so $B < H$. The susceptibility is small, negative and independent of the temperature.

Remark 64.8.2. All materials exhibit diamagnetic behaviour.

Definition 64.8.3 (Paramagnetism). The susceptibility is small, positive and inversely proportional to the temperature.

Definition 64.8.4 (Ferromagnetism). Spontaneous magnetization can occur. The susceptibility is large and dependent on the applied field and temperature. Above a certain temperature, the **ferromagnetic Curie temperature**, the materials will behave as if they were only paramagnetic.

64.8.1 Paramagnetism

Formula 64.8.5 (Curie's law). If the interactions between the particles can be neglected, the following relation is obtained:

$$\chi = \frac{C}{T}. \tag{64.20}$$

Materials that satisfy this law are called **ideal paramagnetics**.

Formula 64.8.6 (Curie–Weiss law). If the interactions between particles cannot be neglected, a correction of Curie's law is required:

$$\chi = \frac{C}{T - \theta}, \tag{64.21}$$

where $\theta = CN_W$ with N_W the **Weiss-constant**. This deviation from Curie's law is due to the intermolecular interactions that induce an internal magnetic field $H_m = N_W M$.

Formula 64.8.7 (Brillouin function).

$$B_J(y) := \frac{2J+1}{2J} \coth\left(\frac{2J+1}{2J} y\right) - \frac{1}{2J} \coth\left(\frac{y}{2J}\right), \tag{64.22}$$

where $y := \frac{g\mu_B JB}{k_B T}$.

Remark 64.8.8. Because $\lim_{y\to\infty} \coth(y) = 1$, one obtains:

$$T \longrightarrow 0 \implies M = \lim_{y\to\infty} Ng\mu_B J B_J(y) = Ng\mu_B J\,. \tag{64.23}$$

This value is called the **absolute saturation magnetization**.

64.8.2 Ferromagnetism

Ferromagnetics are materials that have strong internal interactions that lead to large scale (with respect to the lattice constant) parallel ordering of the atomic magnetic (dipole) moments. This also leads to the spontaneous magnetization of the material and, consequently, to a nonzero total dipole moment. In reality, however, ferromagnetic materials do not always spontaneously possess a magnetic moment in the absence of an external field. Still, when stimulated by a small external field, they will display a magnetic moment much larger than paramagnetic materials would.

Definition 64.8.9 (Domain). This behaviour is explained by the existence of **Weiss domains**. These are spontaneously magnetized regions in a magnetic material. The total dipole moment is the sum of the moments of the individual domains. If not all the domains have a parallel orientation, the total dipole moment can be 0. A small external field is, however, sufficient to change the domain orientation and produce a large total magnetization.

Definition 64.8.10 (Bloch walls). A wall between two magnetic domains.

Definition 64.8.11 (Ferromagnetic Curie temperature). Above this temperature, the material loses its ferromagnetic properties and it becomes a paramagnetic material following the Curie–Weiss law.

Remark 64.8.12. For ferromagnetic (and ferrimagnetic) materials, it is impossible to define a magnetic susceptibility, because the magnetization is nonzero even in the absence of a magnetic field.[2] Above the critical temperature, it is, however, possible to define a susceptibility, because the materials become paramagnetic in this region.

64.8.3 Antiferromagnetism

When the domains in a magnetic material have an antiparallel order (whenever this is energetically more favourable), the total dipole moment will be small. If the temperature rises, the thermal agitation will disturb the orientation of the domains and the magnetic susceptibility will increase.

Definition 64.8.13 (Néel temperature). At the Néel temperature, the susceptibility will reach a maximum. Above this temperature, the material will become paramagnetic,

[2] This can be seen in Eq. (55.24): $M = \chi H$. The susceptibility should be infinite.

satisfying the following formula:

$$\chi = \frac{C}{T + \theta} \,. \tag{64.24}$$

This resembles a generalization of the Curie–Weiss law with a negative and, therefore, virtual critical temperature.

64.8.4 Ferrimagnetism

Materials that are not completely ferromagnetic nor antiferromagnetic, due to an imbalance between the sublattices, will have a nonzero dipole moment even in the absence of an external field. The magnitude of this moment will, however, be smaller than that of a ferromagnetic material. These materials are called **ferrimagnetic materials**.

Formula 64.8.14 (Néel hyperbola). Above the Néel temperature, it is possible to define a susceptibility given by

$$\frac{1}{\chi} = \frac{T}{C} - \frac{1}{\chi_0} - \frac{\sigma}{T - \theta'} \,. \tag{64.25}$$

@@ EXPLAIN PARAMETERS @@

Chapter 65

Tensor Networks

The main references for the sections on *topological order* are Williamson et al. (2016); Şahinoğlu et al. (2021).

Chapter Contents

65.1 Matrix Product States **1134**
65.1.1 Finite-dimensional lattices 1134
65.1.2 Injectivity 1136
65.1.3 Gauge freedom and canonical forms 1136
65.1.4 Translation-invariant states 1137
65.2 Matrix product operators **1137**
65.2.1 MPO-injectivity 1138
65.2.2 MPO-symmetries for SPT phases 1139

65.1 Matrix Product States

65.1.1 Finite-dimensional lattices

Definition 65.1.1 (Matrix product state). Let $\mathcal{H}_n$ be the local Hilbert spaces of dimension $d_n \in \mathbb{N}$, where $n \in \{1, \dots, N\}$. A state $|\psi\rangle$ in the total Hilbert space $\bigotimes_i \mathcal{H}_i$ is called a matrix-product state (MPS) with periodic boundary conditions if there exist matrices $A^{i_n}(n) \in \mathcal{L}(\mathbb{C}^{D_n}, \mathbb{C}^{D_{n-1}})$ with $i_n \leq d_n$ such that

$$|\psi\rangle = \sum_{\{i_k\}} \mathrm{tr}\left(\prod_n^N A^{i_n}(n)\right) |i_1 \dots i_N\rangle \,. \tag{65.1}$$

For each lattice site $n \in \mathbb{Z}$, the set of matrices $\{A^{i_n}_{\alpha\beta}(n)\}$ can be regarded as forming a rank-3 tensor. The periodic boundary condition requires that $D_0 = D_N$, otherwise

the trace would be ill-defined. Different boundary conditions can be implemented by inserting an additional factor X at the end of the trace.

Remark 65.1.2 (Physical and virtual spaces). For each 'physical' index i_n, one can regard the matrix $A^{i_n}(n)$ as a linear map between 'virtual' (or **ancilla**) spaces $\mathbb{C}^{D_n}$.

Formula 65.1.3 (MPS projector). Consider an MPS defined by tensors $\{A(n)\}_{n\leq N}$. The associated MPS projector is defined as follows:

$$\mathcal{P}_A := \sum_{i,\alpha,\beta} A^i_{\alpha\beta}(n)|i\rangle\langle\alpha\beta| \,. \tag{65.2}$$

Formula 65.1.4 (Transfer operator). Given MPS tensors $\{A(n)\}_{n\leq N}$, one can define a transfer operator:

$$\mathbb{E}(n) := \sum_{i=1}^{d_n} A^i(n) \otimes \overline{A^i}(n) \,. \tag{65.3}$$

Formula 65.1.5 (Superoperator). More generally, one can define for every local observable $\widehat{O}_n$ a superoperator in $\mathcal{L}(\mathbb{C}^{D_n} \otimes \overline{\mathbb{C}^{D_n}}, \mathbb{C}^{D_{n-1}} \otimes \overline{\mathbb{C}^{D_{n-1}}})$:

$$\mathbb{E}_{O_n}(n) := \sum_{i,i'=1}^{d_n} \langle i \mid \widehat{O}_n \mid i'\rangle A^{i'}(n) \otimes \overline{A^i}(n) \,. \tag{65.4}$$

Comparing with the definition of the transfer operator, it can be seen that $\mathbb{E}$ is given by the superoperator associated to the unit operator. Given two sets of MPS tensors $\{A(n)\}, \{B(n)\}$, one can define a generalized superoperator as follows:

$$\mathbb{E}^A_B(n) := \sum_{i=1}^{d_n} A^i(n) \otimes \overline{B^i}(n) \,. \tag{65.5}$$

Example 65.1.6. Using these definitions, one can rewrite the formula for expectation values more efficiently. Given a product operator $\widehat{O} = \bigotimes_i^N \widehat{O}_i$, one finds that

$$\langle \psi[A] \mid \widehat{O} \mid \psi[A]\rangle = \mathrm{tr}\big(\mathbb{E}_{O_1}(1) \cdots \mathbb{E}_{O_N}(N)\big) \,. \tag{65.6}$$

Formula 65.1.7. Associated to the superoperator $\mathbb{E}_O(n)$, one can define a virtual superoperator:

$$\begin{aligned} \mathcal{E}^{(n)}_{O_n}(\phi) &= \sum_{i,i'=1}^{d_n} \langle s \mid \widehat{O}_n \mid s'\rangle A^{i'}(n)\phi A^i(n)^\dagger \,, \\ \widetilde{\mathcal{E}}^{(n)}_{O_n}(\phi) &= \sum_{i,i'=1}^{d_n} \langle s \mid \widehat{O}_n \mid s'\rangle A^i(n)^\dagger \sigma A^{i'}(n) \,, \end{aligned} \tag{65.7}$$

where $\phi \in \mathcal{L}(\mathbb{C}^{D_n})$ and $\sigma \in \mathcal{L}(\mathbb{C}^{D_{n-1}})$.

Property 65.1.8. The map $\mathcal{E}^{(n)}_{\mathbb{1}}$ associated to the transfer operator is a CP map (Definition 26.2.15) and the associated Kraus operators (Corollary 26.2.46) are the MPS matrices $A^i(n)$.

65.1.2 Injectivity

Definition 65.1.9 (Injective MPS). A translation-invariant MPS is said to be injective if its transfer operator has a unique maximal eigenvalue.

In the next sections, the MPSs will always be assumed to be injective unless stated otherwise.

65.1.3 Gauge freedom and canonical forms

Property 65.1.10 (Gauge freedom). As is clear from the construction of matrix product states, there exists some freedom in the representation of the MPS tensors. One can always perform a transformation of the form $A(n) \to X^{-1}(n)A(n)X(n+1)$.

Remark 65.1.11. If one uses periodic boundary conditions, one must require that $X(L+1) = X(1)$, where L is the lattice size.

Using the gauge freedom in the representation of a generic MPS, one can construct certain forms which have useful properties.

Construction 65.1.12 (Left canonical form[1]). This form is specified by the following property:

(65.8)

Any MPS can be written in this form. First, construct the transfer operator $\mathbb{E}(n)$ for every site and find its maximal eigenvector. By the Perron–Frobenius theorem 22.5.6, this eigenvector (which is in fact itself a matrix) is positive and, hence, allows a decomposition of the form $\lambda(n) = L^\dagger(n)L(n)$. The left orthogonal forms are then defined by

$$A_L(n) := L(n)A(n)L^{-1}(n+1)\,. \tag{65.9}$$

In a similar manner, one can construct the right orthogonal form A_R.

Method 65.1.13 (Vidal). Given a general quantum state in terms of an n-leg tensor, there exists an efficient way of constructing the left (or right) canonical forms due to Vidal (2003). To this end, perform a cut between the first and second site and apply a singular value decomposition to obtain a tensor of the form $U^{[1]}SV^{[2,\dots]}$. Recursively applying this procedure to the product of the singular values S and the right unitary V will result in the canonical form.

[1] Also called the **left isometric form**, **left orthogonal form** or just **left gauge**.

Construction 65.1.14 (Mixed canonical form). The left and right canonical forms can also be combined. Let $L(n)$ and $R(n)$ be the decompositions of the left and right eigenvectors of the transfer operator at site n, i.e. $\lambda(n) = L^\dagger(n)L(n)$ and $\rho(n) = R(n)R^\dagger(n)$. The left and right canonical forms are then related by a matrix $C(n)$ in the following way: $A_L(n)C(n+1) = C(n)A_R(n)$. These matrices are given by

$$C(n) = L(n)R(n)\,. \tag{65.10}$$

65.1.4 Translation-invariant states

Definition 65.1.15 (Uniform MPS). By choosing all MPS tensors $A(n) = B$ for a given tensor B, one obtains a translation-invariant (TI) state. These MPSs form the variational class of uniform MPS.

Remark 65.1.16 (TIMPS). Not every TIMPS is uniform, there should only exist a local gauge transformation $A'(n) = U(n-1)A(n)U(n)^{-1}$ such that $A'(n)$ is uniform (in certain cases, this is only possible by enlarging the bond dimension).

65.2 Matrix product operators

Definition 65.2.1 (Matrix product operator[2]). Starting from the general form of an MPS, one can easily construct more general objects. By replacing the rank-3 tensors $A^i(n)$ with rank-4 tensors $A^{i,j}(n)$ and $|i_1 \cdots i_n\rangle$ by $|i_1\rangle\langle j_1| \otimes \cdots \otimes |i_n\rangle\langle j_n|$, one obtains the notion of a matrix-product operator:

$$\widehat{O} = \sum_{\{i_k, j_l\}} \mathrm{tr}\left(\prod_{m,n=1}^{N} O^{i_m, j_n}(n)\right) |i_1 \cdots i_N\rangle\langle j_1 \cdots j_n|\,. \tag{65.11}$$

In terms of a basis $\{\widehat{O}_i\}$ for the space of local operators, this becomes:

$$\widehat{O} = \sum_{\{i_k\}} \mathrm{tr}\left(\prod_{n}^{N} A^{i_n}(n)\right) \widehat{O}_{i_n}\,. \tag{65.12}$$

Method 65.2.2 (Local Hamiltonian to MPO). Given a local Hamiltonian $\widehat{H} = \sum_i \widehat{H}^{(i)}$, one can build an MPO that generates this Hamiltonian[3]:

$$\widehat{H} := \sum_{\{i_k, j_l\}} \mathrm{tr}\left(\prod_{m,n=1}^{N} A^{i_m, j_n}(n)\right) |i_1 \cdots i_N\rangle\langle j_1 \cdots j_n|\,. \tag{65.13}$$

To obtain this MPO, one can use the concept of a *cellular automaton*. This is a set of possible states together with a set of rules that say how one can go from one state

[2] As in the case of matrix product states this will be abbreviated as **MPO**.

[3] In fact, one can use this procedure to turn any local operator into an MPO.

to another. To obtain the set of states in this case, look at a given site i. All distinct combinations of 1-site operators to the right of i give rise to a distinct state μ. The transition rules are obtained by looking at which operator can be placed at the site i in a way consistent with the form of the given Hamiltonian.

Example 65.2.3. Consider a 2-site Hamiltonian of the form

$$\widehat{A} \otimes \hat{B} \otimes \mathbb{1} \otimes \cdots + \mathbb{1} \otimes \widehat{A} \otimes \hat{B} \otimes \mathbb{1} \otimes \cdots + \cdots . \tag{65.14}$$

For a specific site i, three distinct possibilities can arise:

- There are only identity operators acting to the right of i.
- Immediately to the right there is an operator $\hat{B}$ acting on $i+1$.
- Somewhere to the right there is a combination $\widehat{A} \otimes \hat{B}$.

The transition rules for this automaton are then given by the following list:

- $1 \to 2$: $\mathbb{1}$,
- $1 \to 2$: $\hat{B}$,
- $2 \to 3$: $\widehat{A}$, and
- $3 \to 3$: $\mathbb{1}$.

What is useful here is that this set of transition rules can be turned into a matrix:

$$T = \begin{pmatrix} \mathbb{1} & \hat{B} & 0 \\ 0 & 0 & \widehat{A} \\ 0 & 0 & \mathbb{1} \end{pmatrix}. \tag{65.15}$$

The MPO is then obtained by setting the MPO matrix A equal to T at every site.

65.2.1 MPO-injectivity

Definition 65.2.4 (MPO-injective PEPS). Consider a trivalent PEPS network on a manifold M and select a simply-connected subregion $\Omega \subset \Lambda$. By contracting the tensors withing this region, one obtains a linear map

$$A_\Lambda : (\mathbb{C}^D)^{\otimes|\Lambda|} \to (\mathbb{C}^d)^{\otimes|\partial\Lambda|} \tag{65.16}$$

from the virtual spaces on the edges to the physical space in the bulk. This PEPS is said to be MPO-injective if there exists a linear map (four-leg tensor)

$$M : \mathbb{C}^D \otimes \mathbb{C}^m \to \mathbb{C}^D \otimes \mathbb{C}^m \tag{65.17}$$

such that, for every subregion $\Omega \subset \Lambda$, the linear map A_Λ is injective on a (maximal) subspace S for which the projector onto S can be written as an MPO constructed from

the tensors M living on the boundary $\partial\Lambda$. See Fig. 65.1, where the tensors M are given by crossings of black and red lines.

Figure 65.1: MPO-injective PEPS.

Axiom 65.1 (Pulling-through). One of the key features of topological order is that it cannot be detected locally. Only a global measurement can show the existence of topologically ordered states. To this end, an axiom is introduced that allows to pull an MPO through the lattice. Graphically this is shown in Fig. 65.2.

Figure 65.2: Pulling-through condition.

65.2.2 MPO-symmetries for SPT phases

One can generalize the above framework to also include symmetry-protected topological order. To this end, one has to slightly modify the axioms.

Axiom 65.2 (Pulling-through for SPT phases). When pulling a symmetry-MPO through a tensor one has to act with a unitary on the physical level. Graphically this is shown in Fig. 65.3.

Figure 65.3: Pulling-through condition for SPT phases.

Part XI

Relativity & Quantum Field Theory

Table of Contents

66 Special Relativity **1145**
66.1 Lorentz transformations 1145
66.2 Energy and momentum 1146
66.3 Action principle 1146

67 General Relativity **1149**
67.1 Causal structure 1149
67.2 Einstein field equations 1151
67.3 Black holes 1152
67.4 Conserved quantities 1152
67.5 Tetradic formulation 1153

68 Dirac Equation **1155**
68.1 Dirac matrices 1155
68.2 Spinors 1156
68.2.1 Dirac equation 1156
68.2.2 Chiral spinors 1157
68.2.3 Dirac algebra in $d = 4$ 1160
68.2.4 Fierz identities 1160

69 Gauge Theory **1162**
69.1 Gauge invariance 1163
69.1.1 Gauge principle 1163
69.1.2 Gauge fixing 1164
69.1.3 Anomalies 1166
69.2 Yang–Mills theory 1166
69.2.1 Yang–Mills equations 1166
69.2.2 Currents and matter coupling 1167
69.2.3 2 dimensions 1167
69.3 Spontaneous symmetry breaking 1168
69.3.1 Higgs mechanism 1168
69.4 Instantons 1170

69.4.1 Instanton sectors 1170
69.4.2 Large gauge transformations 1171
69.4.3 Selfduality 1172
69.4.4 't Hooft–Polyakov instantons 1173
69.4.5 Dyons 1173
69.4.6 Electric-magnetic duality 1174
69.5 Chern–Simons theory 1174
69.6 Quantization 1174
69.6.1 Auxiliary fields 1174

70 Quantum Field Theory **1175**
70.1 Klein–Gordon field 1176
70.1.1 Lagrangian and Hamiltonian 1176
70.1.2 Raising and lowering operators 1176
70.1.3 Scalar propagator 1178
70.1.4 Normalization constant 1178
70.1.5 Invariant integration measure 1178
70.2 Contractions and Wick's theorem 1179
70.2.1 Bosonic fields 1179
70.2.2 Fermionic fields 1180
70.3 Gauge theory 1181
70.3.1 Schwinger–Dyson equation 1181
70.3.2 Faddeev–Popov integration 1183
70.3.3 BV-BRST theory 1184
70.4 Quantum chromodynamics 1186
70.5 Feynman rules 1186
70.5.1 Scalar theory 1186
70.5.2 Graph theory 1187
70.5.3 Degree of divergence 1188
70.5.4 Amplituhedron ♣ 1190
70.6 Renormalization 1190
70.6.1 Introduction: Statistical physics 1190
70.6.2 Dimensional regularization 1190
70.6.3 Wilsonian renormalization 1192
70.6.4 Connes–Kreimer renormalization 1192
70.7 Entanglement in QFT 1196
70.7.1 Lattice theories 1196

71 Particle Physics **1198**
71.1 Overview 1198
71.2 Standard model 1200

71.2.1 Group theory 1200
71.2.2 Symmetry breaking 1201
71.3 Shortcomings 1203
71.3.1 Open problems 1203
71.3.2 Might-be problems 1203

72 Axiomatic QFT ♣ 1205
72.1 Topological QFT 1206
72.1.1 Atiyah–Segal axioms 1206
72.1.2 Open-closed TQFT 1207
72.2 Algebraic QFT 1207
72.2.1 Haag–Kastler axioms 1207
72.3 Perturbative AQFT 1209
72.3.1 Haag–Kastler axioms 1210
72.3.2 Functional analysis 1210
72.3.3 Equations of motion 1212
72.3.4 Symmetries 1215
72.3.5 Quantization 1216
72.4 Prequantum field theory 1218

73 Conformal Field Theory 1220
73.1 Introduction 1221
73.2 In dimension $d = 2$ 1221
73.2.1 Minkowski space 1222
73.3 Quantization in $d = 2$ 1223
73.3.1 Radial quantization 1224
73.3.2 Virasoro algebra 1225
73.3.3 Representation theory 1225
73.4 Vertex operator algebras 1226
73.4.1 Definition 1226
73.4.2 Rational CFTs 1227
73.5 Extensions ♣ 1227
73.5.1 TCFT 1227
73.5.2 Strange correlator 1227

74 Extensions and Unification 1229
74.1 Supersymmetry 1230
74.1.1 Supersymmetric quantum mechanics 1230
74.1.2 Loop space mechanics 1231
74.1.3 Extensions of the Standard Model 1231
74.1.4 Spinning particle 1232

74.2 Chern–Simons theory 1233
74.2.1 Holomorphic Chern–Simons theory 1233
74.3 Four-manifolds 1233
74.3.1 Donaldson–Witten theory 1234
74.3.2 Instanton Floer homology 1236
74.3.3 Seiberg–Witten theory 1237
74.4 String theory 1240
74.4.1 Worldsheets 1240
74.4.2 Branes 1241
74.4.3 Duality 1241
74.4.4 Superstrings 1241
74.5 M-theory 1242

75 Astronomy & Cosmology 1243
75.1 Stars 1243
75.2 Galaxies 1243
75.3 Black holes 1244
75.4 Inflation 1244
75.5 Tegmark multiverses 1244

Chapter 66

Special Relativity

In this chapter, as will be the case in the chapters on quantum field theory, the mostly-minuses convention for the Minkowski signature is adopted unless stated otherwise, i.e. the signature is $(+,-,-,-)$. Furthermore, natural units will be used unless stated otherwise, i.e. $\hbar = c = 1$. This follows the introductory literature such as Greiner and Reinhardt (1996); Peskin and Schroeder (1995).

Chapter Contents

66.1 Lorentz transformations **1145**

66.2 Energy and momentum **1146**

66.3 Action principle **1146**

66.1 Lorentz transformations

Notation 66.1.1. In the context of special relativity, it is often useful to introduce the following quantities:

$$\begin{aligned} \beta &:= \frac{v}{c}, \\ \gamma &:= \frac{1}{\sqrt{1-\beta^2}}. \end{aligned} \tag{66.1}$$

The latter quantity is called the **Lorentz factor**.

Formula 66.1.2 (Lorentz transformations). Let $\mathbf{V}$ be a 4-vector. A Lorentz boost along

the x^1-axis is given by the following transformation:

$$\begin{cases} V'^0 = \gamma\left(V^0 - \beta V^1\right) \\ V'^1 = \gamma\left(V^1 - \beta V^0\right) \\ V'^2 = V^2 \\ V'^3 = V^3. \end{cases} \tag{66.2}$$

Remark 66.1.3. Setting $c = +\infty$ in the previous formulas recovers the Galilei transformations from classical mechanics (cf. Inönü–Wigner contractions (Definition 32.5.1)).

66.2 Energy and momentum

Formula 66.2.1 (4-velocity). In analogy to the definition of velocity in classical mechanics, the 4-velocity is defined as follows:

$$\mathbf{U} := \left(\frac{\mathrm{d}x^0}{\mathrm{d}\tau}, \frac{\mathrm{d}x^1}{\mathrm{d}\tau}, \frac{\mathrm{d}x^2}{\mathrm{d}\tau}, \frac{\mathrm{d}x^3}{\mathrm{d}\tau}\right). \tag{66.3}$$

By applying the formulas for proper time and time dilatation, one obtains:

$$\mathbf{U} = \left(\gamma c, \gamma \vec{\mathbf{u}}\right). \tag{66.4}$$

Formula 66.2.2 (4-momentum). The 4-momentum is defined as follows:

$$\mathbf{p} = m_0 \mathbf{U}, \tag{66.5}$$

or, after defining $E := cp^0$:

$$\mathbf{p} = \left(\frac{E}{c}, \gamma m_0 \vec{\mathbf{u}}\right). \tag{66.6}$$

Definition 66.2.3 (Relativistic mass). The factor $m := \gamma m_0$ in the momentum 4-vector is called the relativistic mass. By introducing this quantity, the classical formula $\vec{\mathbf{p}} = m\vec{\mathbf{u}}$ for the 3-momentum can be generalized to 4-momenta $\mathbf{p}$.

Formula 66.2.4 (Relativistic energy relation).

$$E^2 = p^2c^2 + m^2c^4 \tag{66.7}$$

This formula is often called the **Einstein relation**.

66.3 Action principle

The main guiding principles for writing down a relativistic action for a point particle are locality and invariance. The latter means that one should only use geometric

quantities, i.e. diffeomorphism-invariant quantities, while the former means that these should only depend on local information. For a single particle, the most obvious choice of action would be one that is proportional to the proper time along the worldline of the particle or, more invariantly, the arc length of the worldline:

$$S_{\text{point}} \sim \int_\gamma ds\,. \tag{66.8}$$

To get the units right, one should multiply by suitable Lorentz-invariant constants:

$$S_{\text{point}} := mc \int_\gamma ds\,. \tag{66.9}$$

By reparametrization invariance, one can choose a specific time coordinate, e.g. $\tau = ct$. In this coordinate system, the action becomes

$$S_{\text{point}} = mc \int_\gamma \sqrt{1 - \frac{v^2}{c^2}}\, c\, dt\,, \tag{66.10}$$

with v the speed of the particle. The Lagrangian density can be Taylor expanded as

$$L_{\text{point}} = mc^2 - \frac{1}{2} m v^2 + \cdots\,, \tag{66.11}$$

which recovers (up to a constant) the classical Lagrangian of a massive point particle when the speed is small $v \ll c$.

When trying to quantize this action, however, a problem occurs. After a Legendre transformation, the Hamiltonian becomes $H = \sqrt{p^2 + m^2c^4}$ (this is just the Einstein relation 66.2.4). When applying the ordinary Dirac procedure $p_\mu \longrightarrow i\partial_\mu$, this becomes a nonlocal operator (the whole reason for why Dirac introduced spinors).

Instead of passing to a spinor framework, one can try to write down an equivalent action that gives rise to a local Hamiltonian. One possibility is to pass to **light cone coordinates**:

$$x^\pm := x^0 \pm x^1\,. \tag{66.12}$$

However, here one makes a specific split of coordinates, which ruins Lorentz invariance. A better idea is to introduce a dynamical Lagrange multiplier (from here on, natural units are used):

$$S_{\text{point}} := \frac{1}{2} \int_\gamma \left[\eta^{-1} \left(\frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} \right)^2 - m^2 \eta \right] d\tau\,. \tag{66.13}$$

The equation of motion for η is algebraic and gives

$$\eta = \frac{1}{m} \sqrt{-g_{\mu\nu} \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} \frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}}\,. \tag{66.14}$$

To find an interpretation of this multiplier, it is useful to consider the case where a metric h is introduced on the worldline. This implies that the action has to be 'covariantized':

$$\int_\gamma \left(g_{\mu\nu} \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} + m^2 \right) d\tau \longrightarrow \int_\gamma \sqrt{h} \left(h g_{\mu\nu} \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} + m^2 \right) d\tau \,. \tag{66.15}$$

From this perspective, it is clear that the Lagrange multiplier can be viewed as the square root of a dynamical metric on the worldline. It is an example of a *vielbein* (in this case, an '*einbein*'). Furthermore, this action is a so-called *nonlinear σ-model*.

Chapter 67

General Relativity

References for this chapter are Misner, Thorne, and Wheeler (2017); Rovelli and Vidotto (2014). See Chapter 36 for an introduction to the theory of Riemannian geometry. The mathematical background for the section on the tetradic formulation of GR can be found in Section 35.8.

In this chapter, the signature convention of the previous chapter is reversed. For general relativity, it is often more convenient to use the mostly-pluses convention (this simply reduces the number of minus signs).

67.1 Causal structure

Definition 67.1.1 (Null coordinate). Consider a vector $v \in T_pM$ on a Lorentzian manifold (M, g). This vector is said to be null or **lightlike** if it satisfies the following condition:

$$g_p(v, v) = 0\,. \tag{67.1}$$

One can also define **timelike** and **spacelike** vectors in a similar way as those vectors having negative and positive norm, respectively.[1] Spacelike, lightlike and timelike curves are defined as curves for which every tangent vector is respectively spacelike, lightlike or timelike. A curve is said to be **causal** if its tangent vectors are time- or lightlike.

Remark 67.1.2. Note that Riemannian manifolds do not admit timelike and lightlike vectors as their metric is positive definite.

Definition 67.1.3 (Time orientability). A Lorentzian manifold is said to be time orientable if there exists a nowhere-vanishing timelike vector field. It should be noted

[1] For a mostly-minuses signature, one should interchange these definitions.

that, in contrast to ordinary orientability, this notion is not purely topological. Moreover, neither orientability nor time orientability implies the other. They are independent notions.

The choice of a time-orienting vector field τ divides the set of timelike vectors at a point into two equivalence classes. A curve γ is said to be **future directed** (resp. **past directed**) if $g(\tau, \dot{\gamma}) < 0$ (resp. $g(\tau, \dot{\gamma}) > 0$).

Definition 67.1.4 (Causal cone). Let M be a Lorentzian manifold. The causal cone of a point $p \in M$ is defined as the set $J^-(p) \cup J^+(p) \subset M$ of points that are connected to p by a (smooth) causal curve. The past and future cones $J^\pm(p)$ are respectively defined as the sets of points that can be connected to p by a future-directed or past-directed casual curve. The boundaries of these causal cones are called the causal **lightcones**, sometimes denoted by $V^\pm(p)$.

Definition 67.1.5 (Causal closure). Let S be a subset of a Lorentzian manifold. The **causal complement** of S consists of all points that cannot be causally connected to any point in S. The causal closure of S is defined as the causal complement of the causal complement of S. A **causally closed set** is then defined as a set that is equal to its causal closure.

Definition 67.1.6 (Globally hyperbolic manifold). A Lorentzian manifold M that does not contain closed causal curves and for which $J^+(p) \cap J^-(q)$ is compact for any two points $p, q \in M$.

This concept can also be stated in a different way, through the relation between causal structures and initial value problems.

Definition 67.1.7 (Cauchy surface). Consider a Lorentzian manifold M. A curve $\gamma :]a, b[\to M$ is said to be an **inextensible timelike curve** if:

1. $\dot{\gamma}(t)$ is timelike for all $t \in]a, b[$.
2. γ does not approach a limit towards a or b.

Given a subset $S \subseteq M$, the **domain of dependence** $D(S)$ is given by the points $p \in M$ such that every inextensible timelike curve through p also intersects S.

A subset $S \subset M$ is called a Cauchy surface if every inextensible timelike curve in M intersects S exactly once and $D(S) = M$.

Property 67.1.8. For a Lorentzian manifold M, the following conditions are equivalent:

- M is globally hyperbolic, and
- M admits a Cauchy surface S.

Moreover, in this case, there exists a homeomorphism (or even diffeomorphism if S

is taken to be smooth) $M \cong S \times \mathbb{R}$. A time orientation of M is then equivalent to an orientation of the factor $\mathbb{R}$.

Definition 67.1.9 (Stationary spacetime). A Lorentzian manifod M is said to be stationary if there exists a timelike Killing vector. By the *flowbox theorem*, there always exists a coordinate chart such that, locally, one can choose the Killing vector field to be ∂_0 and, hence, M is stationary if one can find a coordinate system for which the metric coefficients are time independent.

Definition 67.1.10 (Celestial sphere). Consider a 4-dimensional Lorentzian manifold M. The celestial sphere at a point $p \in M$ is given by the projectivization $\mathbb{P}\big(J^+(p)\big) \cong \mathbb{CP}^1 \cong S^2$, i.e. it consists of all rays in the future (or, equivalently, past) lightcone at p. As Schreiber (2008) so nicely phrases it "*your celestial sphere (the one around the point where your head is) is the sphere of which you directly perceive a portion when you look.*".

Property 67.1.11. By the above definition and Property 68.2.9, the celestial sphere is given by $\mathbb{CP}^1$ or, in the spinorial setting, by the Weyl spinors modulo rescalings.

67.2 Einstein field equations

Formula 67.2.1 (Einstein field equations). The Einstein field equations with a cosmological constant Λ read as follows (all fundamental constants are shown for completeness):

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} + \Lambda g_{\mu\nu}\,, \tag{67.2}$$

where $G_{\mu\nu}$ is the Einstein tensor (36.2.9) and $T_{\mu\nu}$ is the stress-energy tensor (53.15).

By taking the trace of both sides, one obtains $T = -R$ and, hence, the Einstein field equations can be rewritten as

$$R_{\mu\nu} = \widehat{T}_{\mu\nu}\,, \tag{67.3}$$

where $\widehat{T}_{\mu\nu} := T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T$ is the **reduced stress-energy tensor**.

Formula 67.2.2 (Einstein–Hilbert action). The (vacuum) field equations can be obtained by applying the variational principle to the following action:

$$S_{\mathrm{EH}}[g_{\mu\nu}] := \int_M \sqrt{-g} R \ \mathsf{Vol}\,. \tag{67.4}$$

For manifolds with boundary, one needs an extra term to make the boundary contributions vanish (as to obtain a well-defined variational problem). This term is due to *Gibbons*, *Hawking* and *York*:[2]

$$S_{\mathrm{GHY}}[g_{\mu\nu}] := \oint_{\partial M} \epsilon \sqrt{h} K \ \mathsf{Vol}, \tag{67.5}$$

[2] *Einstein* had, in fact, already introduced a variant: the $\Gamma\Gamma$-*Lagrangian*.

where h_{ab} is the induced metric on the boundary, K_{ab} is the extrinsic curvature and $\epsilon = \pm 1$ is a 'function' depending on whether the boundary is timelike or spacelike.

67.3 Black holes

Formula 67.3.1 (Schwarzschild metric).

$$ds^2 := \left(1 - \frac{R_s}{r}\right)c^2dt^2 - \left(1 - \frac{R_s}{r}\right)^{-1} dr^2 - r^2 d\Omega^2 \,, \tag{67.6}$$

where R_s is the Schwarzschild radius

$$R_s := \frac{2GM}{c^2} \,. \tag{67.7}$$

Theorem 67.3.2 (Birkhoff). *The Schwarzschild metric is the unique solution of the vacuum field equation under the additional constraints of asymptotic flatness and staticity.*

Formula 67.3.3 (Reissner–Nordström metric). If the black hole is allowed to have an electric charge Q, the Schwarzschild metric must be modified in the following way:

$$ds^2 := \left(1 - \frac{2GM}{r} + \frac{GQ^2}{4\pi r^2}\right)c^2dt^2 - \left(1 - \frac{2GM}{r} + \frac{GQ^2}{4\pi r^2}\right)^{-1} dr^2 - r^2 d\Omega^2 \,. \tag{67.8}$$

Remark 67.3.4. The electric field generated by a Reissner–Nordström black hole is given by

$$E^r = \frac{Q}{4\pi r^2} \,. \tag{67.9}$$

Although the coordinate r is not the proper distance, it still acts as a parameter for the surface of a sphere (as it does in a Euclidean or Schwarzschild metric). This explains why the above formula is the same as the one in classical electromagnetism.

@@ ADD (KERR–NEWMAN, ERGOSPHERE, PENROSE MECHANISM, KRUSKAL) @@

67.4 Conserved quantities

Before raising any hope that conserved quantities are to be found in GR, the following result is given.

Property 67.4.1. By Noether's third theorem 40.1.9, there exist no proper (local) conservation laws such as those of momentum and energy in classical mechanics, since the translation group is a subgroup of the infinite-dimensional symmetry group $\mathrm{Diff}(M)$.

@@ COMPLETE @@

67.5 Tetradic formulation

This section starts from the geometric interpretation of the (weak) equivalence principle, i.e. spacetime is locally modelled on Minkowski space. The natural language for this kind of geometry is that of Cartan geometries (Section 35.8). By the Erlangen program, it is known that Minkowski spacetime M^4 can be described as the coset space $\mathsf{ISO}(3,1)/\mathsf{SO}(3,1)$. The natural generalization is given by a Cartan geometry with model geometry $\big(\mathfrak{iso}(3,1), \mathfrak{so}(3,1)\big)$.

Property 67.5.1 (Cartan connection). This way, a $\mathsf{SO}(3,1)$-structure on the spacetime manifold M is obtained, i.e. a choice of Lorentzian metric g. The Cartan connection $\widetilde{\nabla}$ can also be decomposed as $\nabla + \mathsf{e}$ where:

- ∇ defines a $\mathfrak{so}(3,1)$-valued principal connection, and
- e defines an M^4-valued solder form.

The principal connection ∇ is called the **spin connection** and e is called the **vierbein** or **tetrad**.[3] These objects are well known in general relativity. In case of vanishing torsion, the connection ∇ is the ordinary Levi-Civita connection associated to the Lorentzian manifold M, and e gives the isometry between local (flat) Minkowski coordinates and 'global' coordinates:

$$g := \mathsf{e}^* \eta \tag{67.10}$$

or, locally,

$$g_{\mu\nu} = e^i_\mu e^j_\nu \eta_{ij} \,. \tag{67.11}$$

Using the tetrad field, one can rewrite the Einstein–Hilbert action in a very elegant way. To this end, a new curvature form is defined:

$$F^i{}_{j\mu\nu} := e^i_\rho e^\sigma_j R^\rho{}_{\sigma\mu\nu} \,, \tag{67.12}$$

where $R^\rho{}_{\sigma\mu\nu}$ is the ordinary Riemann curvature tensor. The Einstein–Hilbert action is then equivalent[4] to the following Yang–Mills-like action.

Formula 67.5.2 (Palatini action).

$$S[e,\nabla] := \int_M \mathsf{e} \wedge \mathsf{e} \wedge *F \,. \tag{67.13}$$

This action is sometimes called the **tetradic Palatini action** and the resulting formulation of general relativity is called the **first-order formulation**. If one considers the

[3] In other dimensions, it is called a **vielbein**.
[4] At least in the case of pure gravity (Rovelli & Vidotto, 2014).

same action but only as a functional of the tetrad field, one obtains the **second-order formulation** of gravity.[5]

Variation of the Palatini action gives the following EOM:

- $\delta\nabla$: $T(\mathsf{e}) = 0$ or, equivalently, $\nabla(\mathsf{e}) \equiv \nabla$, i.e. the torsion vanishes and the connection ∇ is, on-shell, equal to the Levi-Civita connection on M.
- $\delta\mathsf{e}$: The metric g satisfies the Einstein field equations.

Because of its importance in general relativity, the first factor in the Palatini action deserves a name.

Definition 67.5.3 (Plebanski form).

$$\Sigma := \mathsf{e} \wedge \mathsf{e} \tag{67.14}$$

Because of its internal antisymmetric Lorentz indices, one can interpret this object as an $\mathfrak{so}(3,1)$-valued 2-form.

As is the case for *4D Yang–Mills theory* (see Section 69.2), one can introduce a topological term that leaves the EOM invariant (up to boundary terms).

Definition 67.5.4 (Holst action[6]).

$$\begin{aligned} S[\mathsf{e},\nabla] :&= \int_M \mathsf{e} \wedge \mathsf{e} \wedge *F + \frac{1}{\gamma}\int_M \mathsf{e} \wedge \mathsf{e} \wedge F \\ &= \int_M \left(*\mathbf{e} \wedge \mathsf{e} + \frac{1}{\gamma}\mathsf{e} \wedge \mathsf{e} \right) \wedge F. \end{aligned} \tag{67.15}$$

The coupling constant γ is called the **Barbero–Immirzi** constant.

@@ COMPLETE @@

[5] These formulations are equivalent for pure gravity. However, when coupling the theory to fermions, they differ by a four-fermion vertex. This follows from the introduction of torsion due to the fermions.

[6] *Holst* was actually the second author to include this term.

Chapter 68

Dirac Equation

References for this chapter are Van Proeyen and Freedman (2012). (Note that these authors use the mostly-pluses signature.) For the mathematical background of Clifford algebras and Spin-groups, see Chapter 27 and, in particular, Section 27.4.2. For the extension to (pseudo-)Riemannian manifolds, see Section 36.3.

Chapter Contents

68.1 Dirac matrices **1155**

68.2 Spinors **1156**

68.2.1 Dirac equation 1156

68.2.2 Chiral spinors 1157

68.2.3 Dirac algebra in $d = 4$ 1160

68.2.4 Fierz identities 1160

68.1 Dirac matrices

Definition 68.1.1 (Dirac matrices). The Dirac (or **gamma**) matrices are defined by the following equality:

$$\{\gamma^\mu, \gamma^\nu\}_+ = 2\eta^{\mu\nu}\mathbb{1}\,, \tag{68.1}$$

where $\eta^{\mu\nu}$ is the Minkowski metric. This has the form of Eq. (27.9), i.e. the Dirac matrices form the generating set of a Clifford algebra, called the **Dirac algebra**.

There exist multiple distinct representations of the Clifford generators in signature $(1,3)$. The first one is called the **Dirac representation**. Here, the timelike Dirac ma-

trix γ^0 is defined as

$$\gamma^0 := \begin{pmatrix} \mathbb{1}_2 & 0 \\ 0 & -\mathbb{1}_2 \end{pmatrix}. \tag{68.2}$$

The spacelike Dirac matrices γ^k $(k = 1, 2, 3)$ are defined using the Pauli matrices (Definition 57.6.9) σ^k:

$$\gamma^k := \begin{pmatrix} 0 & \sigma^k \\ -\sigma^k & 0 \end{pmatrix}. \tag{68.3}$$

The **Weyl** or **chiral** representation[1] is defined by replacing the timelike matrix γ^0 by

$$\gamma^0 := \begin{pmatrix} 0 & \mathbb{1}_2 \\ \mathbb{1}_2 & 0 \end{pmatrix}. \tag{68.4}$$

In signature $(3, 1)$, one obtains the Weyl representation by defining $\sigma^\mu := (\mathbb{1}, \sigma_i)$ and $\overline{\sigma}^\mu := (\mathbb{1}, -\sigma_i)$:

$$\gamma^\mu := \begin{pmatrix} 0 & \sigma^\mu \\ \overline{\sigma}^\mu & 0 \end{pmatrix}. \tag{68.5}$$

Remark. In the remainder of this compendium, the Weyl representation will be used.

Notation 68.1.2 (Feynman slash notation). Let $\mathbf{a} \equiv a^\mu \mathbf{e}_\mu \in M^4$ be a general 4-vector. The Feynman slash is defined as follows:

$$\not{a} := a^\mu \gamma_\mu \,. \tag{68.6}$$

In fact, this is just the embedding of Minkowski space in its Clifford algebra:

$$/ : M^4 \to C\ell(M^4, \eta) : a^\mu \mathbf{e}_\mu \mapsto a^\mu \gamma_\mu \,. \tag{68.7}$$

68.2 Spinors

68.2.1 Dirac equation

Formula 68.2.1 (Dirac equation). In covariant form, the Dirac equation reads as

$$(i\hbar \not{\partial} - mc)\psi = 0 \,, \tag{68.8}$$

where m denotes the mass and c denotes the speed of light. In the case of massless particles, the **Weyl equation** is recovered:

$$\not{\partial}\psi = 0 \,. \tag{68.9}$$

[1] This representation is preferred in quantum field theory and supergravity.

Definition 68.2.2 (Dirac adjoint).

$$\overline{\psi} := i\psi^\dagger\gamma^0 \tag{68.10}$$

When working in the Dirac representation, the factor i should be dropped.

Definition 68.2.3 (Majorana adjoint). In the context of supersymmetry, it is often convenient to work with a different adjoint spinor. Let $\widehat{\mathcal{C}} := i\gamma^3\gamma^1$ denote the charge conjugation operator. The Majorana adjoint is defined by

$$\overline{\psi} := \psi^t\widehat{\mathcal{C}}\,. \tag{68.11}$$

Formula 68.2.4 (Parity). The parity operator is defined as follows:

$$\widehat{P}(\psi) := \gamma^0\psi\,. \tag{68.12}$$

68.2.2 Chiral spinors

In an even number of dimensions, one can define an additional matrix that satisfies Eq. (68.1).

Definition 68.2.5 (Chiral matrix). Assume that the dimension $d = m + n$ is even. The chiral (helicity) matrix can be defined as follows:[2]

$$\gamma^{d+1} := \gamma^1\gamma^2\cdots\gamma^d\,. \tag{68.13}$$

In an odd number of dimensions ($d = 2m+1$), a generating set for the Clifford algebra can be obtained by taking the generating set from one dimension lower and adjoining the element $k\gamma^{d+1}$, where $k^2 = (-1)^{n+d/2}$. This gives two inequivalent representations of the Clifford algebra (depending on the sign). From here on, the following redefinition will be used:

$$\gamma^{d+1} \longrightarrow k\gamma^{d+1}\,. \tag{68.14}$$

This has the benefit that $(\gamma^{d+1})^2 = \mathbb{1}$. In $d = 3 + 1$, one generally takes the following representation for γ^5 (in the Dirac representation):[3]

$$\gamma^{d+1} := \begin{pmatrix} 0 & \mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix}. \tag{68.15}$$

Remark 68.2.6 (Weyl representation). Note that γ^5 in the Dirac representation corresponds to γ^0 in the Weyl representation. The chiral matrix in the Weyl representation is given by

$$\gamma^{d+1} := \begin{pmatrix} -\mathbb{1} & 0 \\ 0 & \mathbb{1} \end{pmatrix}. \tag{68.16}$$

[2]Some authors add a constant factor to this definition.

[3]Such a block diagonal form can always be chosen by working in a *helicity-adapted basis*.

Passing from the Dirac representation to the Weyl representation, hence, corresponds to the following transformation:

$$\begin{aligned} \gamma^0 &\longrightarrow \gamma^5 \,, \\ \gamma^5 &\longrightarrow -\gamma^0 \,. \end{aligned} \tag{68.17}$$

Property 68.2.7. The chiral matrix anticommutes with all other gamma matrices:

$$\{\gamma^{d+1}, \gamma^\mu\}_+ = 0 \tag{68.18}$$

for all $\mu \in \{0, \dots, d\}$.

Definition 68.2.8 (Chiral projection). The chiral projections of a spinor ψ are defined as follows:

$$\psi_- := \frac{1 + \gamma_{d+1}}{2} \psi \tag{68.19}$$

and

$$\psi_+ := \frac{1 - \gamma_{d+1}}{2} \psi \,. \tag{68.20}$$

Every spinor can then be written as a sum of its chiral parts:

$$\psi = \psi_+ + \psi_- \equiv \begin{pmatrix} \psi_+ \\ \psi_- \end{pmatrix} . \tag{68.21}$$

In general, since the Dirac equation is linear, it has a plane wave solution of the form

$$\psi(x) = u(p) e^{-ip\cdot x} + v(p) e^{ip\cdot x} \,, \tag{68.22}$$

the first term corresponding to (incoming) fermions and the second corresponding to (outgoing) antifermions. (Taking the Dirac adjoint switches incoming and outgoing states.) For massless particles, the Dirac equation reduces to the Weyl equation and has two solutions (of opposite **helicity**):

$$\not{p} u_\pm(p) = 0 \qquad \text{and} \qquad \not{p} v_\pm(p) = 0 \,, \tag{68.23}$$

where the helicity is defined as follows:

$$\hat{\lambda} := \frac{\hat{p} \cdot \vec{\sigma}}{|\hat{p}|} \,. \tag{68.24}$$

Since massless particles move at the speed of light, their helicity is independent of a choice of reference frame and, hence, the helicity is a well-defined quantity. Moreover, in the case of massless particles, a *crossing symmetry* is found:

$$u_\pm = v_\mp \,. \tag{68.25}$$

Property 68.2.9 (Bispinor representation). The exceptional isomorphism $\mathsf{Spin}(3,1) \cong \mathsf{SL}(2,\mathbb{C})$ from Example 27.4.16 allows to express the Lorentz norm as a determinant:

$$2\|(x^\mu)\| = \det(\not{x}) = 2\det((x^{\alpha\beta}))\,, \tag{68.26}$$

where

$$x_{\alpha\beta} := x_\mu(\sigma^\mu)_{\alpha\beta}\,. \tag{68.27}$$

It follows that if x is lightlike, it is equivalent to the outer product of a (Weyl) spinor ξ and its conjugate $\overline{\xi}$:

$$x^{\alpha\beta} = u_+^\alpha \overline{u}_+{}^\beta\,. \tag{68.28}$$

For future use, some new notations are introduced (akin to the standard Dirac notation for quantum states).

Notation 68.2.10 (Weyl notation). By the property above, every lightlike 4-vector can be expressed as the product of a Weyl spinor and its adjoint. Given this representation, one defines

$$\begin{aligned} |p\rangle &:= u_+(p)\,, \\ [p| &:= \overline{u}_+(p)\,, \end{aligned} \tag{68.29}$$

and, accordingly,

$$\begin{aligned} \langle p| &= \overline{u}_-(p)\,, \\ |p] &= u_-(p)\,, \end{aligned} \tag{68.30}$$

This also implies that $\not{p}$, which essentially contains two copies of $(p^{\alpha\beta})$ can also be expressed in the bispinor representation through the following completeness relation:[4]

$$\not{p} = |p\rangle[p| + |p]\langle p| = u_+(p)\overline{u}_+(p) + u_-(p)\overline{u}_-(p)\,. \tag{68.31}$$

The Weyl equation itself becomes:

$$p^{\alpha\beta}|p\rangle_\beta = 0\,. \tag{68.32}$$

Using the Weyl notation, one can also introduce spinor bilinears that are similar in nature to inner products, with the main difference that they are antisymmetric:

$$\begin{aligned} \langle p\,q\rangle &:= \langle p\,|\,q\rangle = \overline{u}_-(p)u_+(q)\,, \\ [p\,q] &:= [p\mid q] = \overline{u}_+(p)u_-(q)\,. \end{aligned} \tag{68.33}$$

[4]The addition in the first step should be interpreted as a direct sum.

Remark 68.2.11. Mixed bilinears, with one angle variable and one bracket variable, vanish. In fact, if they would not vanish, they would not be Lorentz invariant. To obtain an invariant quantity, a Dirac matrix should be introduced, e.g. $\langle p|\gamma^\mu|q]$.

Property 68.2.12 (Inner products). The inner product of two lightlike 4-vectors can be expressed in terms of spinor bilinears:

$$p \cdot q = \frac{1}{2}(p+q)^2 = \frac{1}{2}\langle p\, q\rangle[p\, q]\,. \tag{68.34}$$

Property 68.2.13 (Schouten identity). Consider four lightlike 4-vectors $\{p_1, p_2, p_3, p_4\}$. Given the bispinor decomposition $p_i = |i\rangle[i|$, the following equality is satisfied:

$$\langle 1\,2\rangle\langle 3\,4\rangle + \langle 1\,3\rangle\langle 4\,2\rangle + \langle 1\,4\rangle\langle 2\,3\rangle = 0\,. \tag{68.35}$$

68.2.3 Dirac algebra in $d = 4$

For a lot of calculations, especially in quantum electrodynamics, one needs the properties of the gamma matrices. The most relevant relations in $d = 3+1$ are listed below:

Formula 68.2.14 (Trace algebra).

$$\mathrm{tr}(\gamma^\mu) = \mathrm{tr}(\gamma^\mu\gamma^\nu\gamma^\rho) = 0 \tag{68.36}$$

$$\mathrm{tr}(\gamma^\mu\gamma^\nu) = 4\eta^{\mu\nu} \tag{68.37}$$

$$\mathrm{tr}(\gamma^\mu\gamma^\nu\gamma^\kappa\gamma^\lambda) = 4(\eta^{\mu\nu}\eta^{\kappa\lambda} - \eta^{\mu\kappa}\eta^{\nu\lambda} + \eta^{\mu\lambda}\eta^{\nu\kappa}) \tag{68.38}$$

$$\mathrm{tr}(\gamma^5) = \mathrm{tr}(\gamma^\mu\gamma^5) = \mathrm{tr}(\gamma^\mu\gamma^\nu\gamma^5) = \mathrm{tr}(\gamma^\mu\gamma^\nu\gamma^\rho\gamma^5) = 0 \tag{68.39}$$

$$\mathrm{tr}(\gamma^\mu\gamma^\nu\gamma^\kappa\gamma^\lambda\gamma^5) = -4i\varepsilon^{\mu\nu\kappa\lambda} \tag{68.40}$$

$$\mathrm{tr}(\gamma^{\mu_1}\cdots\gamma^{\mu_k}) = \mathrm{tr}(\gamma^{\mu_k}\cdots\gamma^{\mu_1}) \tag{68.41}$$

Formula 68.2.15 (Contraction identities).

$$\gamma^\mu\gamma_\mu = 4 \tag{68.42}$$

$$\gamma^\mu\gamma^\nu\gamma_\mu = -2\gamma^\nu \tag{68.43}$$

$$\gamma^\mu\gamma^\nu\gamma^\rho\gamma_\mu = 4\eta^{\nu\rho} \tag{68.44}$$

$$\gamma^\mu\gamma^\nu\gamma^\kappa\gamma^\lambda\gamma_\mu = -2\gamma^\lambda\gamma^\kappa\gamma^\nu \tag{68.45}$$

68.2.4 Fierz identities

Using a spinor $u \in S$ and a cospinor $\overline{v} \in S^*$, one can build a bilinear form $\overline{v}u$. However, for two spinors u, ω and two cospinors $\overline{v}, \overline{\rho}$, one can interpret the expression $(\overline{v}u)(\overline{\rho}\omega)$ either as a quadrilinear form on $u \otimes \overline{v} \otimes \omega \otimes \overline{\rho}$ or as a quadrilinear form on $\omega \otimes \overline{v} \otimes u \otimes \overline{\rho}$. Because $C\ell_{3,1}(\mathbb{C})$ is isomorphic to the endomorphism ring on S, there must exist coefficients a^{ij}, with $i, j = 1, \dots, 2^d$, such that

$$(\overline{v}u)(\overline{\rho}\omega) = \sum_{i,j=1}^{2^d} a^{ij}(\overline{v}\gamma_i\omega)(\overline{\rho}\gamma_j u)\,. \tag{68.46}$$

By using the orthogonality relations of the trace, one can find that

$$\alpha^{ij} = \begin{cases} 0 & \text{if } i \neq j\,, \\ \dfrac{1}{2^{\lfloor d/2 \rfloor}} & \text{if } i = j\,. \end{cases} \tag{68.47}$$

The above equality can then also be rewritten as follows:

$$\delta^a_b \delta^c_d = \frac{1}{2^{\lfloor d/2 \rfloor}} \sum_{i=1}^{2^d} (\gamma_i)^a_d (\gamma_i)^c_b\,. \tag{68.48}$$

This expression, and the techniques used to find it, allows one to rearrange almost any multilinear expression involving spinors and cospinors.

Chapter 69

Gauge Theory

References for this chapter are Belgun (2024); Nash and Sen (2011); Schuller (2016); Sontz (2015a). The section on the Higgs mechanism is mainly based on Choquet-Bruhat and DeWitt-Morette (2000). Using the tools of differential geometry, as presented in Chapter 33 and onwards, one can introduce a general formulation of gauge theories and, in particular, Yang–Mills theories.

Chapter Contents

69.1 Gauge invariance **1163**
69.1.1 Gauge principle 1163
69.1.2 Gauge fixing 1164
69.1.3 Anomalies 1166
69.2 Yang–Mills theory **1166**
69.2.1 Yang–Mills equations 1166
69.2.2 Currents and matter coupling 1167
69.2.3 2 dimensions 1167
69.3 Spontaneous symmetry breaking **1168**
69.3.1 Higgs mechanism 1168
69.4 Instantons **1170**
69.4.1 Instanton sectors 1170
69.4.2 Large gauge transformations 1171
69.4.3 Selfduality 1172
69.4.4 't Hooft–Polyakov instantons 1173
69.4.5 Dyons 1173
69.4.6 Electric-magnetic duality 1174

69.5 Chern–Simons theory **1174**
69.6 Quantization **1174**
69.6.1 Auxiliary fields 1174

69.1 Gauge invariance

69.1.1 Gauge principle

Consider a general Lie group G, often called the **gauge group**, acting on a vector bundle E with typical fibre $\mathcal{H}$ over a base manifold M (be wary of Definition 35.1.9). This bundle is, in general, obtained as a bundle associated to the frame bundle FM. Locally, a general gauge transformation has the form

$$\psi'(x) = U(x)\psi(x)\,, \tag{69.1}$$

where $\psi, \psi' : M \to \mathcal{H}$ are trivializations of local sections of E and $U : M \to G$ encodes the local behaviour of the gauge transformation. It is assumed to be a unitary representation with respect to the Hilbert structure on $\mathcal{H}$. Globally, the gauge transformations are given by the vertical automorphisms.

Axiom 69.1 (Local gauge principle). The Lagrangian functional $\mathcal{L}[\psi]$ is invariant under the action of the gauge group G:

$$\mathcal{L}[U\psi] = \mathcal{L}[\psi]\,. \tag{69.2}$$

Generally, this gauge invariance can be achieved in the following way. Denote the Lie algebra corresponding to G by $\mathfrak{g}$. Because the gauge transformation is local, the information on how it varies from point to point should be able to propagate through space(time). This is done by introducing a new (local) field $A(x)$, called the **gauge field**. The most elegant formulation uses the concept of covariant derivatives.

Definition 69.1.1 (Covariant derivative). When gauging a symmetry group, the ordinary partial derivatives are replaced by the covariant derivative

$$\mathrm{D}_\mu = \partial_\mu + igA_\mu(x)\,, \tag{69.3}$$

where $A_\mu : M \to \mathfrak{g}$ are the components of the ($\mathfrak{g}$-valued) gauge field (often expressed through coefficients A^a_μ with respect to a chosen basis of $\mathfrak{g}$). This procedure is called **minimal coupling** (conform Definition 57.7.1). It should be noted that the explicit action of the covariant derivative depends on the chosen representation of $\mathfrak{g}$ (or G) on $\mathcal{H}$. Furthermore, one should pay attention to the fact that the physicist convention was used, where one multiplies the gauge field A by a factor ig.[1]

[1] The imaginary unit turns anti-Hermitian fields into Hermitian fields.

So, to achieve gauge invariance, one should replace all derivatives by covariant derivatives. However, for this to be a well-defined operation, one should check that the covariant derivative itself satisfies the local gauge principle, i.e. $\mathsf{D}'\psi' = U\mathsf{D}\psi$ (from here on the coordinate-dependence of all fields will be suppressed).

$$\begin{aligned} U^{-1}\left(\frac{\partial}{\partial x^\mu} + igA'_\mu\right)\psi' &= U^{-1}\left(\frac{\partial}{\partial x^\mu} + igA'_\mu\right)U\psi \\ &= U^{-1}\frac{\partial U}{\partial x^\mu}\psi + \frac{\partial \psi}{\partial x^\mu} + igU^{-1}A'_\mu U\psi\,. \end{aligned} \tag{69.4}$$

This expression can only be equal to $\mathsf{D}\psi$ if

$$igA_\mu = U^{-1}\frac{\partial U}{\partial x^\mu} + igU^{-1}A'_\mu U\,, \tag{69.5}$$

which can be rewritten as

$$A'_\mu = UA_\mu U^{-1} - \frac{1}{ig}(\partial_\mu U)U^{-1} \tag{69.6}$$

or, in coordinate-independent form, as

$$A' = UAU^{-1} - \frac{1}{ig}\mathsf{d}UU^{-1}\,. \tag{69.7}$$

Up to scale conventions, this is exactly the content of Eq. (35.42) and Eq. (35.44) appearing in the study of principal connections. This should not come as a surprise, since the physical fields are sections of associated vector bundles and, hence, the principal bundle structure lurks in the background. Adding interactions is mathematically equivalent to coupling the space(time) manifold to a principal bundle.

Example 69.1.2 (QED). For quantum electrodynamics, which has $\mathsf{U}(1) \cong S^1$ as its gauge group, the parametrization $U(x) = e^{iq\chi(x)}$ is used with $\chi : \mathbb{R}^n \to \mathbb{R}$. Minimal coupling is given by

$$\partial_\mu \longrightarrow \mathsf{D}_\mu = \partial_\mu + iqA_\mu\,, \tag{69.8}$$

under the consistency condition

$$A_\mu \longrightarrow A'_\mu = A_\mu - \partial_\mu\chi\,, \tag{69.9}$$

where A_μ is the classic electromagnetic potential. These are the formulas introduced in Chapter 55.

69.1.2 Gauge fixing

Definition 69.1.3 (Gauge-fixing condition). Consider the space of all gauge connections $\mathcal{A}$. The group of gauge transformations $\mathcal{G} \equiv \mathsf{Aut}_V(P)$ acts on $\mathcal{A}$ and the quotient

space $\mathcal{A}/\mathcal{G}$ is called the **moduli space of gauge connections**. When calculating, for example, the path integral, it is important to select one representative connection from each orbit.

Algebro-differential conditions of the form

$$\Theta(A) = 0\,, \tag{69.10}$$

for a **gauge-fixing function** $\Theta : \mathcal{A} \to \mathbb{C}$, are called a gauge-fixing conditions.

Example 69.1.4 (Temporal gauge[2]). On any globally hyperbolic manifold $M \cong \mathbb{R} \times \Sigma^{n-1}$, the gauge field can be transformed to the temporal gauge, i.e. $A_0 = 0$.

Equation (69.6) gives the ODE

$$\partial_0 U = -A_0 U \tag{69.11}$$

for the required gauge transformation, which, similar to (35.54), can be integrated to obtain

$$U(t,x) = \mathcal{T}\exp\left(i\int_0^t A_0(t',x)\,dt'\right). \tag{69.12}$$

Property 69.1.5. The temporal gauge still leaves some residual gauge freedom. It is preserved by any time-independent gauge transformation.

The following two gauge-fixing conditions are very similar and both are common in the literature.

Example 69.1.6 (Lorenz gauge).

$$\partial_\mu A^\mu = 0 \tag{69.13}$$

Example 69.1.7. Consider a globally hyperbolic manifold $M \cong \mathbb{R} \times \Sigma^{n-1}$.

$$\partial_i A^i = 0 \tag{69.14}$$

Note that the Coulomb gauge only looks at the 'spatial components' of the connection.

Definition 69.1.8 (Gribov ambiguity). Given a gauge-fixing function $\Theta : \mathcal{A} \to \mathbb{C}$, the level set $\Theta^{-1}(0)$ does not necessarily intersect the orbits of the moduli space $\mathcal{A}/\mathcal{G}$ exactly once. It can occur that some orbits are intersected multiple times or that orbits are never reached. These situations, especially the former, are called Gribov ambiguities.

In the path integral, gauge freedom is usually accounted for through the Fadeev–Popov procedure (but this issue persists in the BRST treatment). However, when Gribov ambiguities are present, this procedure still overcounts the number of inequivalent solutions. @@ DISCUSS possible solutions (Gribov horizon, fundamental region, Gribov–Zwanziger framework, ...) + work by N. Vandersickel @@

[2] Also known as the **Weyl gauge**.

The following important result is due to *Singer* and shows how generic Gribov ambiguities acutally are.

Property 69.1.9. If the gauge group G is non-Abelian and attention is restricted to gauge fields that extend to the one-point compactification of spacetime, any gauge-fixing condition suffers from Gribov ambiguities.

69.1.3 Anomalies

Definition 69.1.10 (Anomaly). A classical symmetry that is broken when passing to quantum theory (either due to radiative corrections or nonperturbative corrections).

@@ COMPLETE (e.g. ABJ, Witten/SU(2)) @@

69.2 Yang–Mills theory

69.2.1 Yang–Mills equations

The general Yang–Mills action for a compact gauge group G (in vacuum) reads as follows:

$$S_{\mathrm{YM}}[A] := -\frac{1}{4}\langle F, F\rangle = -\frac{1}{4}\int_M \mathrm{tr}(F \wedge *F) = -\frac{1}{4}\int_M F_{\mu\nu}F^{\mu\nu} \tag{69.15}$$

for some curvature 2-form $F \in \Omega^2(M; \mathrm{ad}(P))$ as defined through Formula 35.4.17. In the above equation, $\langle\cdot,\cdot\rangle$ combines integration over M and the Killing form (Definition 32.4.1) on $\mathfrak{g}$.

The variational principle leads to the following equations of motion:

$$\mathrm{D}_A * F = 0\,, \tag{69.16}$$

where D_A is the exterior covariant derivative (Section 35.4.2). WIth the notation $\tilde{F} := *F$ (common in physics), the Yang–Mills equations can be written more explicitly as

$$\mathrm{d}\tilde{F} + A \barwedge \tilde{F} = 0\,. \tag{69.17}$$

By taking the Hodge dual of this expression, the following local coordinate expression can be found:

$$\partial_\mu F^{\mu\nu} + [A_\mu, F^{\mu\nu}] = 0\,. \tag{69.18}$$

Example 69.2.1 (QED). Consider the case $G = \mathrm{U}(1)$ as in the previous section. Since this group is Abelian, the second term in the Yang–Mills equation vanishes and one obtains

$$\partial_\mu F^{\mu\nu} = 0\,, \tag{69.19}$$

Together with the second Bianchi identity (Property 35.4.24), this gives the Maxwell equations (Section 55.3) in disguise, conform Section 55.4.3.

@@ COMPLETE @@

69.2.2 Currents and matter coupling

For most purposes (such as the Standard Model), one wants to couple the gauge fields to matter. Matter fields are given by sections of associated bundles: $\phi \in \Gamma(P \times_\rho V)$. The kinematic matter Lagrangian is given by

$$S_{\text{matter}}[\phi] := \langle \mathsf{D}\phi, D\phi \rangle \,. \tag{69.20}$$

The resulting Euler–Lagrange equation is a generalization of the Laplace equation:

$$\mathsf{D} * \mathsf{D}\phi \,. \tag{69.21}$$

If the full Lagrangian $\mathcal{L}[A, \phi] := \mathcal{L}_{\text{YM}}[A] + \mathcal{L}_{\text{matter}}[\phi]$ is used, the Yang–Mills equation (69.16) gets modified by a matter term:

$$\mathsf{D} * F + *[A, *\mathsf{D}\phi] = 0 \,. \tag{69.22}$$

These are also sometimes called the **Yang–Mills–Higgs equations**.

@@ ADD gerbes for magnetic currents @@

69.2.3 2 dimensions

In 2D, Yang–Mills theory has some interesting properties. Note that the action still makes sense in 2D, F is a 2-form and $*F$ is a 0-form. Using the results for Gaussian integrals (Section 16.6.3), it can be shown that the following action gives the same partition function (and, hence, the same physics) as the 2D Yang–Mills action:

$$S[\phi, A] := -\frac{g^2}{4} \int_M \langle \phi, \phi \rangle_{\mathfrak{g}} \, \mathsf{Vol}_M - i \int_M \langle \phi, F \rangle_{\mathfrak{g}} \,, \tag{69.23}$$

where $\phi \in \Omega^0(M; \mathfrak{g})$ is called the **background field**. This first-order formulation of 2D Yang–Mills theory as a background field (BF) theory shows that it is quasi-topological, i.e. topological when the coupling $g \longrightarrow 0$. Even though the Yang–Mills term itself depends on the metric and, hence, is not diffeomorphism-invariant, the second term is entirely topological. In this BF theory, the background field ϕ acts as a Lagrange multiplier and leads, as for Chern–Simons theory, to flat connections.

69.3 Spontaneous symmetry breaking

Theorem 69.3.1 (Goldstone). *Consider a field theory with gauge group G and denote the generators of the corresponding Lie algebra by X_a. Generators that do not annihilate the vacuum, i.e. $X_a v \neq 0$, or, equivalently, transformations that leave the vacuum invariant, correspond to massless scalar particles.*

The massless bosons from this theorem are called **Goldstone bosons**.

Theorem 69.3.2 (Elitzur). *The only operators in a lattice gauge theory*[3] *with a nonvanishing VEV are those invariant with respect to local gauge transformations.*

Corollary 69.3.3. Gauge symmetries cannot be spontaneaously broken.

69.3.1 Higgs mechanism

In Property 35.1.21, the equivariant maps corresponding to global sections of a principal bundle were called Higgs fields. In this section, a clarification for this terminology is given. The **Higgs vacuum** of a G-gauge theory, described by a principal bundle P, with a G-invariant potential V is given by the solutions of the following equations:

$$V(\phi) = 0 \tag{69.24}$$

$$\nabla\phi = 0\,, \tag{69.25}$$

where ∇ is a covariant derivative on P and ϕ is a section of some associated (finite-rank) vector bundle $P \times_\rho \mathcal{H}$. If the space of solutions $\mathcal{M}$ to the first equation admits a transitive G-action, i.e. it is a homogeneous space, then, by Property 2.3.14, one can write

$$\mathcal{M} \cong G/H\,, \tag{69.26}$$

where H is the isotropy group of any given solution. More generally, when the action is not transitive, the solution manifold is still the union of G-orbits, all of the form G/H with H the isotropy group of a point in the orbit.

Now, consider a specific choice of vacuum $m_0 \in \mathcal{M}$. If the whole theory were to be G-invariant, like the potential V, this corresponds to an equivariant map

$$\phi : P \to \mathcal{M}_0 \cong G/H\,, \tag{69.27}$$

where $\mathcal{M}_0$ is the orbit of m_0. This field is called the **Higgs field** in the physics literature. (For this reason, all such equivariant morphism and their associated sections are called Higgs fields.) The specific choice of vacuum, which generically has the smaller

[3] A proof for continuum field theories does not exist. However, since nonperturbative field theories are usually constructed through a limit procedure of lattice theories, this does have an impact.

symmetry group H, induces by Property 35.5.7 a reduction of the structure group from G to H and, consequently, the symmetry group is said to be **broken** to H.

After reduction, the G-connection can locally be decomposed as follows:

$$\iota^*\omega_{\mathfrak{g}} = \omega_{\mathfrak{h}} + \gamma\,, \tag{69.28}$$

where $\iota : P_H \hookrightarrow P$ denotes the reduction morphism and γ is a tensorial $(\mathrm{Ad}_H, \mathfrak{m})$-form with $\mathfrak{m}$ the complement of $\mathfrak{h}$ in $\mathfrak{g}$.[4]

For the Higgs field $\phi : P \to \mathcal{M}_0$ and, in fact, for any equivariant map $\phi : P \to \mathcal{M}_0$ such that $\nabla^H(\phi \circ \iota) = 0$, the covariant derivative satisfies

$$\nabla_X \phi = (\rho_{e,*} \circ \gamma)(X) m_0\,. \tag{69.29}$$

The generators $\rho_{e,*}(\gamma^i_\mu)m_0$, for $i = 1, \dots, \dim(\mathfrak{m})$, are called the **(Nambu–)Goldstone bosons**. Since $\dim(\mathfrak{m}) = \dim(G) - \dim(H)$, there are $\dim(G) - \dim(H)$ Goldstone fields. As seen above, after reduction, the connection form (gauge field) splits into a connection form for the smaller symmetry group and a set of new (massive) fields. The new connection form is obtained by trivially extending $\omega_{\mathfrak{h}}$ to a connection on P through G-equivariance. For such connections, Property 35.5.10 implies that $\nabla\phi = 0$ (this also follows from the expression above since γ vanishes for this kind of connection). This is exactly the second condition for the Higgs vacuum.

Now, what about Elitzur's theorem 69.3.2? If its generalization to field theories holds, the above considerations should not hold. Two solutions exist:

1. Realize that the Higgs mechanism can be restated without symmetry breaking.
2. Realize that the symmetry breaking applies to a global symmetry group and not a local one.

Although the first option is a very interesting approach, only the second one will be covered here. @@ MIGHT ADD FIRST OPTION TOO @@

The crucial point is that the group being broken is the ‚residual symmetry group, i.e. the symmetry group that remains after gauge-fixing the theory. When fully fixing a gauge, this residual group coincides with the center of the gauge group G, e.g. $\mathrm{U}(1)$ for $\mathrm{U}(1)$ or $\mathbb{Z}_n$ for $\mathrm{SU}(n)$. Consider for simplicity the typical Mexican hat potential in Yang–Mills theory:

$$\mathcal{L} := -\frac{1}{2}\mathrm{tr}\big(F_{\mu\nu}F^{\mu\nu}\big) + |D_\mu\phi|^2 - V(\phi)\,, \tag{69.30}$$

[4] To make Ad_H a well-defined representation on $\mathfrak{m}$, the latter is usually constructed as an orthogonal complement with respect to an Ad-invariant metric on $\mathfrak{g}$. In general, the pair $(\mathfrak{g}, \mathfrak{h})$ is required to be reductive (Definition 35.8.5).

where $V(\phi) := m^2|\phi|^2 + \lambda|\phi|^4$ with $\lambda > 0$ (note that $m^2 < 0$ is required for symmetry breaking, i.e. the mass term is required to be 'tachyonic'). The minimum of this potential is achieved for fields of modulus

$$|\phi|^2 = -\frac{m^2}{\lambda} \equiv \frac{v}{\sqrt{2}}\,. \tag{69.31}$$

The states within the level set $V^{-1}(v)$ can be parametrized as follows:

$$\phi(x) = \frac{v}{\sqrt{2}} e^{i\pi(x)/v}\,, \tag{69.32}$$

where $\pi(x)$ represents the (massless) Nambu–Goldstone boson. Inserting this expression into the Lagrangian gives an expression where the gauge fields A_μ are replaced by $A_\mu - \frac{1}{qv}\partial_\mu \pi$, where q is the charge factor. To get to the Lagrangian of a massive gauge field, the unitary gauge is chosen at this point:

$$A_\mu - \frac{1}{qv}\partial_\mu \pi \longrightarrow W_\mu\,. \tag{69.33}$$

This gives

$$\mathcal{L} \sim -\frac{1}{2}\mathrm{tr}\big(F_{\mu\nu}F^{\mu\nu}\big) + (qv)^2 W_\mu W^\mu\,. \tag{69.34}$$

However, it should be noted that the gauge-fixed field W_μ actually has a gauge-invariant representation:

$$W_\mu = \frac{i}{q}\widetilde{\phi}^* D_\mu \widetilde{\phi}\,, \tag{69.35}$$

with

$$\widetilde{\phi} := \frac{\phi}{|\phi|}\,. \tag{69.36}$$

@@ COMPLETE @@

69.4 Instantons

69.4.1 Instanton sectors

Definition 69.4.1 (Instanton). A Yang–Mills solution with finite Euclidean action.

To attain finite Euclidean energy, the solution should asymptotically be equal to the vacuum, i.e. the field strength should vanish at infinity and, hence, the solutions are asymptotically (locally) 'pure gauge':

$$A_\mu = ig(x)\partial_\mu g^{-1}(x) \tag{69.37}$$

for some smooth function $g : \mathbb{R}^n \to G$. Now, since F should vanish at infinity to obtain a finite (Euclidean) action, it extends to a curvature 2-form on the one-point compactification S^4. By the clutching construction (Remark 33.2.4), the underlying principal bundles are classified by maps $\chi : S^{n-1} \to G$ up to homotopy or, equivalently, by $\pi_{n-1}(G)$. For example, when working with $G = \mathsf{SU}(2) \cong S^3$ over Minkowski spacetime, the vacua are classified by their first Pontryagin number (which coincides with the second Chern number for traceless Lie algebras) since $\pi_3(S^3) \cong \mathbb{Z}$, which, in this context, is also called the **instanton number** $k \in \mathbb{Z}$.[5] In fact, the instanton number can also be calculated through the clutching function χ (this is essentially the Chern–Simons construction from Section 35.6.5):

$$k = -\frac{1}{24\pi^2}\int_{S^3} \mathrm{tr}\big(\chi \mathrm{d}\chi^{-1} \wedge \chi \mathrm{d}\chi^{-1} \wedge \chi \mathrm{d}\chi^{-1}\big)\,. \tag{69.38}$$

Remark 69.4.2 (Literature). In the literature, one can often find 'physical' explanations for the clutching construction where the 'spacetime boundary' at infinity is considered, often also working in the temporal gauge to remove the time dependence. It should be noted that that is only a rough analogy and should not be taken literally.

69.4.2 Large gauge transformations

Definition 69.4.3 (Topological vacuum). A Yang–Mills solution where $F = 0$ everywhere. All vacua belong to the zeroth instanton sector.

As stated in Definition 35.1.9, the gauge transformations in a general gauge theory are given by the vertical automorphisms of the underlying principal bundle. When quantizing the theory as in Section 59.5, one always starts with a constraint algebra. By Property 32.1.4 and Corollary 32.2.26, however, the exponential map only generates the identity component of the full gauge group. It follows that only the transformations in this connected component give rise to physically equivalent states. The gauge transformations not homotopic to the identity are called **large gauge transformations**.

Even though all vacua belong the same instanton sector, there are topological distinctions. Locally, vacua admit the form

$$A_\mu = ig(x)\partial_\mu g^{-1}(x) \tag{69.39}$$

as remarked for instantons at infinity. As such, one can also calculate their winding number through Eq. (69.38). If the total gauge group were connected, the winding number of a vacuum would be fixed. However, in general, large gauge transformations can map vacua with different winding numbers into each other.

[5] In fact, this holds for all $\mathsf{SU}(N)$ with $N \geq 2$ since the homotopy groups $\pi_3(\mathsf{SU}(N))$ stabilize at $N = 2$.

Because the topological vacua are not gauge invariant, they are actually not proper vacua (the dinstinction is merely a remnant from the nonconnectedness of gauge symmetries after all). The solution is to take a coherent superposition:

$$|\theta\rangle := \sum_{k\in\mathbb{Z}} e^{in\theta}|0\rangle_k \,, \tag{69.40}$$

with $\theta \in [0, 2\pi[$. These gauge-invariant vacua are called **θ-vacua**. It should be noted that states with different values of θ belong to different superselection sectors.

The different topological vacua can actually also be connected by instanton solutions, which are, here, related to tunneling amplitudes in path integral calculations. Consider a chunk of 4-dimensional spacetime of the form $[-T, T]\times S^3$ for some $T \in \mathbb{R}_0$.[6] Through Chern–Simons theory, the instanton number of a solution A on this manifold can be calculated as

$$k \sim \int_{[-T,T]\times S^3} \mathrm{dCS}(A_-, A_+) = \int_{\{T\}\times S^3} \mathrm{CS}(A_+) - \int_{\{-T\}\times S^3} \mathrm{CS}(A_-) \,. \tag{69.41}$$

If the initial and final states (the S^3's) represent vacua, the gauge fields are locally pure gauge, i.e.

$$A_\pm = i\chi_\pm \mathrm{d}\chi_\pm^{-1} \,, \tag{69.42}$$

the Chern–Simons forms give the winding numbers of the 'clutching functions' $\chi_\pm : S^3 \to G$:

$$k = k_+ - k_- \,. \tag{69.43}$$

The instanton (with instanton number $k \in \mathbb{Z}$) on the cylinder can, hence, be seen as interpolating between spatial slices of vacua of winding numbers k_- and $k_+ = k_- + k$, respectively. (This interpretation will be recalled in Section 69.5.)

Example 69.4.4 (BPST instanton). Instantons that connect vacua $|0\rangle_n$ to $|0\rangle_{n\pm1}$ and, accordingly, have winding number 1, are called **Belavin–Polyakov–Schwarz–Tyupkin (BPST) instantons**.

@@ COMPLETE (e.g. n-vacua, BPST solution) @@

69.4.3 Selfduality

Consider the Yang–Mills equations (in vacuum):

$$\mathrm{D}F = 0$$
$$\mathrm{D} * F = 0 \,.$$

[6] In *TQFTs* (see Section 72.1), this can be interpreted as a propagator between spaces of states on spatial slices S^3.

The first equation, the Bianchi identity, holds for all connection 1-forms and, hence, any (anti)selfdual connection (Definition 23.4.33) is a solution to the Yang–Mills equations. In fact, one has more.

Property 69.4.5 (Minimizers). Consider $G = \mathsf{SU}(N)$. Within a topological sector, i.e. for a fixed second Chern number

$$c_2(F) \sim \frac{\mathsf{tr}(F \wedge F)}{8\pi^2} = \frac{\mathsf{tr}(F^+ \wedge F^+)}{8\pi^2} - \frac{\mathsf{tr}(F^- \wedge F^-)}{8\pi^2}, \tag{69.44}$$

the Yang–Mills action is minimized for (anti)selfdual connections, since

$$F \wedge *F = (F^+ \wedge *F^+) + (F^- \wedge *F^-),$$

where the notation of Definition 23.4.33 is used. For negative $c_2(F)$, the antiselfdual connections are minimizers, whereas for positive $c_2(F)$ the selfdual connections are minimizers. For $c_2(F) = 0$, one recovers the flat connections.

69.4.4 't Hooft–Polyakov instantons

The definition of instantons can, of course, be generalized to gauge theories with matter couplings (or, in fact, to any Lagrangian theory). Consider Yang–Mills–Higgs theory, i.e. ordinary Yang–Mills theory coupled to a Higgs field $\phi : M \to V$. The Lagrangian of this theory reads as follows:

$$S_{\text{YMH}} = S_{\text{YM}} + S_{\text{Higgs}} = -\frac{1}{4}\langle F \wedge *F\rangle + \langle \nabla\phi, \nabla\phi\rangle + V(\phi). \tag{69.45}$$

As mentioned before, instantons are defined as the finite-action (in Euclidean signature) solutions of the equations of motions. This is equivalent to finite-energy solutions.

69.4.5 Dyons

Definition 69.4.6 (Dyon). A particle carrying both an electric and a magnetic charge.

Property 69.4.7 (Schwinger–Zwanziger condition). Recall Dirac's quantization condition for the electric charge. @@ This condition is not present in the document yet. (Only reference in prequantization.) @@ It can be extended to dyons as follows:

$$q_1 g_2 - g_1 q_2 \in 2\pi\mathbb{Z}. \tag{69.46}$$

This condition implies the following constraints on the charges:

1. The magnetic charges are quantized.
2. The electric charges differ by an integral multiple of the minimal electric charge e.

@@ ADD (e.g. Witten effect) @@

Now, recall the definition of θ-vacua in Eq. (69.40). These give the vacuum for a Yang–Mills Lagrangian to which the (second) Chern number has been added:

$$S_\theta := S_{\mathrm{YM}} + \frac{\theta}{16\pi^2}\langle F \wedge F\rangle\,. \tag{69.47}$$

The problem here is that this extra (topological) term is not invariant under, for example, inversions (i.e. parity transformations), unlike the standard Yang–Mills term $\mathrm{tr}(F \wedge *F)$. The same issue holds for charge conjugation. The only situation where no **CP-problem** occurs, is when $\theta = 0$. In nature, however, no such violation is observed and, hence, a fine-tuning problem arises (the magnitude of θ is experimentally bounded by 10^{-10})!

However, assuming the topological Chern term is included, the Schwinger–Zwanziger is modified.

Property 69.4.8 (Witten effect). When the CP-violating θ-term is added to the Lagrangian, the quantization condition for the electric charge is modified as follows:

$$q = ne + \frac{e\theta}{2\pi}N \tag{69.48}$$

for some $n, N \in \mathbb{Z}$. In particular, when $\theta \neq 0$, the electric charges are integral multiples of the minimal charge e.

69.4.6 Electric-magnetic duality

@@ ADD (see blog post) @@

69.5 Chern–Simons theory

The Chern–Simons action on a 3-manifold Σ is given by integration of the Chern–Simons form (Section 35.6.5 and Example 35.6.20 in particular):

$$S_{\mathrm{CS}}[A_\mu] := \frac{k}{4\pi}\int_\Sigma \mathrm{tr}\left(\mathrm{d}A \wedge A + \frac{2}{3}A \wedge A \wedge A\right)\,. \tag{69.49}$$

69.6 Quantization

69.6.1 Auxiliary fields

Chapter 70

Quantum Field Theory

The main reference for this chapter is Peskin and Schroeder (1995). For the section on the Connes–Kreimer algebra, see the original paper by Connes and Kreimer (1998) or the paper by Ebrahimi-Fard and Kreimer (2005). Another good reference on the formal structure of Feynman diagrams is Yeats (2017). The main references for the section on entanglement are Rangamani and Takayanagi (2017); Tuybens (2017).

Chapter Contents

70.1 Klein–Gordon field **1176**

70.1.1 Lagrangian and Hamiltonian 1176

70.1.2 Raising and lowering operators 1176

70.1.3 Scalar propagator 1178

70.1.4 Normalization constant 1178

70.1.5 Invariant integration measure 1178

70.2 Contractions and Wick's theorem **1179**

70.2.1 Bosonic fields 1179

70.2.2 Fermionic fields 1180

70.3 Gauge theory **1181**

70.3.1 Schwinger–Dyson equation 1181

70.3.2 Faddeev–Popov integration 1183

70.3.3 BV-BRST theory 1184

70.4 Quantum chromodynamics **1186**

70.5 Feynman rules **1186**

70.5.1 Scalar theory 1186

70.5.2 Graph theory 1187

70.5.3 Degree of divergence 1188
70.5.4 Amplituhedron ♣ 1190
70.6 Renormalization **1190**
70.6.1 Introduction: Statistical physics 1190
70.6.2 Dimensional regularization 1190
70.6.3 Wilsonian renormalization 1192
70.6.4 Connes–Kreimer renormalization 1192
70.7 Entanglement in QFT **1196**
70.7.1 Lattice theories 1196

70.1 Klein–Gordon field

70.1.1 Lagrangian and Hamiltonian

The 'simplest' Lagrangian (density) is given by

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2 \,. \tag{70.1}$$

Using the principle of least action, the following Euler–Lagrange equation is obtained:

$$\left(\partial^\mu\partial_\mu + m^2\right)\phi = 0 \,. \tag{70.2}$$

This can be rewritten using the **d'Alembertian** $\Box = \partial_\mu\partial^\mu$:

$$(\Box + m^2)\phi = 0 \,. \tag{70.3}$$

This equation is called the **Klein–Gordon equation**. In the limit $m \longrightarrow 0$, this equation reduces to the well-known wave equation (Formula 21.1.3).

From the Lagrangian (70.1), one can also derive a Hamiltonian function using Definition 52.2.3 and Definition 52.3.6:

$$H = \frac{1}{2}\int_{\mathbb{R}^3}\left[\pi^2(x) + (\nabla\phi(x))^2 + m^2\phi^2(x)\right] d^3x \,. \tag{70.4}$$

70.1.2 Raising and lowering operators

Fourier transforming the scalar field $\phi(\mathbf{x})$ and inserting it into the Klein–Gordon equation gives

$$\left(\partial_t^2 + p^2 + m^2\right)\phi(\mathbf{p}) = 0 \,. \tag{70.5}$$

This is the equation for a simple harmonic oscillator with frequency $\omega = \sqrt{p^2 + m^2}$.

Analogous to ordinary quantum mechanics, raising and lowering operators $a^\dagger_{\vec{\mathbf{p}}}$ and $a_{\vec{\mathbf{p}}}$ can be defined such that

$$\phi(\vec{\mathbf{x}}) = \int_{\mathbb{R}^3} \frac{d^3p}{(2\pi)^{3/2}} \frac{1}{\sqrt{2\omega_{\vec{\mathbf{p}}}}} \left(a_{\vec{\mathbf{p}}} e^{i\vec{\mathbf{p}}\cdot\vec{\mathbf{x}}} + a^\dagger_{\vec{\mathbf{p}}} e^{-i\vec{\mathbf{p}}\cdot\vec{\mathbf{x}}}\right), \tag{70.6}$$

$$\pi(\vec{\mathbf{x}}) = \int_{\mathbb{R}^3} \frac{d^3p}{(2\pi)^{3/2}} (-i)\sqrt{\frac{\omega_{\vec{\mathbf{p}}}}{2}} \left(a_{\vec{\mathbf{p}}} e^{i\vec{\mathbf{p}}\cdot\vec{\mathbf{x}}} - a^\dagger_{\vec{\mathbf{p}}} e^{-i\vec{\mathbf{p}}\cdot\vec{\mathbf{x}}}\right). \tag{70.7}$$

An equivalent definition is obtained by performing the transformation $\vec{\mathbf{p}} \longrightarrow -\vec{\mathbf{p}}$ in the second term of $\phi(\vec{\mathbf{x}})$ and $\pi(\vec{\mathbf{x}})$:

$$\phi(\vec{\mathbf{x}}) = \int_{\mathbb{R}^3} \frac{d^3p}{(2\pi)^{3/2}} \frac{1}{\sqrt{2\omega_{\vec{\mathbf{p}}}}} \left(a_{\vec{\mathbf{p}}} + a^\dagger_{-\vec{\mathbf{p}}}\right) e^{i\vec{\mathbf{p}}\cdot\vec{\mathbf{x}}}, \tag{70.8}$$

$$\pi(\vec{\mathbf{x}}) = \int_{\mathbb{R}^3} \frac{d^3p}{(2\pi)^{3/2}} (-i)\sqrt{\frac{\omega_{\vec{\mathbf{p}}}}{2}} \left(a_{\vec{\mathbf{p}}} - a^\dagger_{-\vec{\mathbf{p}}}\right) e^{i\vec{\mathbf{p}}\cdot\vec{\mathbf{x}}}. \tag{70.9}$$

When the commutation relation

$$[a_{\vec{\mathbf{p}}}, a^\dagger_{\vec{\mathbf{q}}}] := \delta(\vec{\mathbf{p}} - \vec{\mathbf{q}}) \tag{70.10}$$

is imposed, the following commutation relation for the scalar field and its conjugate momentum is obtained:

$$[\phi(\vec{\mathbf{x}}), \pi(\vec{\mathbf{y}})] = i\delta(\vec{\mathbf{x}} - \vec{\mathbf{y}}). \tag{70.11}$$

Now, the Hamiltonian can be calculated explicitly:

$$H = \int_{\mathbb{R}^3} \frac{d^3p}{(2\pi)^3} \omega_{\vec{\mathbf{p}}} \left(a^\dagger_{\vec{\mathbf{p}}} a_{\vec{\mathbf{p}}} + \frac{1}{2}[a_{\vec{\mathbf{p}}}, a^\dagger_{\vec{\mathbf{p}}}]\right). \tag{70.12}$$

However, it is clear from Eq. (70.10) that the second term in this integral diverges. There are two reasons for this divergence. First, space is infinite, i.e. the d^3x integral in Eq. (70.4) diverges. This problem can be resolved by restricting the system to a (finite) part of space or by considering the energy density instead of the energy itself. Second, by including very large values for p in the integral, a parameter range is explored where the theory is likely to break down. To resolve this problem a 'high p'-cut-off should be introduced. A more practical solution, however, is to note that only energy differences are physical and so one can simply drop the second term altogether as it is merely a 'constant' (albeit an infinite one).

A corollary of Eq. (70.12) together with the canonical commutation relations is

$$[H, a^\dagger_{\vec{\mathbf{p}}}] = \omega_p a^\dagger_{\vec{\mathbf{p}}}, \tag{70.13}$$

$$[H, a_{\vec{\mathbf{p}}}] = -\omega_p a_{\vec{\mathbf{p}}}. \tag{70.14}$$

As was the case for the quantum harmonic oscillator, the creation and annihilation operators deserve their names and one can write:

$$\left|\vec{\mathbf{k}}_1, \dots, \vec{\mathbf{k}}_n\right\rangle = a^\dagger(\vec{\mathbf{k}}_1) \cdots a^\dagger(\vec{\mathbf{k}}_n)|0\rangle\,. \tag{70.15}$$

Furthermore, this equation together with the canonical commutation relations imply that the Klein–Gordon fields are bosonic fields. This way of generating the full Hilbert (or, in fact, Fock) state space is axiomatized by the GNS construction (Construction 26.2.41), where the vacuum plays the role of cyclic vector.

70.1.3 Scalar propagator

Formula 70.1.1 (Pauli–Jordan function).

$$i\Delta(\mathbf{x}-\mathbf{y}) := i[\phi(\mathbf{x}), \phi(\mathbf{y})] = \int_{\mathbb{R}^3} \frac{d^3p}{(2\pi)^3} \frac{1}{2\omega_{\mathbf{x}}} \left(e^{-i\mathbf{p}\cdot(\mathbf{x}-\mathbf{y})} - e^{i\mathbf{p}\cdot(\mathbf{x}-\mathbf{y})}\right) \tag{70.16}$$

When $x^0 = y^0$ (ETCR) or $(\mathbf{x}-\mathbf{y})^2 < 0$ (spacelike curves), the Pauli–Jordan function is identically 0. (See also the *axiom of microcausality* 72.2).

70.1.4 Normalization constant

By Property 19.1.23, the delta function $\delta^{(3)}(\vec{\mathbf{p}} - \vec{\mathbf{q}})$ transforms as $\delta^{(3)}(\Lambda\vec{\mathbf{p}} - \Lambda\vec{\mathbf{q}})\frac{\Lambda E}{E}$ under a general Lorentz boost Λ. Although this is clearly not Lorentz invariant, the quantity $E_{\mathbf{x}}\delta^{(3)}(\vec{\mathbf{p}}-\vec{\mathbf{q}})$ can be seen to be invariant. From this observation it follows, that the correct normalization in the momentum representation is

$$|\mathbf{p}\rangle = \sqrt{2E_{\mathbf{x}}}a^\dagger_{\mathbf{p}}|0\rangle \tag{70.17}$$

and, hence,

$$\langle \mathbf{p} \mid \mathbf{q}\rangle = 2E_{\mathbf{x}}(2\pi)^3\delta^{(3)}(\vec{\mathbf{p}} - \vec{\mathbf{q}})\,, \tag{70.18}$$

where the constants are a matter of convention to cancel the constants in Eq. (70.6).

70.1.5 Invariant integration measure

The factor $2E_p$ does not only occur in the normalization conditions. To find a Lorentz-invariant integration measure in spacetime, the following integral can be studied:

$$\int_{\mathbb{R}^3} \frac{d^3p}{2E_p} = \int_{\mathbb{R}^+\times\mathbb{R}^3} \delta(p^2 - m^2)\, d^4p\,, \tag{70.19}$$

where the timelike component p^0 is restricted to be (strictly) positive. By using this measure, it is ensured that the integral of any Lorentz-invariant function is again Lorentz invariant.

Example 70.1.2 (One-particle identity operator).

$$\mathbb{1} := \int_{\mathbb{R}^3} \frac{d^3p}{2E_p} |\mathbf{p}\rangle\langle\mathbf{p}| \tag{70.20}$$

70.2 Contractions and Wick's theorem

70.2.1 Bosonic fields

In the following definitions, (field) operators will be decomposed as

$$\phi = \phi^{(+)} + \phi^{(-)},$$

where the + symbol denotes the 'positive-frequency' part, i.e. the part consisting of annihilation operators. The 'negative-frequency' part is defined analogously. This terminology stems from classic Fourier theory. By looking at Eq. (70.6) and remembering that the (1, 3)-signature is adopted, the annihilators can be seen to always occur together with a positive-frequency exponential.

Definition 70.2.1 (Contraction for neutral bosonic fields).

$$\overbracket{\phi(\mathbf{x})\phi(\mathbf{y})} := \begin{cases} [\phi(\mathbf{x})^{(+)}, \phi(\mathbf{y})^{(-)}] & \text{if } x^0 > y^0 \\ [\phi(\mathbf{y})^{(+)}, \phi(\mathbf{x})^{(-)}] & \text{if } y^0 > x^0 \end{cases} \tag{70.21}$$

Formula 70.2.2 (Feynman propagator).

$$i\Delta_F(\mathbf{x}-\mathbf{y}) := \overbracket{\phi(\mathbf{x})\phi(\mathbf{y})} := i \lim_{\varepsilon\to 0^+} \int_{\mathbb{R}^4} \frac{d^4p}{(2\pi)^4} \frac{e^{-i\mathbf{p}\cdot(\mathbf{x}-\mathbf{y})}}{p^2 - m^2 + i\varepsilon} \tag{70.22}$$

Definition 70.2.3 (Contraction for charged bosonic fields).

$$\overbracket{\phi(\mathbf{x})\overline{\phi}(\mathbf{y})} := \begin{cases} [\phi(\mathbf{x})^{(+)}, \overline{\phi}(\mathbf{y})^{(-)}] & \text{if } x^0 > y^0 \\ [\phi(\mathbf{y})^{(+)}, \overline{\phi}(\mathbf{x})^{(-)}] & \text{if } y^0 > x^0 \end{cases} \tag{70.23}$$

Definition 70.2.4 (Normal ordering). The normal ordering $\mathcal{N}$, often denoted by colons $:\ :$, of a sequence of field operators is defined as the permuted sequence in which all annihilation operators appear on the right of the creation operators, e.g.:

$$\mathcal{N}\big(\phi(\mathbf{x})\phi^\dagger(\mathbf{y})\phi(\mathbf{z})\big) = \phi^\dagger(\mathbf{y})\phi(\mathbf{x})\phi(\mathbf{z}). \tag{70.24}$$

Note that the normal ordering operator is not a morphism between CCR-algebras since this would lead to a contradiction:

$$b_i b_j^\dagger = b_j^\dagger b_i + \delta_{ij} \implies \mathcal{N}(b_i b_j^\dagger) = \mathcal{N}(b_j^\dagger b_i + \delta_{ij}) = \mathcal{N}(b_j^\dagger b_i) + \delta_{ij} \implies \delta_{ij} = 0. \tag{70.25}$$

The solution is given by the fact that inside the normal ordering all operators commute and, hence, this ordering can be axiomatized as an algebra morphism $\mathcal{N} : \mathsf{Sym}^\bullet A \to A$, where A is the CCR-algebra of the theory (see Miwa, Jimbo, Jimbo, and Date (2000)).

@@ IS THIS REFERENCE CORRECT? @@

Property 70.2.5. From this definition, it immediately follows that the vacuum expectation value of a normal ordered sequence is 0.

Formula 70.2.6 (Wick's theorem for bosonic fields).

$$\mathcal{T}\big(\phi(\mathbf{x}_1)\cdots\phi(\mathbf{x}_n)\big) = \mathcal{N}\big(\phi(\mathbf{x}_1)\cdots\phi(\mathbf{x}_n) + \text{all possible contractions}\big) \tag{70.26}$$

When acting with a time-ordered product on the vacuum, this relation implies that only fully contracted terms will remain. Moreover, by Formula 70.2.2, every such action can be expressed solely in terms of propagators.

Remark 70.2.7. In the case of charged bosons, only contractions of the form $\overbracket{\phi(\mathbf{x})\phi}(\mathbf{y})$ will remain because $[a, b^+] = 0$.

Corollary 70.2.8.

$$\overbracket{\phi(\mathbf{x})\phi}(\mathbf{y}) = \mathcal{T}\big(\phi(\mathbf{x})\phi(\mathbf{y})\big) - \mathcal{N}\big(\phi(\mathbf{x})\phi(\mathbf{y})\big) \tag{70.27}$$

Wick's theorem has an analogue in probability theory.

Theorem 70.2.9 (Isserlis). *Let $X_1, \dots, X_n$ be a set of random variables following a multinormal distribution with mean 0.*

$$\mathsf{E}\big[X_1 \cdots X_n\big] = \sum_{\sigma \in P_{n,2}} \prod_{\{i,j\}\in\sigma} \mathsf{E}\big[X_i X_j\big], \tag{70.28}$$

where $P_{n,2}$ denotes the set of binary partitions of $\{1, \dots, n\}$. Because the mean of the distribution is zero, this expression is equal to

$$\mathsf{E}\big[X_1 \cdots X_n\big] = \sum_{\sigma \in P_{n,2}} \prod_{\{i,j\}\in\sigma} \mathsf{cov}\big[X_i, X_j\big]. \tag{70.29}$$

70.2.2 Fermionic fields

Definition 70.2.10 (Contraction).

$$\overbracket{\psi(\mathbf{x})\overline{\psi}}(\mathbf{y}) := \begin{cases} \{\psi(\mathbf{x})^{(+)}, \overline{\psi}(\mathbf{y})^{(-)}\}_+ & \text{if } x^0 > y^0 \\ -\{\psi(\mathbf{y})^{(+)}, \overline{\psi}(\mathbf{x})^{(-)}\}_+ & \text{if } y^0 > x^0 \end{cases} \tag{70.30}$$

Remark 70.2.11. Only contractions of the form $\overbracket{\psi(\mathbf{x})\overline{\psi}}(\mathbf{y})$ will remain because $\{a, b^\dagger\}_+ = 0$.

Formula 70.2.12 (Feynman propagator).

$$i\Delta_F(\mathbf{x}-\mathbf{y}) := \overbracket{\psi(\mathbf{x})\overline{\psi}}(\mathbf{y}) := i \lim_{\varepsilon\to 0^+} \int_{\mathbb{R}^4} \frac{d^4\mathbf{p}}{(2\pi)^4} \frac{\not{\mathbf{p}} + m}{p^2 - m^2 + i\varepsilon} e^{-i\mathbf{p}\cdot(\mathbf{x}-\mathbf{y})} \tag{70.31}$$

Remark 70.2.13 (Ordering operators). One should take into account the Fermi–Dirac statistics when permuting fermionic field operators under a normal ordering. A general factor $\mathsf{sgn}(\sigma)$, where σ is the permutation of the operators, will arise in every term, e.g.:

$$\mathcal{N}\big(\psi(\mathbf{x})\overline{\psi}(\mathbf{y})\psi(\mathbf{z})\big) = -\overline{\psi}(\mathbf{y})\psi(\mathbf{x})\psi(\mathbf{z})\,. \tag{70.32}$$

A similar remark should be made for the time-ordering operator $\mathcal{T}$. As was the case for bosonic theories, one should pay attention to the nature of the normal ordering. It is not a morphism between CAR-algebras, but instead it is an algebra morphism between a free (odd) algebra and a CAR-algebra.

70.3 Gauge theory

In this section, the theories from Chapter 69 and Chapter 54 will be quantized. Due to the gauge invariance and, hence, the redundancies in the fields, the path integral diverges. Several solutions exist, ranging from rather ad-hoc procedures (as described first) to very general methods. The general gauge structure of Lagrangian (field) theories was considered in Chapter 54. The resulting structure was the BV-BRST complex, which gives a homological resolution of both the critical locus of the field equations and the gauge orbits of the gauge symmetries. The main object was a gauge-invariant action and an extended covariant phase space containing both fields and antifields with the structure of a symplectic dg-manifold (Corollary 44.1.11). The quantum analogue of this structure will be discussed below.

70.3.1 Schwinger–Dyson equation

Consider the following path-integral representation of the expectation value of a (polynomially bounded) functional F:[1]

$$\langle F\rangle = \int F[\phi]\,\exp\big(i/\hbar S[\phi]\big)\,D\phi\,. \tag{70.33}$$

If the field measure $D\phi$ is an appropriate generalization of the Lebesgue/Haar measure on $\mathbb{R}^n$, it is invariant under translations $\phi \longrightarrow \phi+\xi$.[2] It follows that the expectation value is also given by

$$\langle F\rangle = \int F[\phi+\xi]\,\exp\big(i/\hbar S[\phi+\xi]\big)\,D\phi\,. \tag{70.34}$$

[1]The normalization factor has been left out for convenience.

[2]This is, for example, not the case for nonlinear σ-models, where the target space is a manifold and not a vector space.

Equating these two expressions and expanding to first order gives the Schwinger–Dyson equations:

$$\left\langle \frac{\delta^L F}{\delta \phi^I} \right\rangle = -i\hbar \left\langle \frac{\delta^L S}{\delta \phi^I} F \right\rangle \tag{70.35}$$

for all $I \in \mathcal{J}$. This can be recast in geometric form through Property 54.3.15. A polynomial functional that is linear in antifields is of the form $\lambda^I \mathcal{P}_I$. Geometrically, this corresponds to the vector field

$$X_\lambda = \lambda^I \frac{\delta^L}{\delta \phi^I} \,. \tag{70.36}$$

The Schwinger–Dyson equation then becomes

$$\mathcal{T}\left(\mathcal{L}_{X_\lambda} e^{i/\hbar S[\phi]}\right) = 0 \,. \tag{70.37}$$

Remark 70.3.1 (Nonperturbative QFT). The Schwinger–Dyson equations can be taken as the fundamental starting point of QFT for a nonperturbative approach.

Note that, even on shell, the time-ordered product $\mathcal{T}\left(F \frac{\delta S}{\delta \phi}\right)$ does not generally vanish. However, when F has a different spacetime support than ϕ, there is no operator-ordering issue and, accordingly, the expectation value does vanish. As such, the equations of motion can be seen to hold quantum mechanically away from the diagonal.

Property 70.3.2 (Observable equivalence). Classically, two functionals F, G are said to be equivalent on shell whenever

$$F - G \approx 0 \tag{70.38}$$

or, equivalently (Property 54.1.7),

$$F - G = \frac{\delta^L S}{\delta \phi^I} \lambda^I \tag{70.39}$$

Note that the DeWitt convention (54.86) is used here. Through the Schwinger–Dyson equation, this can be extended to QFTs:

$$F \sim G \iff F - G = \frac{\delta^L S}{\delta \phi^I} \lambda^I - i\hbar \frac{\delta^L \lambda^I}{\delta \phi^I} \,. \tag{70.40}$$

Two functionals in the same **quantum equivalence class** will have the same expectation value. Note that, for $\hbar \longrightarrow 0$, this recovers the classical situation. In this way, the Schwinger–Dyson equation also characterizes the kernel of the time-ordering operator $\mathcal{T}$ sending (classical) off-shell observables to (quantum) on-shell observables (cf. Remark 70.2.13).

In Section 54.3, it was shown how the classical equivalence of functionals can be implemented homologically through the BV-BRST complex. Of course, the question arises whether the quantum equivalence relation can also be resolved in a similar way. An essential property of the antibracket was that it constituted a (graded) derivation. However, because of quantum corrections, the quantum BRST operator $\mathcal{S}$ cannot be a derivation. The product of $\mathcal{S}$-exact will not, in general, be $\mathcal{S}$-exact, since $\langle A\rangle = 0 = \langle B\rangle$ does not imply $\langle AB\rangle = 0$ (in fact, it is $\langle AB\rangle \sim \hbar$).

70.3.2 Faddeev–Popov integration

Introducing ghost variables is not only useful for obtaining a resolution of the gauge orbits. It also helps with path integral quantization of gauge theories, where it removes divergences from the path integral. (Of course, this boils down to the same purpose, since the divergence arise from overcounting gauge orbits.)

First, consider a classical theory with second-class constraints. These can be taken care of by passing from the Poisson bracket to the Dirac bracket (Definition 54.1.16). With respect to a triangulation of the path integral, one, hence, has to replace the ordinary Liouville measure by a modified version:

$$\int \prod_{\mathbf{x}} \mathsf{Ber}(\sigma_{ij})^{1/2} \exp\big(i/\hbar S[\phi]\big)[D\phi]\,, \tag{70.41}$$

where σ_{ij} is the structure matrix of the Dirac bracket and Ber denotes the Berezinian (Definition 23.5.13). In terms of the constraint matrix $C_{ab} = \{\chi_a, \chi_b\}$, this becomes:

$$\int \prod_{a,\mathbf{x}} \delta(\chi_a)\mathsf{Ber}(C_{ab})^{1/2} \exp\big(i/\hbar S[\phi]\big)[D\phi]\,. \tag{70.42}$$

For first-class constraints, the approach is slightly different and one needs to construct the reduced phase space. (From now on, it is assumed that there are no second-class constraints left.) Consider a complete set of variables $\tilde{z}^\alpha$ with their associated structure matrix $\sigma^{\alpha\beta} := \{\tilde{z}^\alpha, \tilde{z}^\beta\}$. The reduced path integral is

$$\int \prod_{\mathbf{x}} \mathsf{Ber}(\sigma_{\alpha\beta})^{1/2} \exp\big(i/\hbar S[\tilde{z}]\big)[D\tilde{z}]\,, \tag{70.43}$$

Remark 70.3.3 (Gribov ambiguity). The reduced path integral is gauge invariant and, more importantly, requires no gauge-fixing conditions. Accordingly, it does not suffer from the Gribov problem (Definition 69.1.8).

To obtain the reduced variables $\tilde{z}$ is a difficult and, possibly, impossible task. However, sometimes, there exist gauge conditions $\chi_a = 0$ that carve out the reduced phase

space. The combined constraints $\chi_\alpha \equiv (G_a, \chi_a)$ are always second class. Moreover, the Berezinian satisfies

$$\mathrm{Ber}\big(\{\chi_\alpha, \chi_\beta\}\big) = \mathrm{Ber}\big(\{G_a, \chi_b\}\big)\,. \tag{70.44}$$

The reduced path integral can, hence, be rewritten as

$$\int \prod_{\mathbf{x}} \delta(\chi)\delta(G)\mathrm{Ber}\big(\{G_a, \chi_b\}\big) \exp\big(i/\hbar \tilde{S}[z]\big)[Dz]\,, \tag{70.45}$$

where $\tilde{S}$ contains the boundary term M in Eq. (54.26). Introducing Fourier transforms recovers (part of) Eq. (54.79):

$$\int \prod_{\mathbf{x}} \delta(\chi) \exp \frac{i}{\hbar}\left(\tilde{S}[z] - \int v^a G_a \,\mathrm{Vol}_M - \int \overline{C}^b \delta_C \chi_b \,\mathrm{Vol}_M\right)[Dz][Dv][DC][D\overline{C}]\,. \tag{70.46}$$

This is the **Faddeev–Popov path integral** and the $C, \overline{C}$ are the **Faddeev–Popov ghosts**.

70.3.3 BV-BRST theory

Definition 70.3.4 (BV Laplacian). Consider a BV manifold (M, ω, S). A BV Laplacian is an operator Δ that satisfies the following conditions:

1. **Nilpotency**: $\Delta^2 = 0$, and
2. **Product rule**: $\Delta(fg) = (\Delta f)g + (-1)^{\varepsilon(f)} f(\Delta g) + (-1)^{\varepsilon(f)}(f, g)$,

where $(\cdot, \cdot)$ is the induced Poisson/Gerstenhaber bracket. Although Δ is not a derivation with respect to ordinary multiplication — the product rule says that the bracket measures the deviation of Δ from being a derivation — these relations do imply that the BV Laplacian acts as a graded derivation with respect to the bracket:

$$\Delta(f, g) = (\Delta f, g) + (-1)^{\varepsilon(f)+1}(f, \Delta g)\,. \tag{70.47}$$

The canonical BV Laplacian in field-antifield coordinates is given by (the DeWitt convention is still used here)

$$\Delta_{\mathrm{BV}} := (-1)^{\varepsilon_I+1} \frac{\partial^R}{\partial \phi^I} \frac{\partial^R}{\partial \mathcal{P}_I} + (-1)^{\varepsilon_a+1} \frac{\partial^R}{\partial \eta^a} \frac{\partial^R}{\partial \mathcal{P}_a}\,. \tag{70.48}$$

For functionals that are linear in antifields[3], the BV Laplacian recovers the correction term in the Schwinger–Dyson equations:

$$\Delta\big(\mathcal{P}_I \lambda^I[\phi]\big) = (-1)^{\varepsilon(\lambda^I)-1} \frac{\delta^L \lambda^I}{\delta \phi^I}\,. \tag{70.49}$$

[3]Terms of higher order can be reduced to this case through the product rule.

Hence, the quantum-mechanical analogue of the BRST operator is

$$\mathcal{S}F := (F, S) - i\hbar\Delta F\,, \tag{70.50}$$

where S is the classical action. The BV Laplacian gives exactly the quantum corrections to the classical BRST operator.

Formula 70.3.5 (Schwinger–Dyson equations). For functionals that are linear in the antiields, i.e. vector fields on the BV manifold, Eq. (70.35) can be rewritten as

$$\langle \mathcal{S}F \rangle = 0\,. \tag{70.51}$$

For this reason, the quantum BRST operator is also called the **Schwinger–Dyson operator**.

Formula 70.3.6 (Quantum master equation). Consider the function $f := e^{iS/\hbar}$ on a BV manifold (M, ω, S). Because

$$\Delta_{\mathrm{BV}} f = \frac{i}{\hbar}\Delta S e^{iS/\hbar} + \left(\frac{i}{\hbar}\right)^2 \frac{1}{2}(S, S) e^{iS/\hbar}\,, \tag{70.52}$$

the condition that f is BV-harmonic is equivalent to S satisfying

$$\frac{1}{2}(S, S) - i\hbar\Delta_{\mathrm{BV}} S = 0\,. \tag{70.53}$$

This equation is called the quantum master equation. Expanding S as a power series in $\hbar$ shows that the order-0 term satisfies the classical master equation (44.8).

Definition 70.3.7 (BV integral). Consider the shifted cotangent bundle ΠT^*M on an m-dimensional smooth manifold. Let the gauge-fixing fermion K be an odd function of the even coordinates of a Darboux basis (denote these coordinates by q^i). This function determines a projectable Lagrangian submanifold $L_K \subset \Pi T^*M$ by the Maslow–Hörmander theorem 37.3.10 given by

$$L_K = \left\{ (q^i, p_i) \in \Pi T^*M \,\middle|\, p_i = \frac{\partial K}{\partial q^i} \right\}. \tag{70.54}$$

The Batalin–Vilkovisky integral of a function $f \in C^\infty(M)$ with respect to K is defined as

$$\int_{L_K} f \sqrt{\mathsf{Vol}_\omega}\,, \tag{70.55}$$

where $\sqrt{\mathsf{Vol}_\omega}$ is the volume form on Lagrangian submanifolds induced by ω. The BV integral and BV Laplacian interact in the following way (for BV integrable f):

1. If $f = \Delta_{\mathrm{BV}} g$, then $\int_{L_K} f = 0$ for all gauge-fixing fermions K.

2. If $\Delta_{\mathrm{BV}} f = 0$, then $\frac{\mathrm{d}}{\mathrm{d}t} \int_{L_{K_t}} f = 0$, where $\{K_t\}_{t\in\mathbb{R}}$ is a continuous family of gauge-fixing fermions.

The second property implies that the BV integral is invariant under deformations of the domain of integration.

Property 70.3.8 (Divergence). Consider a superdensity μ (or a volume form) on a BV manifold (any supermanifold actually suffices). This density induces a divergence on vector fields (i.e. antifields) by generalizing Eq. (36.19):

$$\mathrm{div}(F) := -\frac{1}{\mu}\frac{\delta^R}{\delta\phi^I}\left((-1)^{\varepsilon_I}\mu\frac{\delta^R F}{\delta\mathcal{P}_I}\right). \tag{70.56}$$

When $\mu = 1$, one obtains $\div(F) = \Delta F$, i.e. the BV Laplacian is exactly the divergence.

Remark 70.3.9 (de Rham differential). The BV Laplacian can also be obtained as the dual to the de Rham differential when identifying p-vector fields with $(n-p)$-forms through the volume form.

70.4 Quantum chromodynamics

Property 70.4.1 (OZI rule[4]). Decay processes for which the corresponding Feynman diagrams become disconnected (initial states and final states are disconnected) when removing internal gluon lines, are suppressed with respect to other processes.

@@ COMPLETE or REMOVE @@

70.5 Feynman rules

70.5.1 Scalar theory

By expanding the correlation functions in perturbation theory and applying Wick's theorem, one can rewrite every term using the following dictionary for Feynman diagrams (it is assumed that every coupling constant in the Lagrangian is divided by the necessary permutation factors, e.g. in ϕ^4-theory it is assumed that the constant is of the form $\lambda/4!$):

- Propagator $\Delta_F(\mathbf{x}-\mathbf{y})$:

[4] *Okubo, Zweig* and *Iizuka*

- Interaction vertex[5] $-i\lambda \int d^4z$:

The main idea behind these rules is to draw all possible diagrams consistent with the given interaction Lagrangian and translate them into analytic expressions. However, to obtain the correct normalization, one should take the following remark into account.

Remark 70.5.1. Symmetry factors of diagrams should be accounted for in analytic expressions. As an example, consider the following vacuum bubble:

Because the two legs can be interchanged, this diagram has **symmetry factor** 2 and, hence, it gives the analytic expression $-\frac{i\lambda}{2}\int d^4z\,\Delta_F(\mathbf{z}-\mathbf{z})$.

70.5.2 Graph theory

Given a decorated graph Γ, representing a Feynman diagram, collapsing all vertices leads to its residue $\mathrm{res}(\Gamma)$. Such a one-vertex graph $\underline{r}$, where $r \in \mathbb{N}$ is the number or remaining edges, is sometimes also called the **external leg structure** of all graphs Γ with $\mathrm{res}(\Gamma) = \underline{r}$. Two other important properties of a graph Γ are its degree and its symmetry factor.

Definition 70.5.2 (Degree). Consider a graph Γ. Its degree $|\Gamma| \in \mathbb{N}$ is given by the number of independent loops in Γ or, equivalently, the rank of its first homology group $H_1(\Gamma;\mathbb{Z})$ (Definition 7.2.16), i.e. its first Betti number.

Definition 70.5.3 (1PI graph). A graph that remains connected when an internal edge is removed.

Several different ways exist to determine the number of relevant permutations of the Feynman diagram. One is to directly use Wick's theorem and get the combinatorics out of that result. However, two other useful approaches exist.

Definition 70.5.4 (Symmetry factor). Consider a Feynman diagram Γ consisting of $n \in \mathbb{N}$ internal vertices of valence $\nu \in \mathbb{N}_{\geq 3}$.[6] The symmetry factor is given by:

$$\mathrm{sym}(\Gamma) := \frac{1}{n!}\left(\frac{1}{\nu!}\right)^n C\,, \tag{70.57}$$

[5] Four legs were drawn as an example but this can be generalized to any order of interaction term.

[6] It is here assumed that all vertices are of the same valence and type. The definition is easily extended to more complex diagrams.

where $C \in \mathbb{N}$ denotes the number of different ways that the vertices can be contracted while resulting in the same diagram.

A more structured way is to consider the suitable notion of graph automorphism for Feynman diagrams. For ordinary graphs, two common definitions are either through an *adjacency matrix* or through a relation on edges. A third, equivalent, way, however, is through the notion of 'half-edge'.

Alternative Definition 70.5.5 (Graph). Consider a set of **half-edges** Γ. A graph structure on Γ consists of a partition $\mathcal{E}$ by subsets of cardinality 2, called the **edge** set, and a partition $\mathcal{V}$, the **vertex** set.

By slightly modifying this definition, a Feynman diagram is obtained.

Alternative Definition 70.5.6 (Symmetry factor). Consider a set of **half-edges** Γ. A Feynman diagram structure on Γ consists of a disjoint collection $\mathcal{E}$ of cardinality-2 subsets, called the **internal edge** set, and a partition $\mathcal{V}$ by subsets of cardinality at least 3, the **vertex** set. Half-edges not part of the internal edge set are called external edges. These determine the residue $\mathrm{res}(\Gamma)$.

An automorphism of a Feynman diagram consists of a bijection $\Gamma \to \Gamma$ that fixes the external edges and $\mathcal{E}, \mathcal{V}$. The symmetry factor of Γ is then given by:

$$\mathrm{sym}(\Gamma) = |\mathrm{Aut}(\Gamma)| \,. \tag{70.58}$$

Aside from the obvious algebra structure on Feynman diagrams, given by formal linear combinations and juxtaposition, there also exists a pre-Lie algebra structure (Definition 32.2.1):

$$(\Gamma_1, \Gamma_2) = \sum_{\Gamma} n(\Gamma_1, \Gamma_2; \Gamma)\Gamma \,, \tag{70.59}$$

where $n(\Gamma_1, \Gamma_2; \Gamma) \in \mathbb{N}$ counts the number of ways that, if Γ_2 is a subgraph of Γ, $\Gamma/\Gamma_2 \cong \Gamma_1$ and Γ ranges over 1PI diagrams.

70.5.3 Degree of divergence

Definition 70.5.7 (Mass dimension). Through the Compton wavelength

$$\lambda = \frac{h}{mc} \,, \tag{70.60}$$

the dimensions of length and mass are related. In particular, in natural units $\hbar = c = 1$, one obtains $[m] = [\lambda]^{-1}$. This dimension is called the mass dimension. @@ ADD Compton and de Broglie wavelength to [QM] @@

Now, consider a general Lagrangian density $\mathcal{L}$. Since this density is integrated against a volume form to obtain the dimensionless action, the mass dimension of $\mathcal{L}$ has to equal the spacetime dimension:

$$[\mathcal{L}] = d\,. \tag{70.61}$$

To obtain the mass dimension of a field ϕ, the general procedure, is to look at the mass term in the Lagrangian. For example, for scalar theories:

$$\begin{aligned}[\mathcal{L}_{\text{mass}}] &= \left[\frac{1}{2}m^2\phi^2\right] \\ &= 2 + 2[\phi]\end{aligned}$$

and, hence

$$[\phi] = \frac{d-2}{2}\,. \tag{70.62}$$

Definition 70.5.8 (Degree of divergence). Consider a general Feynman diagram Γ and its associated integral I_Γ. The (superficial) degree of divergence of Γ is defined as

$$D_\Gamma := \text{power of momenta in numerator} - \text{power of momenta in denominator}\,. \tag{70.63}$$

This can also be calculated at the level of the graph Γ:

$$D_\Gamma = dL - 2I_b - I_f - \sum_{i\in I} n_i d_i\,, \tag{70.64}$$

where L is the number of loops in Γ, n_i is the number of vertices of type $i \in I$, d_i is the number of derivatives in the interaction term corresponding to the vertex type $i \in I$ and I_b, I_f are the number of internal boson and fermion lines, respectively.

A relation between the mass dimension $[\Gamma]$ and the degree of divergence D_Γ is given by the following relation:

$$D_\Gamma = [\Gamma] - \sum_{n=3}^{+\infty} k_n[g_n]\,, \tag{70.65}$$

where k_n denotes the number of vertices of valence $n \in \mathbb{N}$.

@@ CHECK THIS (e.g. Peskin) @@

Property 70.5.9 (Renormalizability). A Feynman diagram Γ is UV divergent if $D_\Gamma \geq 0$, with $D_\Gamma = 0$ corresponding to logarithmic divergences, and, hence, $D_\Gamma < 0$ implies a UV-convergent diagram. If the maximal degree of divergence that occurs for a given theory is finite, the theory is said to be **renormalizable**. By Eq. (70.65), a necessary condition for renormalizability is that the mass dimension of all coupling constants is positive (or zero).

70.5.4 Amplituhedron ♣

In the 80s, *Parke* and *Taylor* discovered that, in QCD (Section 70.4), when scattering gluons, certain tree-level amplitudes consistently vanished. For $n \in \mathbb{N}$ gluons, if either all n or $n-1$ of them have the same helicity, the scattering amplitudes vanish. The **maximally helicity-violating** (MHV) amplitudes arise for $n-2$ gluons of the same helicity and can, in fact, be expressed in a single term, the **Parke–Taylor formula**:

$$\mathcal{A}(p_1^+ \cdots p_i^- \cdots p_j^- \cdots p_n^+) = i(-g)^{n-2} \frac{\langle p_i\, p_j\rangle^4}{\langle p_1\, p_2\rangle\langle p_2\, p_3\rangle \cdots \langle p_{n-1}\, p_n\rangle\langle p_n\, p_1\rangle}\,, \tag{70.66}$$

where the convention of two gluons with negative helicity was chosen. Note that the bilinear operations in this formula are not inner products, but spinor bilinears (Eq. (68.33)) in the spinor-helicity formalism.

70.6 Renormalization

One of the biggest issues in (quantum) field theory are the divergences that arise everywhere in calculations involving loop diagrams. Renormalization theory tries to find a way around these (nonphysical) divergences.

70.6.1 Introduction: Statistical physics

Before introducing renormalization theory in the context of quatum field theory, it is helpful to study some applications in statistical physics, in particular in the study of lattice systems. To this end, this section will be focused on the study of the Ising model 63.6.1 on a lattice Λ:

$$\widehat{H} := -\sum_{\langle i,j\rangle\in\Lambda} J_{ij}\hat{S}_i\hat{S}_j - h\sum_{i\in\Lambda}\hat{S}_i\,. \tag{70.67}$$

70.6.2 Dimensional regularization

When calculating transition amplitudes and scattering cross-sections, three situations can arise:

1. Convergence: Only low-energy/low-momentum contributions matter.
2. UV divergence: These are sensivity to high-energy/high-momentum contributions, but only to these. Given a momentum cut-off, the integrals are fine.
3. Logarithmic divergence: These get contributions from both low- and high-energy scales and, hence, a cut-off does not solve the issue.

Consider the one-dimensional integral[7]

$$f_1(\theta) := \int_0^{+\infty} \frac{1}{x+\theta}\, dx\,. \tag{70.68}$$

For nonzero θ, this integral has no IR divergence, but the upper bound still leads to trouble since $\ln(x) \xrightarrow{x\to\infty} +\infty$. To this end, one can (formally) analytically continue the integral into a space '$\mathbb{R}^{1-\varepsilon}$'. This leads to the following regularization scheme:

$$f_{1,\varepsilon}(\theta) := \int_0^{+\infty} \frac{x^{-\varepsilon}}{x+\theta}\, dx\,. \tag{70.69}$$

Looking at Formula 16.6.8, this integral can be seen to be

$$f_{1,\varepsilon}(\theta) = \frac{B(\varepsilon, 1-\varepsilon)}{\theta^{\varepsilon}}\,. \tag{70.70}$$

Although this formula is finite for all $\theta \in \mathbb{R}$, it is now a function of ε and actually diverges for $\varepsilon \longrightarrow 0$ (which recovers the initial divergence).

This is where renormalization comes in, where counterms are subtracted to remove the dependence on ε and return a physical value. Here, the *on-shell renormalization scheme* is adopted:

$$f_1^R(\theta) := \lim_{\varepsilon\to 0}\big(f_{1,\varepsilon}(\theta) - f_{1,\varepsilon}(1)\big) = \lim_{\varepsilon\to 0} B(\varepsilon, 1-\varepsilon)(\theta^{-\varepsilon} - 1)\,. \tag{70.71}$$

Property 16.6.14, together with a Taylor expansion around $x = 0$, show that this is finite for $\varepsilon \longrightarrow 0$. If this were all, one could simply add counterterms for all such integrals. However, one also encounters interated integrals, corresponding to higher-order diagrams, and these should be regularized and renormalized as well. For a double integral, dimensional regularization gives:

$$f_{2,\varepsilon}(\theta) := \int_0^{+\infty}\int_0^{+\infty} \frac{x_1^{-\varepsilon}}{x_1+\theta}\frac{x_2^{-\varepsilon}}{x_2+x_1}\, dx_2\, dx_1 = \int_0^{+\infty} \frac{x_1^{-\varepsilon}}{x_1+\theta} f_{1,\varepsilon}(x_1)\, dx_1\,. \tag{70.72}$$

Naively subtracting a counterterm as for f_1, will, however, not work in this case since there are two (logarithmic) divergences. One for fixed x_1, when $x_2 \longrightarrow +\infty$ and one when both $x_1, x_2 \longrightarrow +\infty$. The former is called a **subdivergence**.

Remark 70.6.1 (Nonlocality). The naive subtraction scheme $f_{2,\varepsilon}^R(\theta) := f_{2,\varepsilon}(\theta) - f_{2,\varepsilon}(1)$ can be found to be proportional to $\ln(\theta)$. In practice, the scale parameter θ is often related to the external momentum q^2. A term $\ln(q^2)$, however, points towards nonlocal behaviour since this term involves arbitrary high powers in q^2 or, after Fourier transforming back to position space, arbitrary high powers of a differential operator.

[7] Any integrand $g(x) \sim O(\frac{1}{x})$ will do.

First, one has to remove subdivergences. For f_2, this goes as follows:

$$\overline{f}_{2,\varepsilon}(\theta) := f_{2,\varepsilon}(\theta) - f_{1,\varepsilon}(\theta) f_{1,\varepsilon}(1)\,. \tag{70.73}$$

A term consisting of the original function with the subdivergent factor set to 1, times that same contribution evaluated at $\theta = 1$ is subtracted. In the next section, it will be shown how this procedure is generalized to arbitrary (logarithmically) divergent integrals. In the next step, one applies the 'naive' subtraction scheme as for f_1:

$$f_2^R(\theta) := \lim_{\varepsilon\to 0}\big(\overline{f}_{2,\varepsilon}(\theta) - \overline{f}_{2,\varepsilon}(1)\big)\,. \tag{70.74}$$

To keep track of subdivergences, a diagrammatic method can be adopted. General integrals can be represented by rooted trees. Some examples are shown in Fig. 70.1. These trees correspond to the following divergent integrals:

- f_{t_1}:

$$\int_0^{+\infty} \frac{x^{-\varepsilon}}{x+\theta}\,dx\,. \tag{70.75}$$

- f_{t_2}:

$$\int_0^{+\infty} \frac{x_1^{-\varepsilon} f_{t_1}(x_1)}{x_1+\theta}\,dx_1\,. \tag{70.76}$$

- $f_{t_{3,1}}$:

$$\int_0^{+\infty} \frac{x_1^{-\varepsilon} f_{t_2}(x_1)}{x_1+\theta}\,dx_1\,. \tag{70.77}$$

- $f_{t_{3,2}}$:

$$\int_0^{+\infty} \frac{x_1^{-\varepsilon} f_{t_1}(x_1) f_{t_1}(x_1)}{x_1+\theta}\,dx_1\,. \tag{70.78}$$

70.6.3 Wilsonian renormalization

@@ COMPLETE @@

70.6.4 Connes–Kreimer renormalization

The trees by which divergences and subdivergences in Feynman diagrams can be organized, admit an interesting algebraic structure. Consider the free commutative algebra

Figure 70.1: First four rooted trees.

$\mathcal{H}_R$ on rooted trees, where multiplication corresponds to juxtaposition (creating a 'forest') with the empty tree as unit.

$\mathcal{H}_R$ admits more structure then simply the addition and multiplication. It will be equipped with a Hopf algebra structure (Definition 28.2.3). The operations are defined as follows:

- Counit:

$$\varepsilon(\mathbf{1}) := 1 \qquad\qquad \varepsilon(T) := 0\,. \tag{70.79}$$

- Coproduct:

$$\Delta(T) := T \otimes \mathbf{1} + \mathbf{1} \otimes T + \sum_c P_c(T) \otimes R_c(T)\,, \tag{70.80}$$

  where the last term is a sum over all possible combinations of cuts for which any path from a vertex to the root only passes through at most one cut, i.e. the **admissible cuts**.[8] Given a cut c, the **trunk** $R_c(T)$ is the part of the tree that was attached to the root and the **branches** $P_c(T)$ is the forest given by all remaining components (this forest will consist of more than one tree if and only if c contained multiples cuts).

- Antipode (defined recursively):

$$S(\mathbf{1}) := \mathbf{1} \qquad\qquad S(T) := -T - \sum_c S\big(P_c(T)\big)R_c(T)\,. \tag{70.81}$$

Remark 70.6.2 (Sweedler notation). If the set of admissible cuts would be enlarged with the trivial options, the coproduct could be rewritten in Sweedler notation (Nota-

[8] Although, in theory, the two trivial cuts, where either $R_c(T) = \emptyset$ or $R_c(T) = T$, are also admissible, these are not included.

tion 28.1.9) as

$$\Delta(T) = \sum_{(T)} T_1 \otimes T_2 \,. \tag{70.82}$$

For clarity's sake, this will not be done in this section unless explicitly noted.

Now, consider the operator $B_- : \mathcal{H}_R \to \mathcal{H}_R$ that removes the root from a tree (resulting in a forest) and the operator $B_+ : \mathcal{H}_R \to \mathcal{H}_R$ that takes a forest of trees and grafts all of them to a new root. These can easily seen to be inverses. With these operators, the coproduct can be rewritten as follows:

$$\Delta(T) = T \otimes \mathbf{1} + (\mathbb{1} \otimes B_+) \circ \Delta \circ B_+(T) \,. \tag{70.83}$$

So, now, how to proceed? A Hopf algebra structure on Feynman graphs has been provided, but this does not say how the resulting Feynman amplitudes should be handled. These amplitudes, usually expressed through *Feynman rules*, are given by **characters** of $\mathcal{H}_R$, i.e. elements of $\mathbb{C}\mathbf{Alg}(\mathcal{H}_R, \mathbb{C})$. These form a group under the convolution product (Property 28.2.8):

$$\phi_1 * \phi_2 = \mu_{\mathbb{C}} \circ (\phi_1 \otimes \phi_2) \circ \Delta \,. \tag{70.84}$$

Now, when regularizing these amplitudes, the characters do not take values anymore in $\mathbb{C}$ but in some $\mathbb{C}$-algebra A, such as the Laurent series $\mathbb{C}[[\varepsilon, \varepsilon^{-1}]]$ in the case of dimensional regularization, and, hence, one obtains the group of regularized characters $\mathbb{C}\mathbf{Alg}\big(\mathcal{H}_R, A\big)$.

The recursive expression for the antipode is the equation that ties in the Connes–Kreimer algebra with the substraction of subdivergent integrals from the previous section. The renormalized integrals are obtained through a recursive formula, akin to that of the antipode.

$$\phi^R(T) := S^R_\Phi * \phi(T) \,, \tag{70.85}$$

where S^R_ϕ is a deformation of $\phi \circ S$ obtained recursively as follows:

$$S^R_\phi(T) := -R\big[m \circ (S^R_\phi \otimes \pi) \circ \Delta\big] \,, \tag{70.86}$$

where $R : A \to A$ is a linear operator, m is the multiplication on A and $\pi := \mathbb{1}_{\mathcal{H}_R} - \eta \circ \varepsilon$ is the augmentation projection (the projection that kills nontrivial forests), or, more explicitly:

$$S^R_\phi(T) = -R\big[\phi(T)\big] - R\left[\sum_c S^R_\phi\big(P_c(T)\big)\phi\big(R_c(T)\big)\right] . \tag{70.87}$$

The assignment $\mathcal{R} : \phi \mapsto m \circ (S_\phi^R \otimes \pi) \circ \Delta$ is **Bogoliubov's R-operation**. Using the bialgebra laws, this can be rewritten as follows:

$$\phi^R(T) = \mathcal{R}[\phi](\Gamma) + S_\phi^R(\Gamma)\,. \tag{70.88}$$

Now, consider a monomial in the Lagrangian or, equivalently, an external leg structure $\underline{r}$. The total (combinatorial) Green's function corresponding to $\underline{r}$ is given by the formal sum of Feynman diagrams

$$\Gamma^{\underline{r}} := 1 + \sum_{\mathrm{res}(\Gamma)=\underline{r}} \alpha^{|\Gamma|} \frac{\Gamma}{\mathsf{sym}(\Gamma)}\,. \tag{70.89}$$

The renormalized term in the Lagrangian is then given by

$$Z^{\underline{r}} = S_\phi^R(\Gamma^{\underline{r}})\,. \tag{70.90}$$

The R-operation corresponds to the choice of renormalization scheme and isolates the singular part of the Feynman diagram, e.g.:

- On-shell renormalization:

$$R[f(\theta)] := f(1)\,. \tag{70.91}$$

- BPHZ[9]/MS[10] renormalization:@@ IS THIS CORRECT (or is this for DimReg?) @@

$$R[f(\theta)] := \mathsf{PolePart}_\varepsilon[f(1)]\,. \tag{70.92}$$

Property 70.6.3 (Butcher group). The Lie group $\mathbb{C}\mathbf{Alg}(\mathcal{H}_R, \mathbb{C})$ of characters has a Lie algebra the space of Feynman diagrams with the bracket induced by the pre-Lie algebra operation (70.59). This group is known as the (complex) Butcher group.[11]

Theorem 70.6.4 (Birkhoff decomposition). *Let $\mathcal{H}_R$ be the Hopf algebra of Feynman diagrams for a renormalizable pQFT and consider a commutative, unital Rota–Baxter algebra (A, R) of weight 1 (Definition 22.7.12) such that R is idempotent. Denote the space of A-regularized characters by $\mathcal{G}_A := \mathsf{Char}(\mathcal{H}_R, A)$ with unit $e := \eta_A \circ \varepsilon_{\mathcal{H}_R}$. The following statements hold:*

- $\big(\mathbf{Vect}_{\mathbb{C}}(\mathcal{H}_R, A), \mathcal{R}\big)$ *is a (complete filtered) Rota–Baxter algebra of weight 1, where $\mathcal{R}(\phi) := R \circ \phi$.*

- *For every $\phi \in e + \mathcal{G}_A$, there exists a unique decomposition*

$$\phi = \phi_-^{-1} * \phi_+\,, \tag{70.93}$$

where $\phi_- \in e + \mathcal{R}(\mathcal{A}_1)$ and $\phi_+ \in e + \widetilde{\mathcal{R}}(\mathcal{A}_1)$, with $\widetilde{\mathcal{R}} := \mathbb{1} - \mathcal{R}$ as in Property 22.7.14.

[9] *Bogoliubov, Parasiuk, Hepp* and *Zimmermann*

[10] minimal subtraction

[11] This group first arose in the combinatorial study of *Runga–Kutta method* in numerical analysis.

- *The Birkoff decomposition of ϕ is obtained recursively as*

$$\phi_- = e - \mathcal{R}\big(\phi_- * (\phi - e)\big) \qquad \phi_+ = e - \widetilde{\mathcal{R}}\big(\phi_+ * (\phi^{-1} - e)\big)\,. \tag{70.94}$$

These can be rewritten as follows when acting on a nontrivial Feynman diagram Γ:

$$\begin{aligned} \phi_-(\Gamma) &= -R\left(\phi(\Gamma) + \sum_{(\Gamma)} \phi_-(\Gamma_1)\phi(\Gamma_2)\right), \\ \phi_+(\Gamma) &= \widetilde{R}\left(\phi(\Gamma) + \sum_{(\Gamma)} \phi_-(\Gamma_1)\phi(\Gamma_2)\right), \end{aligned} \tag{70.95}$$

where Sweedler's notation (Remark 70.6.2) was used.

- *Spitzer's identity for noncommutative Rota–Baxter algebra (cf. Formula 22.7.15) implies that ϕ_-, ϕ_+ are algebra morphisms and that*

$$\phi_- = \exp^*\big(-\mathcal{R}\big(\chi(Z)\big)\big) \qquad \phi_+ = \exp^*\big(\widetilde{\mathcal{R}}\big(\chi(Z)\big)\big)\,, \tag{70.96}$$

where $Z \in \mathfrak{g}_A$ is the generator of ϕ and $\chi : \mathcal{A}_1 \to \mathcal{A}_1$ is recursively defined as follows:

$$\chi(Z) := Z - \mathsf{BCH}\big(\mathcal{R}(\chi(Z)), \widetilde{\mathcal{R}}(\chi(Z))\big)\,, \tag{70.97}$$

with BCH *the Baker–Campbell–Hausdorff operator (cf. Formula 32.2.29)*

$$\exp(x)\exp(y) = \exp\big(x + y + \mathsf{BCH}(x, y)\big)\,. \tag{70.98}$$

This theorem implies that the form of the regularized character $\phi_+ \equiv \phi^R$ and the counterterm $\phi_- \equiv S_\phi^R$ are completely fixed by the decomposable nature of (noncommutative) idempotent Rota–Baxter algebras of weight 1. The freedom in the choice of renormalization scheme only corresponds to the freedom in the exact choice of Rota–Baxter algebra A.

@@ COMPLETE @@

70.7 Entanglement in QFT

This section should be seen as a generalization of Chapter 58 to the continuum setting and, in particular, of the characterization and computation of entanglement.

70.7.1 Lattice theories

In this section, the most important definitions and constructions in ordinary quantum information theory are recalled and applied to a lattice theory. Taking the lattice spacing to zero will (formally) allow to extend the definitions to continuum field theories

(up to some technicalities that will be explained when necessary). For simplicity, it will be assumed that the local Hilbert space is finite-dimensional.

Consider a bipartite subdivision $A \cup A^c$ of the lattice, given by a codimension-1 hypersurface ∂A, called the **entangling surface**. This induces a binary factorization of the total Hilbert space (all degrees of freedom are assumed to be confined to individual vertices) and, hence, one can compute the reduced density matrix for both A and its complement A^c. The eigenvalues, which solely depend on the entangling surface ∂A, allow to calculate the von Neumann entropy:[12]

$$S(\rho_A) := -\operatorname{tr}(\rho_A \ln \rho_A) = -\sum_i \rho_i \ln \rho_i \,. \tag{70.99}$$

In the same way, one can also introduce the Rényi q-entropy:

$$S_q(\rho_A) := \frac{1}{1-q} \ln\Big(\sum_i \rho_i^q\Big) \,. \tag{70.100}$$

Property 70.7.1 (Limiting case). First of all, one can analytically continue the definition of the q-entropy to arbitrary positive real numbers. The limit $q \longrightarrow 1$ coincides with the von Neumann entropy.

@@ COMPLETE (Ryu–Takayanagi, MERAs, ...) @@

[12]Certain assumptions ought to be made as to keep the entropy finite whenever the state-space is infinite-dimensional, since it can be shown that the set of states with infinite von Neumann entropy is trace norm-dense (see Eisert, Simon, and Plenio (2002)).

Chapter 71

Particle Physics

This chapter summarizes the content of the foregoing chapters and tries to relate it to a more phenomenological approach to modern physics.

Chapter Contents

71.1 Overview **1198**

71.2 Standard model **1200**

71.2.1 Group theory 1200

71.2.2 Symmetry breaking 1201

71.3 Shortcomings **1203**

71.3.1 Open problems 1203

71.3.2 Might-be problems 1203

71.1 Overview

In this section we give a short overview of the particles in the Standard Model of particle physics. This will not include any possible extensions that are currently being researched.

Definition 71.1.1 (Leptons). The leptons are fermionic particles, i.e. characterized by the Dirac equation 68.2.1. They are spin-$\frac{1}{2}$ particles that are charged under both the electromagnetic and weak interactions:

$$\begin{pmatrix} e^- \\ \nu_e \end{pmatrix} \qquad \begin{pmatrix} \mu^- \\ \nu_\mu \end{pmatrix} \qquad \begin{pmatrix} \tau^- \\ \nu_\tau \end{pmatrix}$$

and their antiparticles

$$\begin{pmatrix} e^+ \\ \overline{\nu}_e \end{pmatrix} \qquad \begin{pmatrix} \mu^+ \\ \overline{\nu}_\mu \end{pmatrix} \qquad \begin{pmatrix} \tau^+ \\ \overline{\nu}_\tau \end{pmatrix}.$$

The particles are grouped in weak doublets. The upper particles in a doublet are always electromagnetically charged, while the lower particles are neutral. From left to right one has the **electron**, **mu(on)** and **tau** particles (and their corresponding **neutrinos**).

Definition 71.1.2 (Quarks). A second class of fundamental particles is given by those that not only act through the electromagnetic and weak interactions, but also through the strong interaction. These quarks are also fermionic spin-$\frac{1}{2}$ particles:

$$\begin{pmatrix} u \\ d \end{pmatrix} \qquad \begin{pmatrix} s \\ c \end{pmatrix} \qquad \begin{pmatrix} t \\ b \end{pmatrix}.$$

The quarks are again paired in weak doublets. The electromagnetic charge of the upper particles is $\frac{2}{3}$, while that of the lower particles is $-\frac{1}{3}$. Every quark also has strong (**colour**) charge. The antipartners of these doublets are not shown, but also exist. From left to right and top to bottom one has the **up**, **strange**, **top**, **down**, **charm** and **bottom** quarks.

Definition 71.1.3 (Gauge bosons). The last class of particles is that of the force carriers. These are bosonic in nature and are thus not characterized by the Dirac equation.

- The electromagnetic interaction mediated by an uncharged spin-1 boson, the **foton** γ.
- The weak interaction is mediated by three particles:

  $$Z^0 \qquad \begin{pmatrix} W^+ \\ W^- \end{pmatrix}.$$

  The Z^0-boson is again uncharged, while the the W-doublet carries both electromagnetic and weak charges (± 1). These particles are sometimes called **intermediate vector bosons**.
- The strong interaction is mediated by bosons that are only strongly charged, the **gluons** g.

Remark 71.1.4. Note that gravity is not mentioned in the above classification. Although some of the particles above have a mass, gravity is not part of the Standard Model. Mass is simply a parameter in the model.

Definition 71.1.5 (Higgs boson). A last constituent of the Standard Model is the Higgs boson. This scalar field is responsible for giving a small contribution to the lepton masses. It is a spin-0 particle that has no charges at all.

Remark 71.1.6 (Origin of mass: Yukawa interaction). Contrary to what many popular sources state, the Higgs boson is not responsible for 'the' mass of matter. The mass of fundamental particles indeed comes from the **Yukawa interaction** with the Higgs field, but most of the mass of compound particles actually comes from the 'nuclear force' that holds them together (gluon and virtual meson exchange). So it is CQD, rather than the Higgs mechanism that is important for the mass of ordinary matter.

It should be noted that the Yukawa couplings

$$\mathcal{L}_{\text{Yukawa}}(\phi, \psi) \sim \overline{\psi}\phi\psi \tag{71.1}$$

are introduced in the Standard Model by hand. As of yet, there is no accepted mechanism that generates these terms in the Lagrangian.

The Higgs particle is, however, important for another aspect of the Standard Model. In the next section it will be explained how this particle is responsible for electroweak symmetry breaking and the mass of the intermediate vector bosons.

71.2 Standard model

71.2.1 Group theory

In this section, the structure of the different particles and multiplets in the Standard Model is related to group theory. Recall that the fundamental interactions are described by the following (Lie) groups:

- Electromagnetic interaction: $\mathsf{U}(1)$,
- Weak interaction: $\mathsf{SU}(2)$, and
- Strong interaction: $\mathsf{SU}(3)$.

The fact that all particles can be organised in terms of representations of these groups has some important consequences.

Property 71.2.1 (Electromagnetism). All (electromagnetically interacting) fundamental particles transform under the fundamental representation of $\mathsf{U}(1)$. Since this representation is one-dimensional, interactions with a photon do not change particle types.

Property 71.2.2 (Weak interactions). All (weakly interacting) fundamental particles can be organised into fundamental representations of $\mathsf{SU}(2)$. Since this representation is two-dimensional, all particles can be organised into doublets. These are the pairs from the previous section. Interactions with the $W^{\pm}$-bosons interchange the particles in a doublet.

At this point we have to make an important remark. As was discovered in the 50s and 60s, the weak interaction behaves in a special way with respect to parity transformations. Left-handed and right-handed (Weyl) spinors are treated in a different way. Only the left-handed particles couple to $W^{\pm}$-bosons. So the doublets shown in the previous section are actually only the left-handed particles. The right-handed particles transform in the trivial, one-dimensional representation of SU(2), e.g.:

$$\begin{pmatrix} e^- \\ \nu_e \end{pmatrix}_L \qquad e_R^- \qquad \nu_{e,R}\,.$$

This is characterized by a so-called V-A theory (*vector minus axial*-theory), where the coupling to the weak interaction is governed by the chirality operator $1-\gamma$ (right-handed spinors are eigenvectors of γ with eigenvalue 1).

In the next section it is explained why the intermediate vector bosons do not form an adjoint triplet (W^-, W^0, W^+) as would be expected from ordinary gauge theory, but rather form a doublet $W^{\pm}$ and a singlet Z^0. To understand this one needs to unify the forces into an electroweak theory and then break symmetry.

Property 71.2.3 (Strong interaction). All (strongly interaction) fundamental particles can be organised into fundamental representations of SU(3). Since this representation is three-dimensional all particle types come in three version: red, green and blue. Contrary to the weak interaction, the strong interaction does not interchange particle types, it only changes the 'colour' of a particle.

The gauge bosons mediating the strong interaction form an adjoint representation of SU(3). The dimension formula for the adjoint representation gives $N^2-1=8$ gluons. Note that free colour charges do not occur outside of interactions due to confinement in QCD, so in practice all particles that can be observed will be white. This also explains why, with three colours, one does not obtain 9 colour-anticolour gluon states. If the singlet state $r\bar{r}+g\bar{g}+b\bar{b}$ would exist, it could couple colourless states to colourful states, which is not allowed.

71.2.2 Symmetry breaking

At our energy scale, the gauge bosons represent three distinct fundamental forces. However, at higher energies, these forces are in fact not distinct. They come from a unified force, but due to spontaneous symmetry breaking at lower energies, one observes this distinction.

The 'true' gauge group of the Standard Model is

$$\mathrm{U}(1)\times\mathrm{SU}(2)\times\mathrm{SU}(3)/\mathbb{Z}_6\,. \tag{71.2}$$

The center $\mathbb{Z}_6$ acts trivially on all particle types. It is generated by the element

$$(e^{\pi i/3}, -1, e^{2\pi i/3}) . \tag{71.3}$$

As noted in the previous section, the second factor only acts on left-handed particles, so a better notation would be $\mathsf{SU}(2)_L$. However, the first factor is also not what it seems. The $\mathsf{U}(1)$ gauge group is not simply that of the electromagnetic interaction. If $\mathsf{U}(1)$ was the electromagnetic gauge group, by the direct product structure the electromagnetic and weak interactions would commute and, accordingly, the particles in a weak multiplet would necessarily have the same charge. This is clearly not the case. Correctly interpreting these groups will also explain why the weak gauge bosons ($W^\pm$ and Z^0) seemingly do not form an adjoint triplet.

In the 50s a new formula was proposed to be able to organize the dozens of particles that were experimentally detected. This was the so-called **Gell-Mann–Nishijima formula**:

$$Q = I_3 + \tfrac{1}{2}Y , \tag{71.4}$$

where the new **hypercharge** quantum number Y was introduced. Originally introduced for the strong interaction, it later became apparent that this formula is actually better suited for describing the weak interaction, Q being the electric charge and I_3 being the weak isospin. The weak hypercharge Y_W is now exactly the quantum number related to the $\mathsf{U}(1)$ gauge group. After spontaneous symmetry breaking, the associated B-particle mixes up with the W^0-particle from the weak adjoint triplet to give the photon γ and the Z^0-boson:

$$\gamma = W^0 + \tfrac{1}{2}B , \tag{71.5}$$

$$Z^0 = W^0 - \tfrac{1}{2}B . \tag{71.6}$$

This recovers the Gell-Mann–Nishijima formula.

The reason for electroweak symmetry breaking from $\mathsf{SU}(2)_L \times \mathsf{U}(1)_Y$ to $\mathsf{U}(1)_{\text{em}}$ is given by the Higgs mechanism (see Section 69.3.1). Before symmetry breaking, the Higgs field is a $\mathsf{SU}(2)_L$-doublet. This couples to the electroweak interaction (the doublet has $I = \frac{1}{2}$ and $Y_W = 1$). However, if the Higgs field has a nonzero vacuum expectation value, the total symmetry group is broken. This has two consequences. First of all, the gauge fields are reduced and the new generator γ is obtained (the photon). Secondly, the remaining components of the gauge fields give rise to three massive Goldstone bosons: $W^\pm$ and Z^0. It follows that the $\mathsf{SU}(2)_L$-doublets and the intermediate vector bosons in the Standard model do not constitute a gauge theory. Only the electromagnetic sector or the complete unified electroweak theory are proper gauge theories.

@@ COMPLETE @@

71.3 Shortcomings

Although the Standard Model if often celebrated as the most accurate theory every written down, it might not be flawless. In fact, we know it is not perfect. There are a couple of well-known phenomena that cannot be explained by the model and there are also some hints toward problems within the model itself.

71.3.1 Open problems

There are two very big open problems in the field of particle physics at this time. Both are related to cosmology and are mainly of importance on larger scales.

The first, and the most obvious one, is the unification problem for the Standard Model with gravity. No quantummechanical model exists for gravity, but it is known that (almost) all particles interact gravitationally. So the ultimate physical theory should describe both aspects of nature. The current way around this problem is the fact that gravity is extremely weak, the coupling constant (Newton's constant G) is much smaller than even the weak coupling constant, and can be ignored for particle physics. However, when trying to describe for example baryogenesis shortly after the Big Bang, an era where the universe was completely different in terms of energy scales, it is expected that all interactions have to be taken into account. (See Chapter 74 for possible extensions of the Standard Model.)

Another problem, related to the gravitational one, is the existence of **dark matter** and **dark energy**. Astronomical studies have indicated that the visible content of the universe, hadronic matter and light, cannot explain the gravitational phenomena that are observed. The mass required for this phenomena just cannot be accounted for only by matter and energy as we know them. In fact, it is estimated that only about 5% of the known universe exists out of hadronic matter and light. The remaining 95% contains about 20% dark matter and 75% dark energy.

71.3.2 Might-be problems

Since the advent of the Standard Model scientists have tried to test its limits. One of the biggest surprises was exactly how robust and accurate this theory was. Even the most precise measurements seemed to agree with the theoretical predictions. However, as was to be expected, there are hints that even this theory might not be the end (even for particle physics).

The first issue is related to neutrinos. As currently understood, these particles are not charged under the electromagnetic and strong interactions. However, because of weak symmetry breaking, the right-handed neutrinos also do not interact weakly. This

means that the only possible way to observe them would be through gravitational interactions. Because of *neutrino oscillations* it is known that neutrinos have an intrinsic mass, but this does not really increase the hope of detecting right-handed particles. One needs an extremely large amount of matter in the detectors in order to have a reasonable chance of detecting gravitational interactions with neutrinos.

@@ STERILE NEUTRINOS @@

A second possible issue arose a coupe of years ago after FermiLab released measurements of the magnetic moment of muons, the muon $g-2$ experiment. From basic quantum theory it is known that particle spin induces a magnetic dipole moment:

$$\vec{\mathbf{B}} = g\frac{e}{2m}\vec{\mathbf{S}}\,. \tag{71.7}$$

The factor g is called the **anomalous magnetic moment**. Without quantum field theory this factor would be equal to 2, but due to higher-order loop corrections, the value is slightly larger than 2. The latest measurements at FermiLab, however, indicated that the true value differs from the predicted one.

Even more recently, a third possible deviation from the Standard Model was found. Measurements of the W-boson mass seemed to indicate that it did not exactly match the predictions of the Standard Model, the mass was found to be slightly larger. If found to be true, this is especially problematic since the value of the weak mixing angle is fixed and the mass of the Z-boson is measured very accurately.

Chapter 72

Axiomatic QFT ♣

For the sections on TQFTs and open-closed TQFTs, the reader is referred to the original papers Atiyah (1988) and Lauda and Pfeiffer (2008), respectively. For the sections on the Haag–Kastler framework and perturbative AQFT, the reader is referred to the work of Brunetti, Fredenhagen, and Verch (2003) and Fewster and Rejzner (2020); Rejzner (2016). A reference for the remaining sections on algebraic QFT is Baez et al. (2014). The sections on cohomological quantization and quantization from a sheaf point of view are based on Nuiten (2013).

Chapter Contents

72.1 Topological QFT **1206**

72.1.1 Atiyah–Segal axioms 1206

72.1.2 Open-closed TQFT 1207

72.2 Algebraic QFT **1207**

72.2.1 Haag–Kastler axioms 1207

72.3 Perturbative AQFT **1209**

72.3.1 Haag–Kastler axioms 1210

72.3.2 Functional analysis 1210

72.3.3 Equations of motion 1212

72.3.4 Symmetries 1215

72.3.5 Quantization 1216

72.4 Prequantum field theory **1218**

72.1 Topological QFT

For convenience, the reader is reminded of some of the notations that are used. **FinVect** denotes the category of finite-dimensional vector spaces over $\mathbb{C}$ and $\mathbf{Bord}^d_{d-1}$ denotes the category of d-dimensional (framed[1]) cobordisms (Definition 31.4.5).

72.1.1 Atiyah–Segal axioms

Axiom 72.1 (Atiyah–Segal). A d-dimensional **topological quantum field theory** (TQFT) is a symmetric monoidal functor $F : \mathbf{Bord}^d_{d-1} \to \mathbf{FinVect}$. This means (among other things) that F is a map satisfying the following axioms:

1. **Normalization**: $F(\emptyset) = \mathbb{C}$.
2. **Stacking**: $F(M \sqcup M') = F(M) \otimes F(M')$.
3. **Composition**: If $N = M \cup M'$, where ∂M and $\partial M'$ have opposite orientations, then
$$F(N) = F(M) \circ F(M').$$
4. **Invariance**: If $f : M \to M'$ is a diffeomorphism rel boundary, then $F \circ f = F$.

Example 72.1.1 (1D). In $d = 1$, the Atiyah–Segal axioms give rise to the following correspondence:

point with orientation $+$	vector space V
point with reversed orientation $-$	dual space V^*
line between points	linear map $f : V \to V$
cap between $\emptyset$ and points $+, -$	coevaluation $\mathbb{C} \to V \otimes V^*$
cup between points $-, +$ and $\emptyset$	evaluation $V^* \otimes V \to \mathbb{C}$

Essentially, this gives the structure of a (finite-dimensional) vector space and its dual.

Example 72.1.2 (2D). In $d = 2$, one can obtain a similar result by drawing all possible configurations. However, the existence and combination of 'pair of pants'-diagrams gives a richer structure. For 2D TQFTs, the corresponding object is a (finite-dimensional) commutative and cocommutative Frobenius algebra (Example 22.7.6).

In dimensions 3 and higher, the definition above is intractable. To allow the construction to be generalized to higher dimensions, one considers the following (extended) definition.

Definition 72.1.3 (Extended TQFT). A d-dimensional extended TQFT is a symmetric monoidal (∞, d)-functor $F : \mathbf{Bord}^d \to \mathbf{FinVect}$ satisfying the Atiyah–Segal axioms, where the invariance axiom is required only at the highest level of k-morphisms.

[1] A framing in this context means a framing of the d-stabilized tangent bundles $TM \oplus \mathbb{R}^{d-\dim(M)}$.

Remark 72.1.4. One can replace the category **FinVect** by any other symmetric monoidal category **C**.

The following conjecture by Baez and Dolan (1995), later (potentially) proven by Grady and Pavlov (2022); Lurie (2009), states that every extended TQFT is defined by a single object. This is motivated by the fact that all cobordisms between smooth manifolds can be obtained from gluing together (the geometric representation) of identity morphisms and taking disjoint unions of them.

Conjecture 72.1.5 (Cobordism hypothesis). $\mathbf{Bord}^d$ is the free symmetric monoidal (∞, n)-category with all duals on a single object pt. In particular, every symmetric monoidal (∞, n)-functor $F : \mathbf{Bord}^d \to \mathbf{C}$ is characterized by a single fully dualizable object $F(\mathrm{pt}) \in \mathrm{ob}(\mathbf{C})$.

72.1.2 Open-closed TQFT

In the case of ordinary TQFTs as defined in the previous section, one considers cobordisms between closed manifolds. Hence, the relevant objects in this category are manifolds with boundary. A generalization is obtained by relaxing the constraint on the cobordisms and allowing for the notion of manifolds with corners (Section 31.4). For simplicity, only the case of 2D TQFTs (as in the original definition) will be considered.

@@ COMPLETE @@

72.2 Algebraic QFT

72.2.1 Haag–Kastler axioms

Axiom 72.2 (Local net of observables). To every bounded open subset of Minkowski space, one associates a C^*-algebra (Definition 26.2.6). This assignment is required to satisfy the following conditions:

1. **Isotony**: If $U_1 \subset U_2$, then $\mathcal{O}(U_1) \hookrightarrow \mathcal{O}(U_2)$ and compositions are preserved.
2. **(Causal) locality**[2]: If U_1 and U_2 are spacelike separated, then

$$\left[\mathcal{O}(U_1), \mathcal{O}(U_2)\right] = 0 \tag{72.1}$$

(as a graded commutator) within a larger algebra $\mathcal{O}(U)$ such that $U_1, U_2 \subset U$.[3]

Remark 72.2.1. The isotony condition implies that local nets of observables are modelled by copresheaves $\mathbf{Mink} \to \mathbf{C^*Alg}$ that map (mono)morphisms to monomorphisms.

[2] Also called **microcausality**.

[3] Usually, U is chosen to be the smallest causally closed open set $U_1 \vee U_2$ containing both U_1 and U_2.

The notion of causal locality can be strengthened as follows.

Definition 72.2.2 (Strong locality). A local net of observables $\mathcal{O} : \mathbf{Mink} \to \mathbf{C^*Alg}$ is said to satisfy **strong causal locality** if it causally local and if, for all spacelike separated regions U_1, U_2 and all commutative subalgebras $A_1 \subseteq \mathcal{O}(U_1), A_2 \subseteq \mathcal{O}(U_2)$,

$$(A_1 \vee A_2) \cap \mathcal{O}(U_1) = A_1 \qquad (A_1 \vee A_2) \cap \mathcal{O}(U_2) = A_2\,, \tag{72.2}$$

where $A_1 \vee A_2$ denotes the subalgebra of $\mathcal{O}(U_1 \vee U_2)$ generated by A_1 and A_2.

Moreover, $\mathcal{O}$ is said to satisfy **Einstein locality** if $A_1 \vee A_2 \cong A_1 \otimes A_2$. Einstein locality implies stronger locality in the sense that algebras corresponding to spacelike separated regions are actually independent in the quantum mechanical sense.

Property 72.2.3. Under suitable conditions (Brunetti, Fredenhagen, Imani, & Rejzner, 2012), a functor $\mathcal{O} : \mathbf{Mink} \to \mathbf{C^*Alg}$ satisfies causal locality if and only if it can be extended to a monoidal functor $\mathcal{O} : \mathbf{Mink}^{\otimes} \to \mathbf{C^*Alg}^{\otimes}$ where **Mink** is endowed with the disjoint union of sets and $\mathbf{C^*Alg}$ with the spatial tensor product (Definition 26.2.32). Einstein locality is obtained for general monoidal

Axiom 72.3 (Poincaré covariance). For all bounded open sets O and Poincaré transformations Λ, there exists an isomorphism $\alpha^O_\Lambda : \mathcal{A}(O) \to \mathcal{A}(\Lambda O)$ such that the following conditions are satisfied:

1. If $O_1 \subset O_2$, then $\alpha_\Lambda \circ \iota_{O_1,O_2} = \iota_{\Lambda O_1,\Lambda O_2} \circ \alpha_\Lambda$.
2. The isomorphisms satisfy a composition rule: $\alpha^{\Lambda O}_{\Lambda'} \circ \alpha^O_\Lambda = \alpha^O_{\Lambda'\Lambda}$.

Axiom 72.4 (Spectrum). For all spacetime regions O, one can construct a faithful C^*-algebra representation ρ_O of $\mathcal{A}(O)$ on a fixed Hilbert space by the GNS construction (Construction 26.2.41). The different representations should be compatible, i.e. if $O_1 \subset O_2$, the restriction of ρ_{O_2} to $\mathcal{A}(O_1)$ should equal ρ_{O_1}. Furthermore, all spacetime translations are implemented unitarily:

$$U(a)\rho_O(c)U(a)^{-1} = \rho_{O+a}\big(\alpha^O_a(c)\big) \tag{72.3}$$

for all $c \in \mathcal{A}(O)$, where U is a unitary representation of the translation subgroup. In addition, the generators of the translation subgroup are required to have a spectrum that is contained in the future light cone.

The following axiom is not part of the standard Haag–Kastler framework but can be added to introduce dynamics.

Axiom 72.5 (Time slice). Consider two spacetime regions O_1, O_2. If O_1 contains a Cauchy surface of O_2, the morphism $\mathcal{A}(O_1 \hookrightarrow O_2)$ of C^*-algebras is an isomorphism.

Axiom 72.6 (Haag duality). Let $\overline{O}$ denote the spacelike complement of O and let $\mathcal{A}'$

denote the commutant of $\mathcal{A}$. Haag duality states that[4]

$$\mathcal{A}(\overline{O})' = \mathcal{A}(O) \tag{72.4}$$

for all causally closed sets O.

Remark 72.2.4. Haag duality is known to hold for all free theories and even for some interacting theories. However, it is also known to fail in the case of symmetry breaking (Roberts, 1974).

To generalize the above axiom system to globally hyperbolic space times, one must enter the realm of category theory. The notation[5] of Calaque and Strobl (2015) will be adopted. Let **Loc** be the category of globally hyperbolic space times with orientation- and causal structure-preserving isometries. Let **Obs** be the category of relevant algebras (commutative algebras for classical physics and C^*-algebras for quantum physics) together with suitable algebra morphisms. The assignment of algebras is then given by a functor $\mathcal{O} : \mathbf{Loc} \to \mathbf{Obs}$.

Remark 72.2.5 (LCAQFT). Given this generalized framework, also called **locally covariant AQFT**, the Haag–Kastler framework, which restricts attention to a given fixed manifold, is recovered when $\mathcal{O}$ is restricted to the subcategory of causally closed, relatively compact open subsets of a globally hyperbolic manifold (with inclusions as morphisms).

72.3 Perturbative AQFT

Perturbative algebraic quantum field theory (pAQFT) tries to formalize the notion of perturbation theory in QFT, as it is nowadays used in particle physics and high energy physics, by combining the formalism of AQFT with the causal perturbation theory as introduced by Epstein and Glaser (1973).

In this section, the main object of interest will be the space of fields. In almost all cases, this will be the space of sections $\mathcal{E}$ of a (graded) vector bundle $\pi : E \to M$. Such spaces are always (graded) Fréchet spaces by generalizing the construction in Definition 19.1.1. Here, one can use the fact that every manifold is σ-compact and admits a connection. Moreover, as before (cf. Remark 72.2.5), the spacetime manifolds M will always be assumed to be globally hyperbolic, i.e. admit a Cauchy surface.

[4]Here, it should be understood that $\mathcal{A}(\overline{O})$ is the algebra generated by all algebras $\mathcal{A}(Q)$, where Q ranges over the causally closed subsets of $\overline{O}$.

[5]This could potentially cause confusion with other notations used in this text. Here, **Loc** has nothing to do with the category of locales from Chapter 6.

72.3.1 Haag–Kastler axioms

The Haag–Kastler axioms (Section 72.2.1) can be generalized to the perturbative setting as follows. Instead of considering functors

$$\mathcal{O} : \mathbf{Caus} \to \mathbf{C^*Alg}\,, \tag{72.5}$$

one considers functors

$$\mathcal{O} : \mathbf{Caus} \to *\mathbf{Alg}\,, \tag{72.6}$$

In fact, the category on the right-hand side does not suffice yet. To consider convergence issues, one should look at topological $*$-algebras and, to work perturbatively, one should look at formal power series. Moreover, to handle gauge theory, the BV formalism (Section 54.3) will be used and homological algebra is required. Hence, the full version of the theory is concerned with functors of the form

$$\mathcal{O} : \mathbf{Caus} \to \mathbf{Ch}_\bullet\big(\mathbf{Top}*\mathbf{Alg}[[\hbar]]\big)\,, \tag{72.7}$$

As such, the isotony condition is still satisfied. However, the locality condition (72.1) has to be modified. Requiring this at the level of the full algebra would be too strong due to the presence of auxiliary fields. However, passing to cohomology suffices.

Axiom 72.7 (Causal locality). Let O_1, O_2 be spacelike separated, then

$$\big[H_\bullet(\mathcal{O}(O_1)), H_\bullet(\mathcal{O}(O_2))\big] = 0\,. \tag{72.8}$$

Again, this should be understood within a larger algebra.

Axiom 72.8 (Time slice). If O_1 contains a Cauchy surface of O_2, the morphism $\mathcal{O}(O_1 \hookrightarrow O_2)$ is a quasi-isomorphism.

@@ COMPLETE (Rejzner et al.) @@

72.3.2 Functional analysis

To handle the dynamics, one first has to generalize the notion of Lagrangians and actions from Chapter 53. Instead of simply considering a local action functional $S : \mathcal{E} \to \mathbb{R}$, one uses a generalization:

$$S : C_c^\infty(M) \to \mathcal{F}_{\mathrm{loc}}(\mathcal{E})\,, \tag{72.9}$$

i.e. one uses local functional-valued distributions on M. Because local functionals are of the form

$$F : \phi \mapsto \int_M (j^\infty\phi)^* L\,\mathrm{Vol}\,, \tag{72.10}$$

the standard example of such generalized action functionals is simply given by taking a pointwise product:

$$S[f] : \phi \mapsto \int_M f(j^\infty \phi)^* L \,\mathsf{Vol}\,. \tag{72.11}$$

This approach has two benefits. First of all, one does not have to restrict to compact sets to have a well-defined integration-by-parts formula, since all functions under the integral sign are now automatically compactly supported on M. Secondly, one has a straightforward way to implement adiabatic switching, an important ingredient for perturbation theory, since the function f can be treated as a position-dependent coupling constant.

Remark 72.3.1 (Variational bicomplex). In Chapter 53, the variational bicomplex was implicitly assumed as the framework for doing variational calculus. The main assumption of this framework, however, is locality. The functional approach, on the other hand, does allow for nonlocal functionals. For completeness, algebraically, a local functional is characterized by the following conditions:

1. For all ϕ, ψ and $\chi \in \mathcal{E}$ such that $\mathsf{supp}(\phi) \cap \mathsf{supp}(\chi) = \emptyset$:

$$F(\phi + \psi + \chi) = F(\phi + \psi) - F(\psi) + F(\psi + \chi)\,. \tag{72.12}$$

2. The wave front set (Definition 36.5.3) of $F^{(n)}$ is orthogonal to the tangent bundle $T\Delta_n$ of the diagonal of M for all $n \in \mathbb{N}$.

For $n = 1$, this implies that $F'(\phi)$ is a smooth function. Moreover, this also implies that derivatives of local functionals are supported on the diagonals Δ_n.

For many purposes in modern field theory, one needs to enlarge the space of sections to a graded vector space (e.g. to include fermionic fields). Functionals on such graded spaces are defined as follows.

Definition 72.3.2 (Graded functional). One can show that, assuming nuclearity of the relevant spaces, the (strong) dual to the space sections of an exterior tensor product is isomorphic to the completed projective tensor product of the (strong) dual of the spaces of sections of the individual bundles:

$$\Gamma(E^{\boxtimes k})' \cong \left(\Gamma(E)'\right)^{\widehat{\otimes}_\pi k}\,. \tag{72.13}$$

By restricting on the right-hand side to the subspaces of (anti)symmetric tensors, one can define the spaces of (anti)symmetric functionals on exterior tensor powers (after completion). The space of antisymmetric functionals on $\mathcal{E}$ is then defined as follows:

$$\mathcal{F}(\mathcal{E}) := C^\infty\left(\mathcal{E}_0, \bigoplus_{\substack{k,l=0\\ i_1+\cdots i_l=k}} \Gamma_{i_1|\underline{i_2}|i_3|\ldots}\left(\boxtimes_{j=1}^l E[-j]_0^{\boxtimes i_j}\right)'\right), \tag{72.14}$$

where the sequence $i_1|\underline{i_2}|i_3|\dots$ denotes symmetric and antisymmetric tensor powers (underlined indices denote antisymmetric tensors). Note that, because of the desuspension operator (Definition 23.5.7), the tensor powers are taken over bundles of degree 0. Equivalently, the desuspension could have been left out such that only the symmetric tensor powers have to be used:

$$\mathcal{F}(\mathcal{E}) = C^\infty\left(\mathcal{E}_0, \bigoplus_{\substack{k,l=0\\ i_1+\dots i_l=k}} \Gamma_{i_1|i_2|i_3|\dots}\left(\boxtimes_{j=1}^l E_j^{\boxtimes i_j}\right)'\right). \tag{72.15}$$

The closure of the subspace of local functionals, those of compact support that are induced by functions on some jet bundle, under wedge products is called the space of **multilocal functionals** $\mathcal{F}_{\mathrm{ml}}(\mathcal{E})$. These model the physical observables of the theory.

Aside from the standard functional derivative, in the sense of Gateaux (Definition 25.3.12), one can now also define a 'graded derivative'.

Formula 72.3.3 (Graded derivative). Denote the functional target space

$$\bigoplus_{\substack{l=0\\ i_1+\dots i_l=k}} \Gamma_{i_1|\underline{i_2}|i_3|\dots}\left(\boxtimes_{j=1}^l E[-j]_0^{\boxtimes i_j}\right)'$$

by $\mathcal{O}^k(\mathcal{E})$ and consider a homogeneous element $F \in \mathcal{O}^k(\mathcal{E})$. The left (graded) derivative is defined as follows:

$$\begin{cases} \left\langle \dfrac{\delta^l F}{\delta\phi}(v), w\right\rangle := F(w \wedge v) & \text{if } k > 0\,, \\[2ex] \dfrac{\delta^l F}{\delta\phi} := 0 & \text{if } k = 0\,. \end{cases} \tag{72.16}$$

Similarly, the right (graded) derivative is defined by

$$\begin{cases} \left\langle \dfrac{\delta^r F}{\delta\phi}(v), w\right\rangle := F(v \wedge w) & \text{if } k > 0\,, \\[2ex] \dfrac{\delta^r F}{\delta\phi} := 0 & \text{if } k = 0\,. \end{cases} \tag{72.17}$$

72.3.3 Equations of motion

Using the functional approach, the Euler–Lagrange derivative $\delta_{\mathrm{EL}} L : \mathcal{E} \to \mathcal{E}_c'$ is given by the following formula:

$$\langle \delta_{\mathrm{EL}} L(\phi), \psi\rangle := \langle L'[f](\phi), \psi\rangle\,, \tag{72.18}$$

for any cutoff function f on $\mathsf{supp}(\psi)$ (Definition 19.3.19).

Remark 72.3.4 (Functoriality). A more categorical formulation is that a generalized Lagrangian is a natural transformation

$$L : C_c^\infty \Rightarrow \mathcal{F}_{\text{loc}} \tag{72.19}$$

that satisfies the following locality conditions:

1. **Additivity**: $L[f+g+h] = L[f+g] - L[g] + L[g+h]$ if $\mathsf{supp}(f) \cap \mathsf{supp}(h) = \emptyset$, and
2. **Support**: $\mathsf{supp}(L[f]) \subseteq \mathsf{supp}(f)$.

To obtain an action functional S, one quotients out by the relation

$$\mathsf{supp}(L_1[f] - L_2[f]) \subset \mathsf{supp}(\mathrm{d}f)\,. \tag{72.20}$$

This identifies Lagrangians that differ up to a total divergence, since the above equation for the Euler–Lagrange derivative is invariant under the addition of such terms. In fact, it is invariant under the addition of any generalized Lagrangian that is supported outside the level set $f^{-1}(1)$. (This is exactly the condition that is quotiented out.)

The solution space of the Euler–Lagrange equations is denoted by $\mathcal{E}_S := \mathsf{ker}(\delta_{\text{EL}} L)$. As in Section 54.2.2, the physically relevant observables are the multilocal functionals on $\mathcal{E}_S$:

$$\mathcal{F}_S := \mathcal{F}_{\text{ml}} / \mathcal{J}_S\,, \tag{72.21}$$

where $\mathcal{J}_S$ is the ideal generated by the functionals vanishing on $\mathcal{E}_S$.

The second variational derivative of the action then satisfies:

$$\langle \delta^2 S(\phi), \psi_1 \otimes \psi_2 \rangle = \langle L''[f](\phi), \psi_1 \otimes \psi_2 \rangle\,, \tag{72.22}$$

where f is a cutoff function on $\mathsf{supp}(\psi_1) \cup \mathsf{supp}(\psi_2)$. Because the Lagrangians are local, the linear maps $S''(\phi)$ can be extended from $\mathcal{E}_c \times \mathcal{E}_c$ to $\mathcal{E} \times \mathcal{E}_c$. By the Kernel Theorem 19.2.6, this gives rise to a linear operator $P_S(\phi) : \mathcal{E} \to \mathcal{E}^*$.

Axiom 72.9 (Normal hyperbolicity). The operator $P_S(\phi)$ is normally hyperbolic, i.e. a hyperbolic differential operator whose principal symbol defines a metric, for all solutions ϕ.

This axiom implies that $P_S(\phi)$ admits unique Green's operators $\Delta^{R/A}_{S,\phi} : \mathcal{E}_c^* \to \mathcal{E}$:

$$\begin{aligned} P_S(\phi) \circ \Delta^{R/A}_{S,\phi} &= \mathbb{1}_{\mathcal{E}_c^*} \\ \Delta^{R/A}_{S,\phi} \circ P_S(\phi) &= \mathbb{1}_{\mathcal{E}_c} \end{aligned} \tag{72.23}$$

with

$$\begin{aligned} \mathsf{supp}(\Delta^R_{S,\phi}(\psi)) &\subset J^-(\mathsf{supp}(\psi)) \\ \mathsf{supp}(\Delta^A_{S,\phi}(\psi)) &\subset J^+(\mathsf{supp}(\psi))\,. \end{aligned} \tag{72.24}$$

These operator also define integral kernels through the kernel theorem:

$$\Delta^R_{S,\phi}(x,y) = \Delta^A_{S,\phi}(y,x)\,. \tag{72.25}$$

Their difference gives the causal propagator:

$$\Delta_{S,\phi} := \Delta^R_{S,\phi} - \Delta^A_{S,\phi}\,. \tag{72.26}$$

Property 72.3.5. If the action S is a quadratic functional, the operator $P_S(\phi)$ is constant:

$$\forall \phi \in \mathcal{E} : P_S(\phi) = P\,. \tag{72.27}$$

Moreover, in this case, the Euler–Lagrange equation becomes $P(\phi) = 0$. In general, the equation $P_S(\phi)\psi = 0$ is called the **linearised equation of motion**.

The next step is to equip the space of functionals with a suitable Poisson bracket. As in Chapter 53, this will be the Peierls bracket.

Definition 72.3.6 (Peierls bracket). Consider two multilocal functionals $F, G \in \mathcal{F}_{\mathrm{ml}}(\mathcal{E})$ together with an action functional S.

$$\{F,G\}_S(\phi) := \left\langle F'(\phi), \Delta_{S,\phi} G'(\phi)\right\rangle. \tag{72.28}$$

The main problem with the Peierls bracket is that it does not define a Poisson algebra structure on $\mathcal{F}_{\mathrm{ml}}(\mathcal{E})$ since this space is not closed under the bracket. The Hörmander criterion for pointwise multiplication of distributions gives some intuition for how to construct such a space.

Property 72.3.7. The integral kernel of the causal propagator has the following wave front set:

$$\mathsf{WF}(\Delta_{(S,\phi)}) = \{(x,k,x',-k') \in T^*M^2 \mid (x,k) \sim (x',k')\}\,, \tag{72.29}$$

where $\sim$ indicates that the points are connected by the lift of a null geodesic whose fibrewise projection is given by the metric dual:

$$\big(x(t),k(t)\big) \equiv \big(\gamma(t), g(\dot\gamma(t),\cdot)\big) \tag{72.30}$$

for some null geodesic γ on M.

Definition 72.3.8 (Microcausal functionals). A functional $F \in \mathcal{E}^*$ that is compactly supported and such that

$$\mathsf{WF}(F^{(n)}(\phi)) \subset \Xi_n \tag{72.31}$$

for all $n \in \mathbb{N}$ and $\phi \in \mathcal{E}$, where

$$\Xi_n := T^*M^n \backslash \{(x,k) \in T^*M^n \mid k \in \overline{V}^+(x) \cup \overline{V}^-(x)\}\,. \tag{72.32}$$

Microcausal functionals hare, hence, exactly those functionals for which the wafront set excludes (the closure of) causal lightcones. Moreover, a microcausal functional is said to be **strongly microcausal** if $F'(\phi)$ is C^1 for all $\phi \in \mathcal{E}$ and $\phi \mapsto F'(\phi)$ is also C^1.

Property 72.3.9. The space of strongly microcausal functionals, equipped with the Peierls bracket, forms a Poisson algebra. If S is quadratic, i.e. the causal propagator does not depend on the fields, the space of microcausal functionals also forms a Poisson algebra under the Peierls bracket.

Definition 72.3.10 (Wightmann propagator). Consider the causal propagator Δ_S. By Property 72.3.7, its wave front set consists of two part lying in the closed past and closed future lightcones. This corresponds to the following decomposition:

$$\frac{i}{2}\Delta_S = \Delta_S^H - H\,, \tag{72.33}$$

where

$$\mathsf{WF}(\Delta_S^H) = \{(x,k,x',-k') \in T^*M^2 \mid (x,k) \sim (x',k') \wedge k \in \overline{V}^+(x)\} \tag{72.34}$$

and

$$H(x,y) = H(y,x)\,. \tag{72.35}$$

The Wightmann or **Hadamard** propagator represents the positive-frequency part of the causal propagator.

72.3.4 Symmetries

Now, to include symmetries, one has to introduce vector fields. On general infinite-dimensional manifolds (Section 44.2), multiple approaches exist. One way is to regard $\mathcal{E}$ as a trivial $\mathcal{E}$-manifold. This way, the space of vector fields, given as sections of the tangent bundle, is simply $C^\infty(\mathcal{E},\mathcal{E})$.

Definition 72.3.11 (Multilocal multivector fields). The space of multivector fields is defined as the space of functionals on the odd cotangent bundle:

$$\mathfrak{X}_{\text{loc}}(\mathcal{E}) := \mathcal{F}_{\text{loc}}\big(\mathcal{E} \oplus \mathcal{E}^*[1]\big)\,. \tag{72.36}$$

The subspace $\mathfrak{X}^1_{\text{loc}}(\mathcal{E})$ consists of the derivations on $\mathcal{F}_{\text{loc}}(\mathcal{E})$. The algebraic completion under wedge products defines the space of multilocal multivector fields $\mathfrak{X}_{\text{ml}}(\mathcal{E})$. The subspace $\mathfrak{X}^1_{\text{ml}}(\mathcal{E})$, the completion as a $\mathcal{F}(\mathcal{E})$-module of the local vector fields, are then the derivations of $\mathcal{F}_{\text{ml}}(\mathcal{E})$.

The action of a vector field $X \in \mathfrak{X}^1_{\text{ml}}(\mathcal{E})$ is given by

$$\partial_X F(\phi) := \langle F'(\phi), X(\phi)\rangle\,. \tag{72.37}$$

In the physics literature, this is often formally written as

$$\partial_X F(\phi) = \int_M X_x(\phi)\frac{\delta F}{\delta\phi(x)}\,. \tag{72.38}$$

Note that, for any multilocal vector field, the following functional vanishes on the zero locus $\mathcal{E}_S$ of $\delta_{\mathrm{EL}}S$:

$$\phi \mapsto \delta_S(X)(\phi) := \partial_X S(\phi) = \langle \delta_{\mathrm{EL}} S(\phi), X(\phi) \rangle \,. \tag{72.39}$$

These functionals form an ideal under pointwise multiplication and the quotient ideal gives the space of physically distinguishable observables, i.e. the on-shell functionals, as in Eq. (72.21). However, in general, this ideal is quite complicated and the subideal $\delta_S(\mathfrak{X}^1)$ is proper. Since the quantized theory is of relevance and, hence, only linear effects are of interest, the solution space will be redefined as:

$$\mathcal{F}_S := \mathcal{F}_{\mathrm{ml}} / \delta_S(\mathfrak{X}^1_{\mathrm{ml}}) \,. \tag{72.40}$$

As in Chapter 54 and Chapter 53, one can now characterize this space in homological terms. To this end, one constructs the Koszul complex

$$\cdots \xrightarrow{\delta_S} \Lambda^2 \mathfrak{X}^1_{\mathrm{ml}}(\mathcal{E}) \xrightarrow{\delta_S} \mathfrak{X}^1_{\mathrm{ml}}(\mathcal{E}) \xrightarrow{\delta_S} \mathcal{F}_{\mathrm{ml}} \longrightarrow 0 \,, \tag{72.41}$$

where δ_S is extended to multivector fields through the (right-)graded Leibniz rule. The first homology group of this complex represents the space of nontrivial symmetries, they vanish identically on $\mathcal{E}_S$. The differential δ_S can also be (locally) generated by a Poisson-like bracket, the antibracket. In this formulation, it is given by the Schouten–Nijenhuis bracket on multivector fields (Definition 34.3.24). Definition 54.3.9 for the antifield (ignoring the ghosts and antighosts for now) then corresponds to the expression

$$\{X, Y\}(\phi) := \left\langle \frac{\delta^r X}{\delta \phi}, \frac{\delta^l Y}{\delta \phi^\dagger} \right\rangle - \left\langle \frac{\delta^r X}{\delta \phi^\dagger}, \frac{\delta^l Y}{\delta \phi} \right\rangle , \tag{72.42}$$

where the derivatives with respect to ϕ and $\phi^\dagger$ should be interpreted as functional (Gateaux) and graded derivatives (Formula 72.3.3) with respect to the base and fibre coordinates of the odd cotangent bundle $\mathcal{E} \oplus \mathcal{E}^*[1]$.[6]

72.3.5 Quantization

To quantize the above classical formalism, the framework of deformation quantization is used. Since all spaces considered in the previous sections are linear, Moyal quantization (Definition 59.2.2) is applicable:

$$F * G := \mu \circ \exp(\tfrac{i\hbar}{2} \pi_\Delta)(F \otimes G) \,, \tag{72.43}$$

where π_Δ is the Poisson bivector induced by the causal propagator (see the definition of Peierls bracket above). However, aside from the causal propagator Δ, one also has the Wightmann propagator Δ^H. This operator also induces a Moyal product:

$$F *_H G := \mu \circ \exp(\hbar \pi_{\Delta^H})(F \otimes G) \,, \tag{72.44}$$

[6]Note that multivector fields, on which the Schouten–Nijenhuis bracket is defined, are essentially (graded) functionals on the odd tangent bundle.

Property 72.3.12. On the space of regular functionals, the causal and Wightmann star products are isomorphic:

$$F *_H G = \alpha_H(\alpha_H^{-1} F * \alpha_H^{-1} G)\,, \tag{72.45}$$

where

$$\alpha_H := \exp\left(\frac{\hbar}{2}\left\langle H, \frac{\delta^2}{\delta\phi^2}\right\rangle\right). \tag{72.46}$$

The transformation from one star product to the other corresponds to Wick's theorem in ordinary QFT.

The space of regular functionals is sequentially dense in the space of microcausal functionals, so it makes sense to consider sequences that converge to an element in $\mathcal{F}_{\mu C}$.

Definition 72.3.13 (Quantum algebra). The quantum algebra $\mathfrak{U}^H$ of the free theory S_0 is defined as the extension of $\mathcal{F}_{\text{reg}}[[\hbar]]$ by elements of the form $\lim_{n\to\infty} \alpha_H^{-1}(F_n)$ for $(F_n)_{n\in\mathbb{N}}$ a convergent sequence in $\mathcal{F}_{\mu C}$. The Moyal product $*_H$ is defined as above and involution is given by

$$F^* := \alpha_H^{-1}\left(\overline{\alpha_H F}\right). \tag{72.47}$$

The support of an element $F \in \mathfrak{U}^H$ is given by

$$\mathsf{supp}(F) := \mathsf{supp}(\alpha_H F)\,. \tag{72.48}$$

Remark 72.3.14. The decomposition of the causal propagator in terms of the Wightmann propagator is not unique. For two such decompositions the difference $H - H'$ is a smooth symmetric bisolution:

$$P_x(H - H')(x,y) = P_y(H - H')(x,y) = 0\,. \tag{72.49}$$

The associated morphism $\alpha_{H-H'}$ is an isomorphism and relates the quantum algebras $\mathfrak{U}^H$ and $\mathfrak{U}^{H'}$. As abstract algebras these are isomorphic, but the choice of Wightmann propagator fixes how classical functionals are realized in the quantum algebra.

Property 72.3.15. Let (M,g) be a globally hyperbolic manifold equipped with a vector bundle $\pi : E \to M$. If S_0 is a quadratic action functional that induces a formally self-adjoint operator P, the net of quantum algebras $O \mapsto \mathfrak{U}(O)$ of elements with support in O satisfies the Haag–Kastler axioms.

As before the on-shell observables are defined as those elements of the quotient algebra

$$\mathfrak{U}_{S_0} := \mathfrak{U}/\mathfrak{J}_{S_0}\,, \tag{72.50}$$

where $\mathfrak{J}_{S_0}$ is generated by the functionals $\langle \delta_{\mathrm{EL}} S, X\rangle$ for $X \in \mathfrak{X}_{\mathrm{loc}}$. The Moyal product is extended to microcausal multivector fields by replacing the pointwise multiplication in Eq. (72.44) by the wedge product:

$$\begin{aligned}&\langle X *_H Y(\phi), v_1 \otimes \cdots \otimes v_{p+q}\rangle := \\ &\sum_{i=0}^{\infty} \frac{\hbar^i}{i!p!q!} \sum_{\sigma \in S_{p+q}} \left\langle\!\!\left\langle \frac{\delta^i X}{\delta \phi^i}(\phi), v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(p)} \right\rangle, (\Delta^H_{S_0})^{\otimes i} \left\langle \frac{\delta^i Y}{\delta \phi^i}(\phi), v_{\sigma(p+1)} \otimes \cdots \otimes v_{\sigma(p+q)} \right\rangle\!\!\right\rangle .\end{aligned} \tag{72.51}$$

@@ COMPLETE @@

72.4 Prequantum field theory

Now, consider an (extended) TQFT. In full generality such a TQFT assigns to every manifold M an (∞-)stack of fields. For a cobordism, this induces a span (Definition 3.4.65) by restriction to the boundary components. Composition of cobordisms corresponds to pullback of spans. So, a TQFT induces a (monoidal) functor $\mathbf{Bord}^d \to \mathbf{Span}(\mathbf{H})$, where $\mathbf{H}$ is the $(\infty, 1)$-category of $\infty\mathbf{Grpd}$-valued ∞-sheaves. When considering cobordisms between cobordisms, one obtains spans between spans. So, eventually, a d-dimensional TQFT can be seen to be equivalent to a symmetric monoidal (∞, n)-functor $\mathbf{Bord}^d \to \mathbf{Corr}_d(\mathbf{H})$, where $\mathbf{Corr}_n(\mathbf{H})$ is the (∞, n)-category of *n-fold correspondences* in $\mathbf{H}$. By a theorem of *Lurie*, correspondences admit a (symmetric) monoidal structure induced by that on $\mathbf{H}$. The cobordism hypothesis now immediately implies that any such functor is uniquely characterized by a (moduli) stack of fields. These theories are called **(local) prequantum field theories (pQFT)**.

Property 72.4.1. Any local pQFT is a (topological) σ-model:

$$\mathsf{Fields}(M) \cong \left[\textstyle\int M, \mathbf{Fields}\right], \tag{72.52}$$

where $\mathbf{Fields}$ is the classifying stack and $\int$ sends a manifold to its homotopy type, i.e. it is the shape modality (Property 44.3.8).

Now, to obtain a reasonable quantum theory, one also needs an action principle. To a d-dimensional manifold it should assign a phase, i.e. a map $\mathbf{H} \to \mathsf{U}(1)$, to its boundary components it should assign a gauge field, i.e. a map $\mathbf{H} \to \mathbf{B}\mathsf{U}(1)$, etc. Combining this with the pQFT, one obtains a functor $\mathbf{Bord}^d \to \mathbf{Corr}_d(\mathbf{H}/_{\mathbf{B}^d\mathsf{U}(1)})$ to the d-fold correspondences in the slice topos over the classifying stack of circle d-bundles. Because the classiyfing stacks are E_∞*-algebras* in $\mathbf{H}$, i.e. commutative monoids in $\mathbf{H}$, the slice topos admits the structure of a symmetric monoidal category. Therefore one can extend the action functional to disjoint unions of cobordisms.

Definition 72.4.2 (Action functional). An action functional for $\mathbf{Bord}^d \to \mathbf{Corr}_d(\mathbf{H})$ is a lift (of monoidal functors) to a symmetric monoidal (∞, n)-functor $\mathbf{Bord}^d \to \mathbf{Corr}_d(\mathbf{H}/_{\mathbf{B}^d\mathsf{U}(1)})$.

By the cobordism hypothesis, such a functional is uniquely defined by a morphism $\mathbf{Fields} \to \mathbf{B}^d\mathrm{U}(1)$, i.e. by a circle d-bundle on the space of fields.

One can also rephrase this assignment in terms of (higher) holonomy functors. Any map $U \to \left[\int M, \mathbf{B}^d\mathrm{U}(1)\right]$ is given by th assignment of a flat d-bundle on $U \times M$, by the adjunction $\int \dashv \flat$. Taking the holonomy of this bundle defines a functor $[\int M, \mathbf{B}^d\mathrm{U}(1)] \to \mathbf{B}^{d-k}\mathrm{U}(1)$, where k is the dimension of U. Composition with the classifying functor $\mathbf{Fields} \to \mathbf{B}^d\mathrm{U}(1)$ induces an assignment of $(d-k)$-bundles to the spaces of fields on M. This is the generalization of the (Lagrangian) action principle, where integration of Lagrangians is replaced by assigning higher holonomies.

Chapter 73

Conformal Field Theory

References for this chapter are Fuchs, Schweigert, Wood, and Yang (2023); Gui (2023); Schottenloher (2008) and lecture notes by *Schellekens*. For an introduction to Riemannian and conformal geometry, see Chapter 36 and, in particular, Section 36.4. The TCFT section is based on the seminal work by Fjelstad, Fuchs, Runkel, and Schweigert (2005); Fuchs, Runkel, and Schweigert (2002, 2004a, 2004b, 2005). The section on strange correlators is mainly based on Vanhove et al. (2018); You, Bi, Rasmussen, Slagle, and Xu (2014).

Chapter Contents

73.1 Introduction . . . **1221**

73.2 In dimension $d = 2$ . . . **1221**

73.2.1 Minkowski space . . . 1222

73.3 Quantization in $d = 2$ . . . **1223**

73.3.1 Radial quantization . . . 1224

73.3.2 Virasoro algebra . . . 1225

73.3.3 Representation theory . . . 1225

73.4 Vertex operator algebras . . . **1226**

73.4.1 Definition . . . 1226

73.4.2 Rational CFTs . . . 1227

73.5 Extensions ♣ . . . **1227**

73.5.1 TCFT . . . 1227

73.5.2 Strange correlator . . . 1227

73.1 Introduction

Property 73.1.1 (Stress-energy tensor). Consider a theory that is invariant under conformal transformations. The generator of general coordinate transformations is the stress-energy tensor T (the associated current is $\mathcal{J}_\mu = T_{\mu\nu}\varepsilon^\nu$). Conformal invariance implies that the stress-energy tensor is traceless:

$$T^\mu_\mu = 0\,. \tag{73.1}$$

@@ MOVE @@

73.2 In dimension $d = 2$

In dimension 2 (in Euclidean signature), something special happens. By inserting $d = 2$ in the conformal Killing equation (36.57), the Cauchy–Riemann equations 17.2.4 are obtained. The scale factor can thus be written as

$$\kappa(z) = \left|\frac{\partial f}{\partial z}\right|^2 \tag{73.2}$$

for some analytic function $f : \mathbb{C} \to \mathbb{C}$. Because of this, complex coordinates will be used from here on. Switching to complex coordinates also has important consequences for the metric and stress-energy tensor:

$$\begin{aligned} g_{zz} = g_{\bar{z}\bar{z}} = 0 \qquad & g_{z\bar{z}} = \frac{1}{2}\,, \\ \partial_z T_{\bar{z}\bar{z}} = \partial_{\bar{z}} T_{zz} = 0 \qquad & T_{z\bar{z}} = 0\,. \end{aligned} \tag{73.3}$$

The stress-energy tensor thus contains a meromorphic[1] component $T_{zz} \equiv T(z)$ and an antimeromorphic component $T_{\bar{z}\bar{z}} \equiv \overline{T}(\bar{z})$.

Definition 73.2.1 (Witt algebra). Infinitesimally, this gives an infinite-dimensional algebra. The generators can be chosen to be

$$\begin{aligned} l_n(z) &:= -z^{n+1}\partial_z\,, \\ \bar{l}_n(\bar{z}) &:= -\bar{z}^{n+1}\partial_{\bar{z}}\,. \end{aligned} \tag{73.4}$$

These generate the transformations $z \mapsto z - z^{n+1}$ and $\bar{z} \mapsto \bar{z} - \bar{z}^{n+1}$, respectively. They also give rise to isomorphic Lie algebras with the following commutation relation:

$$[l_m, l_n] = (m-n)l_{m+n}\,. \tag{73.5}$$

This Lie algebra is called the Witt algebra.

[1] The literature often just calls this holomorphic.

Remark 73.2.2 (Conformal group). Often, one finds in the literature that the conformal group of a 2D CFT is infinite-dimensional. However, this statement is not entirely true. It is true that the Witt algebra is infinite dimensional, but one cannot globally exponentiate all generators l_m. First of all, it should be noted that the space of all holomorphic functions does not even have a group structure because the composition of holomorphic functions does not have to be holomorphic. The correct conformal group for 2D Euclidean CFTs is the Möbius group $\mathsf{PSL}_2(\mathbb{C})$ (Definition 11.2.1). This group is obtained as the Lie group generated by l_0 and $l_{\pm 1}$, which are the only generators that can be exponentiated globally.

What is also true is that the conformal group of two-dimensional Minkowski space is infinite dimensional. It can be shown that $\mathsf{Conf}(\mathbb{R}^{1,1})$ is isomorphic to $\mathsf{Diff}(S^1)_+ \times \mathsf{Diff}(S^1)_+$, where the orientation-preserving diffeomorphism group $\mathsf{Diff}(S^1)_+$ is an infinite-dimensional Lie group (see the intermezzo further below).

At last, it should be noted that, although the conformal group of $\mathbb{R}^{2,0}$ is finite dimensional, the infinite-dimensionality of the Witt algebra (and of its extensions) is sufficient for all physical purposes. The algebraic constraints turn this theory into an integrable theory and allow to solve it exactly.

Definition 73.2.3 (Primary field). A field that transforms tensorially under global conformal transformations:

$$\phi'(z', \bar{z}') = \left(\frac{\partial f}{\partial z}\right)^h \left(\frac{\partial f}{\partial \bar{z}}\right)^{\bar{h}} \phi\big(f(z), \bar{f}(\bar{z})\big) \,. \tag{73.6}$$

The tuple $(h, \bar{h})$ is called the **conformal weight** of the field.

73.2.1 Minkowski space

As mentioned in the remark above, the conformal group of 2D Minkowski space is infinite dimensional. The theory of infinite-dimensional manifolds is, however, a bit more intricate than the theory of finite-dimensional manifolds. A little introduction is, therefore, in order.

Definition 73.2.4 (Fréchet manifold). A Hausdorff space M together with an atlas of coordinate charts (U, φ) such that $\varphi : U \to F_U$ are homeomorphisms onto a Fréchet space and such that the transition functions are smooth maps (of Fréchet spaces).

Remark 73.2.5. Because the dual of Fréchet space that is not Banach is never Fréchet itself, the treatment of the tangent and cotangent bundles becomes problematic. Moreover, existence and uniqueness properties of ODEs can fail in infinite dimensions, hence, many statements of ordinary differential geometry might also fail in this setting.

Using the definition of Fréchet manifolds, one can start to analyze the group $\mathsf{Diff}(S^1)_+$.

First of all, look at the space $C^\infty(S^1, S^1)$. This space has the structure of a Fréchet manifold modelled on the Fréchet space $\mathfrak{X}(S^1)$ of vector fields on the circle[2]:

$$\mathfrak{X}(S^1) = \left\{ \xi(\theta)\frac{\partial}{\partial\theta} \,\middle|\, \theta \in C^\infty(S^1) \right\}. \tag{73.7}$$

Let V_0 be the set of vector fields that have norm $\|v\| \leq \pi$ and let U_0 be the set of smooth mappings $f \in C^\infty(S^1, S^1)$ such that $f(\theta) \neq -\theta$ for all $\theta \in S^1$. There exists a diffeomorphism $\psi : V_0 \to U_0$ that assigns to any vector field v the function $\psi_v : S^1 \to S^1$ such that the arc between θ and $\psi_v(\theta)$ has length $\|v(\theta)\|$. If an open subset $U \subset U_0$ of diffeomorphisms is chosen, then a chart (U, ψ^{-1}) is obtained around the identity map. Charts around any diffeomorphism $f : S^1 \to S^1$ are obtained by left multiplication of U.

The Lie algebra of $\mathrm{Diff}(S^1)_+$ is, therefore, given by $\mathfrak{X}(S^1)$, but the induced Lie bracket is the commutator of vector fields with the opposite sign:

$$[\cdot,\cdot]_{\mathrm{Lie}} = -[\cdot,\cdot]_{\mathfrak{X}(S^1)}.$$

Now, it is interesting to note that the Witt algebra is actually a subalgebra of $\mathfrak{X}(S^1)$. Consider the maps $\xi_n(\theta) := -ie^{in\theta}$ (the minus sign is a convention). The associated vector fields satisfy

$$\left[\xi_k(\theta)\frac{\partial}{\partial\theta}, \xi_l(\theta)\frac{\partial}{\partial\theta}\right] = -i(l-k)\xi_{k+l}(\theta)\frac{\partial}{\partial\theta}. \tag{73.8}$$

These are exactly the relations for the Witt algebra in Definition 73.2.1.

73.3 Quantization in $d = 2$

By Property 36.4.5, the conformal group of a pseudo-Euclidean space of signature (p, q) is given by the special orthogonal group $\mathrm{SO}(p+1, q+1)$. For Minkowski space, this becomes $\mathrm{SO}(4, 2)$. However, with the current section in mind, it is important to look at the interaction between representation theory and quantum mechanics.

When treating the concept of spin quantum-mechanically, one is forced to pass from the rotation group $\mathrm{SO}(3)$ to its double cover $\mathrm{SU}(2)$. Then, when passing to special relativity, the Lorentz group $\mathrm{SO}^\uparrow(3, 1)$ also has to be replaced by its double cover $\mathrm{SL}_2(\mathbb{C})$. The same story repeats itself here. The conformal group should be replaced by its double cover $\mathrm{SU}(2, 2)$. This is how *twistor theory* comes into the picture (Penrose, 1967).

[2] By this definition, it can be seen that $\mathfrak{X}(S^1)$ is isomorphic (as a Lie algebra) to the mapping space $C^\infty(S^1, \mathbb{R})$. Henceforth, the vector fields v will be identified with their corresponding smooth function ξ.

73.3.1 Radial quantization

The charge of a conserved current $\mathcal{J}^\mu$ is generally given by Eq. (53.11):

$$Q = \int_\Sigma \mathcal{J}^0(x,t)\,, \tag{73.9}$$

where Σ is a spacelike hypersurface. Often, the spatial dimension will be compactified (this can be seen as a regularization procedure) and the time dimension will be *Wick-rotated*. By a conformal transformation, one can then go back to the plane, mapping one end of the cylinder to the origin and the other side to a circle at ∞. After these transformations, one obtains the following form for an operator on the plane:

$$Q_\varepsilon = \frac{1}{2\pi i}\oint \varepsilon(z)T(z)\,dz + \frac{1}{2\pi i}\oint \varepsilon(\overline{z})\overline{T}(\overline{z})\,d\overline{z}\,. \tag{73.10}$$

As usual, an infinitesimal transformation of a field ϕ is given by the commutator

$$[Q_\varepsilon, \phi(w,\overline{w})] \tag{73.11}$$

or, in integral form, by[3]

$$\frac{1}{2\pi i}\oint \varepsilon(z)\big[T(z)\phi(w,\overline{w}) - \phi(w,\overline{w})T(z)\big]\,dz\,. \tag{73.12}$$

However, because all the objects in this formula are operators, operator ordering should be taken into account. On the plane, this is given by the so-called **radial ordering**:

$$\mathcal{R}(A(z,\overline{z})B(w,\overline{w})) := \begin{cases} A(z,\overline{z})B(w,\overline{w}) & \text{if } |z| > |w|\,, \\ B(w,\overline{w})A(z,\overline{z}) & \text{if } |w| > |z|\,. \end{cases} \tag{73.13}$$

After a deformation of the integration contour, the following general formula is obtained:

$$[Q_\varepsilon, \phi(w,\overline{w})] = \frac{1}{2\pi i}\oint \varepsilon(z)\mathcal{R}\big(T(z)\phi(w,\overline{w})\big)\,dz\ , \tag{73.14}$$

where the contour is a circle around the point $w \in \mathbb{C}$. For primary fields with conformal weight h, one can also write an infinitesimal transformation as

$$\delta_\varepsilon\phi(z,\overline{z}) = h(\partial_z\varepsilon(z))\phi(z,\overline{z}) + \partial_z\phi(z,\overline{z})\,. \tag{73.15}$$

Comparing these expressions leads to the following form of the operator product:

$$\mathcal{R}(T(z)\phi(w,\overline{w})) = \frac{h}{(z-w)^2}\phi(w,\overline{w}) + \frac{1}{z-w}\partial_w\phi(w,\overline{w}) + \text{higher order in } (z-w)\,. \tag{73.16}$$

An expression of this form is called an **operator product expansion** (OPE).

[3] A holomorphic split is implicitly assumed such that antiholomorphic contributions $\overline{T}(\overline{z})$ can be ignored.

73.3.2 Virasoro algebra

When quantizing classical systems, one has to replace symmetry actions by (unitary) projective representations, which are characterized by central extensions. As an example of the Lie group–Lie algebra correspondence, it can be shown that central extensions of Lie algebras (Section 32.2.6) are in correspondence with central extensions of Lie groups.

To equip a quantum theory with an action of the conformal group, one needs to construct a central extension of the Witt algebra. By applying Construction 32.2.44, one obtains the **Virasoro algebra** as a (universal) central extension of the Witt algebra by[4] $\mathbb{C}$ associated to the cocycle

$$\Theta : (L_m, L_n) \mapsto \frac{c}{12} m(m^2 - 1)\delta_{m+n,0} \,. \tag{73.17}$$

To obtain the Virasoro algebra from a more physical point of view, look at the stress-energy tensor. If the current $\mathcal{J}^n(z) := z^{n+1}T(z)$ is chosen, the transformation $z \longrightarrow z - z^{n+1}$ is obtained. In analogy with the generators of the Witt algebra, this generator is called L_n:

$$L_n := \frac{1}{2\pi i} \oint z^{n+1} T(z)\, dz \,. \tag{73.18}$$

This relation can be inverted using the Residue Theorem 17.5.22 to obtain the following expression for (the holomorphic component of) the stress-energy tensor:

$$T(z) = \sum_{n=-\infty}^{+\infty} z^{-n-2} L_n \,. \tag{73.19}$$

Using the above expression and the product operator expansion of $T(z)T(w)$, one obtains exactly the commutation relations of the Virasoro algebra:

$$[L_m, L_n] = (m - n)L_{m+n} + \frac{c}{12} m(m^2 - 1)\delta_{m+n,0} \,. \tag{73.20}$$

The occurrence of the central charge c gives a conformal anomaly on quantization (see the definition of the vacuum further below). However, the central charge does not affect the $\mathfrak{sl}_2(\mathbb{C})$ subalgebra spanned by L_{-1}, L_0 and L_1. This implies that concepts such as the conformal weight are still well-defined after quantization.

73.3.3 Representation theory

Definition 73.3.1 (Highest weight state). A state with minimal eigenvalue for L_0. Equivalently, a state that is annihilated by all generators L_n for $n \geq 1$:

$$L_n |h\rangle = 0 \,. \tag{73.21}$$

[4] The reason to extend by $\mathbb{C}$ instead of $\mathbb{R}$ is that all Lie algebras are complexified in this section.

Definition 73.3.2 (Vacuum). Consider the Virasoro generators $\{L_n\}_{n\in\mathbb{Z}}$. The vacuum $|0\rangle$ is defined as the maximally symmetric state. In terms of generators, this means that

$$L_n|0\rangle = 0 \tag{73.22}$$

for as many $n \in \mathbb{Z}$ as possible. However, due to the Virasoro commutation relations and, in particular, the central charge, this is not possible for all $n \in \mathbb{Z}$. Instead, one can only require that this expression vanishes for all $n \geq -1$.

Definition 73.3.3 (Descendants). By acting with the generators L_n, for $n \leq -1$, on a highest weight state $|h\rangle$, one obtains a whole family of states. These are called the descendants of $|h\rangle$ and together they span the Verma module (Definition 32.4.62) associated to $|h\rangle$.

73.4 Vertex operator algebras

73.4.1 Definition

Definition 73.4.1 (Vertex algebra). A vertex algebra consists of the following data:

1. a (graded) vector space $V \equiv \bigoplus_{i=0}^{+\infty} V_i$ (typically over $\mathbb{C}$),
2. a vacuum state $1 \in V_0$ (sometimes denoted by $|0\rangle$),
3. a 'shift/translation' operator $T : V \to V$, and
4. a linear multiplication $Y : V \otimes V \to V((z))$ or, equivalently, $Y : V \to \mathsf{End}(V)[[z, z^{-1}]]$, representing the state-field correspondence where $Y(v, z)$ is called a **vertex operator** (or field) at $V \in V$. The convention is that the operator Y_n is the coefficient for z^{-n-1}:

$$Y(v, z) = \sum_{n=-\infty}^{+\infty} v_n z^{-n-1} \,. \tag{73.23}$$

This allows to obtain v_n as the residue of $Y(v, z)z^n$.

These objects are required to satisfy the following conditions:

1. **Identity**: For all $v \in V$, $Y(1, z) = \mathbb{1}_V$ and $Y(v, z)1 \in v + zV[[z]]$.
2. **Translation**: $T(1) = 0$ and, for all $v, w \in V$:

$$[T, Y(v, z)]w = \frac{\mathsf{d}}{\mathsf{d}z} Y(v, z)w \,. \tag{73.24}$$

3. **Locality**[5]: For all $v, w \in V$, there exists a $n \in \mathbb{N}$ such that:

$$(z - z')^n Y(v, z)Y(w, z') = (z - z')^n Y(w, z')Y(v, z) \,. \tag{73.25}$$

[5] Also called the **Borcherds identity**

Definition 73.4.2 (Vertex operator algebra). A VOA is a vertex algebra $(V, 1, T, Y)$ with a **conformal element** $\omega \in V$ such that:

1. $Y(\omega, z)$ corresponds to the stress-energy tensor (73.19), i.e.

$$Y(\omega, z) = \sum_{n=-\infty}^{+\infty} \omega_n z^{-n-1} = \sum_{n=-\infty}^{+\infty} L_n z^{-n-2}, \tag{73.26}$$

   with $\{L_n\}_{n\in\mathbb{Z}}$ the Virasoro generators, i.e. they satisfy Eq. (73.20).

2. L_0 is semisimple, and its eigenvalues are integral and bounded from below (the eigenvalues are called **weights**).

3. Y is homogeneous with respect to the grading induced by L_0, i.e.$\deg\big(Y(v, z)w\big) = \deg(v) + \deg(w) - n - 1$.

4. $\deg(1) = 0$ and $\deg(\omega) = 2$.

5. $L_{-1} = T$.

73.4.2 Rational CFTs

Definition 73.4.3 (Rational CFT). A (2D) CFT for which the spaces of *conformal blocks* are finite dimensional.

73.5 Extensions ♣

73.5.1 TCFT

By the *(F)FRS theorem* (see further below), rational 2D CFTs are classified by pairs of modular tensor categories (Definition 4.6.6) and internal (special symmetric) Frobenius algebras (Definition 28.1.7). By categorifying the latter, Calabi–Yau categories are obtained (Definition 4.4.18). To get a fully weak formulation, one can pass to *Calabi–Yau A_∞-categories*, where a cyclicity condition relates the higher composition maps and the trace functional. It can be shown that these classify topological conformal field theories (by a version of the *Cobordism Hypothesis* 72.1.5).

@@ ADD (F)FRS THEOREM @@

73.5.2 Strange correlator

Consider a, possibly, short-range entangled state $|\Psi\rangle$ on a space M. In general, it can be difficult to determine whether it is actually short-range entangled by looking at the bulk, since SRE states are characterized by edge modes (Section 58.1.3). The usual approach is by analyzing the entanglement spectrum of $|\Psi\rangle$. However, if the SRE is

symmetry protected and the cut made to calculate the entanglement spectrum breaks this symmetry, the spectrum of a trivial state will be obtained. Hence, a more refined approach is needed.

To this end, consider the following quantity:

$$C(p,p') := \frac{\langle\Omega|\phi(p)\phi(p')|\Psi\rangle}{\langle\Omega\,|\,\Psi\rangle}\,. \tag{73.27}$$

Here, $|\Omega\rangle$ represents a trivial, disordered state with the same global symmetries as $|\Psi\rangle$ and $\phi : M \to \mathcal{B}(\mathcal{H})$ is a local operator.

@@ COMPLETE @@

Chapter 74

Extensions and Unification

The main reference for supersymmetric quantum mechanics is the seminal paper by Witten (1982). For an introduction to algebraic superstructures, see Section 23.5.

Chapter Contents

74.1 Supersymmetry **1230**

74.1.1 Supersymmetric quantum mechanics 1230

74.1.2 Loop space mechanics 1231

74.1.3 Extensions of the Standard Model 1231

74.1.4 Spinning particle 1232

74.2 Chern–Simons theory **1233**

74.2.1 Holomorphic Chern–Simons theory 1233

74.3 Four-manifolds **1233**

74.3.1 Donaldson–Witten theory 1234

74.3.2 Instanton Floer homology 1236

74.3.3 Seiberg–Witten theory 1237

74.4 String theory **1240**

74.4.1 Worldsheets 1240

74.4.2 Branes 1241

74.4.3 Duality 1241

74.4.4 Superstrings 1241

74.5 M-theory **1242**

74.1 Supersymmetry

This section is not meant to be an in-depth study of supersymmetry and its implications for (particle) physics. It will only introduce some concepts and constructions that are widely used in the study of supersymmetric theories. It also contains some sections on certain interesting mathematical properties that arise whilst studying supersymmetry.

74.1.1 Supersymmetric quantum mechanics

In this section, a general graded Hilbert space $\mathcal{H}$ will be considered. Assume it comes equipped with an algebra of bounded operator $A \subset \mathcal{B}(\mathcal{H})$ together with a set of $N \in \mathbb{N}$ odd, self-adjoint operators $\{\mathsf{D}^i\}_{i\leq N}$. More precisely, a spectral triple $(\mathcal{H}, A, \{\mathsf{D}^i\}_{i\leq N})$ is considered (Definition 42.3.1).

This data defines an SQM system if the Hamiltonian H satisfies the following condition:

$$\{\mathsf{D}^i, \mathsf{D}^j\}_+ = 2\delta^{ij} H\,. \tag{74.1}$$

For $N = 2$, the whole theory can be rephrased in terms of a nilpotent operator $\mathsf{d} \sim \mathsf{D}^1 + i\mathsf{D}^2$ and its adjoint:

$$\{\mathsf{d}, \mathsf{d}^\dagger\}_+ \sim H\,. \tag{74.2}$$

Example 74.1.1 (Particle on a manifold). The archetypal example of $N = 2$ systems is the situation where d is the exterior derivative on a smooth manifold M, A is the algebra of smooth functions $C^\infty(M)$ and $\mathcal{H}$ is the Hilbert space of square-integrable forms with respect to the Hodge metric

$$\langle \alpha \mid \beta \rangle = \int_M \alpha \wedge *\beta\,. \tag{74.3}$$

The above superalgebra is that of a $d = 1, N = 2$ theory, i.e. there are two generators acting on a one-dimensional manifold. However, this system can be deformed to represent a $d = 2, N = 1$ theory. Let $e^{W(t)}$ be a one-parameter subgroup of invertible operators. The W-deformed operators are defined as follows:

$$\begin{aligned} \mathsf{d}_t^W &:= e^{-W(t)} \circ \mathsf{d} \circ e^{W(t)}\,, \\ (\mathsf{d}_t^W)^\dagger &:= e^{W(t)^\dagger} \circ \mathsf{d}^\dagger \circ e^{-W(t)^\dagger}\,. \end{aligned} \tag{74.4}$$

It is not hard to see that these deformed operators preserve the superalgebra. Although many authors assume W to be a smooth function, this is not necessary. In fact, many interesting examples involve more exotic choices. For example, Schreiber (2005) considers a loop space ΩM — a theory of closed strings — where the deformation operator at a point $\gamma \in \Omega M$ is given by

$$W(\gamma)\omega := \int_\gamma B_{\mu\nu}\big(\gamma(t)\big)\, \mathrm{d}x^\mu(t) \wedge \mathrm{d}x^\nu(t) \wedge \omega(t)\, dt\,, \tag{74.5}$$

i.e. the operator takes the exterior product with a given two-form field (e.g. the *Kalb–Ramond field*) and integrates over the loop γ (after pairing with a set of vector fields). When restrictied to the class of skew-Hermitian operators, these deformations can be shown to be pure gauge.

@@ COMPLETE and CHECK WHETHER CORRECT @@

74.1.2 Loop space mechanics

An interesting setting for supersymmetric quantum mechanics is the situation mentioned at the end of the previous section, namely where the base manifold is a loop space ΩM. In this case, the tangent space $T_\gamma \Omega M$ is the space of vector fields along the path $\gamma \in \Omega M$. As such, the elements carry two indices, one with respect to a (local) frame field on M and one coming from the S^1-parametrization of loops. A holonomic basis is given by functional derivatives:

$$\partial_{\mu,\sigma} := \frac{\delta}{\delta X^\mu(\sigma)}, \tag{74.6}$$

where $X^\mu(\sigma)$ is the μth coordinate of the loop at the parameter $\sigma \in [0, 2\pi[$.

74.1.3 Extensions of the Standard Model

The following theorem states that it is impossible to combine the Poincaré group in a nontrivial way with an internal symmetry group.

Theorem 74.1.2 (Coleman–Mandula). *Consider a quantum field theory satisfying the following conditions:*

1. *There exists a mass gap.*
2. *For every mass scale M, there exist only finitely many particle species with mass $m \leq M$.*
3. *The two-point scattering amplitudes are nonvanishing for almost every energy.*
4. *The (two-point) scattering amplitudes are analytic in the particle momenta.*

If the symmetry group of the S-matrix contains a subgroup isomorphic to the Poincaré group[1]*, it can be written as the direct product of the Poincaré group and an internal gauge group.*

Now, the question arises if one can do better, i.e. is there a nontrivial way to extend this total symmetry group? A first possibility was given by conformal field theories in Chapter 73. CFTs do not admit an S-matrix and, hence, the above theorem is clearly not

[1] Technically, its universal cover should be considered.

applicable. However, a second and more intricate possibility is given by supersymmetry. Here, one does not work with an ordinary symmetry Lie algebra[2] but with a Lie superalgebra (Example 4.9.2). By allowing superspaces or, equivalently, by allowing fermionic symmetry generators, one can generalize the Coleman–Mandula theorem. The resulting generalization was proven by *Sohnius, Lopuszański* and *Haag*.

74.1.4 Spinning particle

In this section, the action of spinning particle is written down in a supersymmetric way (Berezin & Marinov, 1977), and it is explained how the existence of spin structures on a manifold (Section 36.3) is related to a global anomaly in the quantization of the theory (Witten, 1985).

Consider the following action:

$$S_{\text{spin}}[x,\psi] := \frac{1}{2}\int_M g_{\mu\nu}\left(\dot{x}^\mu\dot{x}^\nu + i\psi^\mu\nabla_{\dot{x}}\psi^\nu\right)d\tau = \frac{1}{2}\int_M g_{\mu\nu}\left(\dot{x}^\mu\dot{x}^\nu + i\psi^\mu\dot{\psi}^\nu\right)d\tau\,, \tag{74.7}$$

where ψ is a Grassmann-odd vector field (<u>not</u> a spinor field) and ∇ is the spin connection. The second term is the simplest (Grassmann-even) term that is first order in time derivatives of the Grassmann variables that can be written down.

Although this theory is physically not supersymmetric, it does possess a 'worldline supersymmetry':

$$\begin{aligned}\delta x^\mu &= i\psi^\mu\varepsilon\,,\\ \delta\psi^\mu &= \left(\dot{x}^\mu - i\omega_\kappa{}^\mu{}_\nu\psi^\kappa\psi^\nu\right)\varepsilon\,,\end{aligned} \tag{74.8}$$

where ε is a Grassmann-odd constant and ω is the (local) spin connection one-form. The associated Noether charge is

$$Q = \psi_\mu\dot{x}^\mu\,. \tag{74.9}$$

After canonical quantization, the Grassmann variables ψ^μ satisfy:

$$\{\psi_\mu,\psi_\nu\}_+ = g_{\mu\nu}\,, \tag{74.10}$$

i.e. they can be interpreted as dynamical gamma matrices (68.1). This gives a physical explanation of why Clifford algebras are said to be 'quantizations' of Grassmann algebras.

The conjugate momentum of x^μ is given by

$$p_\mu = g_{\mu\nu}\dot{x}^\nu + \frac{i}{4}\omega_{\mu\nu\kappa}[\psi^\nu,\psi^\kappa]\,. \tag{74.11}$$

[2]See the original paper Coleman and Mandula (1967) for why exactly the algebra plays an essential role.

In the standard representations of quantum mechanics, the relation $p_\mu = -i\partial_\mu$ holds, so the above relation can be rewritten as

$$g_{\mu\nu}\dot{x}^\nu = -i\nabla_\mu = \partial_\mu + \frac{1}{4}\omega_{\mu\nu\kappa}[\psi^\nu, \psi^\kappa]\,, \tag{74.12}$$

i.e. as the spinorial covariant derivative (up to a factor $-i$). The supersymmetry Noether charge is thus given by the Dirac operator (Definition 36.3.19).

@@ FINISH (not clear right now) @@

74.2 Chern–Simons theory

74.2.1 Holomorphic Chern–Simons theory

Consider a 3-dimensional complex manifold equipped with a holomorphic volume form Vol, i.e. a Calabi–Yau three-fold (Definition 39.4.11). For every complex vector bundle $E \to M$ with connection ∇, one can consider the antiholomorphic connection one-form B. This form induces the following holomorphic Chern–Simons action (recall Section 69.5):

$$S_{\mathrm{CS}}[B] := \int_M \left(\langle B, \bar{\partial} B\rangle + \frac{2}{3}\langle B, [B \wedge B]\rangle \right) \wedge \mathsf{Vol}_M\,. \tag{74.13}$$

The critical points of this action are the holomorphically flat connections, i.e. the connections that satisfy $F^{0,2} = 0$. By the Koszul–Malgrange theorem 39.3.1, these correspond exactly to the holomorphic structures on E.

74.3 Four-manifolds

In this section, the base manifold M is assumed to be closed, orientable and Riemannian. The main part will be about the (differential) topology of four-manifolds.

First of all, one can define an **intersection form** $Q : H^2(M;\mathbb{Z}) \times H^2(M;\mathbb{Z})$ by pairing the cup product of two cohomology classes with the fundamental form $[M]$. In de Rham cohomology, this corresponds to integration:

$$Q\big([\omega],[\nu]\big) := \int_M \omega \wedge \nu\,. \tag{74.14}$$

By Poincaré duality, Q can also be extended to homology. It is a quadratic form and the signature (q_+, q_-) of Q is called the **signature** $\sigma(M)$ of M. (Some authors define the signature of M as $q_+ - q_-$.) The rank of Q is equal to the second Betti number of

M: $\mathsf{rk}(Q) = q_+ + q_- = b_2(M)$. An important result by *Freedman* states that two simply-connected, smooth[3] 4-manifolds are homeomorphic if and only if they have the same intersection form.

Because M is assumed to be Riemannian, one can also define the Hodge operator $*$. It can be shown that the spaces of self- and anti-selfdual forms correspond to the positive and negative subspaces of the intersection form. By Hodge theory, the signature (q_+, q_-) corresponds to the number of self- and anti-selfdual harmonic forms on M.

The first Pontryagin class $p_1(M) \in H^4(M;\mathbb{Z})$ induces a characteristic number (after integration) and, by the *Hirzebruch theorem*, one has

$$p_1(M) = 2\sigma(M)\,. \tag{74.15}$$

Moreover, if M is equipped with an almost complex structure, one can calculate its Chern classes. The second Chern class is just the Euler class $c_2(M) = e(M) \in H^4(M;\mathbb{Z})$. The first Chern number admits an expression similar to that of the Pontryagin number by Definition 35.6.17:

$$c_1^2(M) = 2\chi(M) + 3\sigma(M)\,. \tag{74.16}$$

74.3.1 Donaldson–Witten theory

In four dimensions, the classification of manifolds is considerably more intricate than in higher dimensions (cf. Remark 31.1.9). To find invariants, *Donaldson* studied instanton configurations of non-Abelian gauge theories on 4-manifolds.

Recall the physics definition of a Yang–Mills instanton. It is a connection (one-form) A with self-dual curvature vanishing at infinity. The latter means that it extends to the one-point compactification of M. Consider for example the case of a $\mathsf{SU}(n)$-bundle $P \to M$. In this case, the second Chern number reads (by Chern–Weil theory)

$$c_2(P)[M] = \int_M \frac{\mathsf{tr}(F^2) - \mathsf{tr}(F)^2}{8\pi^2}\,. \tag{74.17}$$

Because elements of $\mathfrak{su}(n)$ are traceless, this reduces to

$$c_2(P)[M] = \int_M \frac{\mathsf{tr}(F^2)}{8\pi^2} = -\frac{1}{8\pi^2}\int_M \left(\|F^+\|^2 - \|F^-\|^2\right)\mathsf{Vol}_M\,, \tag{74.18}$$

where the curvature was decomposed in a selfdual and anti-selfdual part $F^\pm$ and the

[3] For topological manifolds and odd intersection forms, two classes exist.

Killing metric $K(X, Y) = -\operatorname{tr}(XY)$ is used. Now, consider the Yang–Mills action

$$\begin{aligned} S_{\mathrm{YM}}[A] &= -\int_M \operatorname{tr}(F \wedge *F) \\ &= \int_M \|F\|^2 \operatorname{Vol}_M \\ &= \int_M \left(\|F^+\|^2 + \|F^-\|^2 \right) \operatorname{Vol}_M \,. \end{aligned} \tag{74.19}$$

The action is bounded below by (the absolute value of) the second Chern number. Depending on the sign of the second Chern number, the following minimization conditions are found:[4]

- $c_2(P)[M] \geq 0$: anti-selfdual connections, i.e. $F^+ = 0$, minimize the action.
- $c_2(P)[M] \leq 0$: selfdual connections, i.e. $F^- = 0$, minimize the action.

When $c_2(P)[M] = 1$, the (anti-)selfdual connections are called **(Yang–Mills) instantons**. In general, when the solution is an absolute minimum, i.e. when the action equals the second Chern number, the solution is called a **Bogomol'nyi–Prasad–Sommerfield instanton** (BPS). @@ CHECK THIS @@

Property 74.3.1. The intersection form Q of M is positive definite if and only if there exist no anti-selfdual harmonic forms on M.

Now, consider the moduli space of (anti-)selfdual connections:

$$\mathcal{M}^{\pm}(P) := \{A \in \mathcal{A}(P) \mid *F_A = \pm F_A\} \,/\, \operatorname{Aut}_V(P) \,. \tag{74.20}$$

First of all, one should pay attention when forming this quotient. In general, the action of the automorphism group will not be free, i.e. there will be connections with a nontrivial stabilizer. By Property 35.4.12, the stabilizer of a connection is given by the covariantly constant sections of $\operatorname{Aut}_V(P)$. Elements for which the stabilizer is a subgroup of the center $Z(G)$ are said to be **irreducible**. In general, the reducible connections correspond to singular points of $\mathcal{M}^{\pm}$. The Lie algebra of this group is given by $\Gamma(\operatorname{ad}(P))$ and, hence, the reducible connections are those that correspond to the nonzero kernel of

$$\nabla^A : \Gamma(\operatorname{ad}(P)) \to \Omega^1(M; \operatorname{ad}(P)) \,, \tag{74.21}$$

where the codomain is exactly the model space of the affine space of connections $\mathcal{A}(P)$ and, accordingly, also the (model) tangent space $T_A\mathcal{A}(P)$.

[4]Note that, on Lorentzian manifolds, the sign of the Yang–Mills action is reversed. Some authors, consequently, also reverse the sign of the Chern class and, hence, also reverse this characterization.

Now, to characterize the tangent space $T_{[A]}\mathcal{M}^{\pm}$, one needs to find the directions in $\mathcal{M}^{\pm}$ that preserve the (anti-)selfdual property and that are not gauge orbits. The second part can be done by orthogonally decomposing the tangent space with respect to ∇^A:

$$\Omega^1(M;\mathsf{Ad}(P)) = \mathsf{im}(\nabla) \oplus \mathsf{ker}(\nabla^\dagger)\,. \tag{74.22}$$

This shows that the neighbourhood of A (consisting of irreducible connections) is modelled on $\mathsf{ker}(\nabla^\dagger)$. Moreover, by linearizing the condition for F_{A+a} to be (anti-)selfdual, where $a \in \Omega^1(M;\mathsf{ad}(P))$, one obtains the condition

$$\mathsf{pr}^{\mp}(\nabla^A a) = 0\,. \tag{74.23}$$

This operator is simply the differential of the function that maps a connection to the (anti-)selfdual part of its curvature. It is also clear that the image of ∇^A lies in the kernel of $\mathsf{pr}^{\mp}\nabla^A$ since $\nabla^2 = F$. This gives rise to the **Atiyah–Hitchin–Singer complex** or **instanton deformation complex** $C^\bullet_{\mathrm{AHS}}$:

$$0 \longrightarrow \Gamma(\mathsf{ad}(P)) \xrightarrow{\nabla^A} \Omega^1(M;\mathsf{ad}(P)) \xrightarrow{\mathsf{pr}^{\mp}\nabla^A} \Omega^{2,\pm}(M;\mathsf{ad}(P)) \longrightarrow 0\,. \tag{74.24}$$

By the above arguments, one obtains the following model:

$$T_{[A]}\mathcal{M}^{\pm} \cong H^1(C^\bullet_{\mathrm{AHS}})\,. \tag{74.25}$$

The above complex is an elliptic complex and its index, minus the Euler characteristic, can be calculated by the *Atiyah–Singer index theorem*.

@@ COMPLETE (dimension, ...) @@

74.3.2 Instanton Floer homology

In the previous sections, two important tools were introduced, that of Chern–Simons theory and that of instantons. The idea of instanton Floer theory is to consider the *Morse–Floer homology* of the Chern–Simons action functional on the moduli space $\mathcal{M}^-$ of (irreducible) anti-selfdual instantons. There are some technical subtleties involved, such as the fact that the functional might not be Morse (this can be resolved by adding a so-called *holonomy perturbation*), which will be omitted.

As in Section 34.10.3, the chain complex is generated (over $\mathbb{Z}$) by the critical points of S_{CS}, i.e. the gauge equivalence classes of flat connections. For every path of connections $A : [0,1] \to \mathcal{M}^-$, one can define the **spectral flow** $\mathsf{sf}(A)$ as the number of eigenvalues that change from negative to positive minus the number of eigenvalues that change from positive to negative along the path. This is the infinite-dimensional analogue of the difference in Morse indices.

Consider the subset of $\mathcal{M}^-(\mathbb{R}\times P)$ consisting of those connections that satisfy the finite-energy condition:

$$\|F_A\|_2^2 := \int_{\mathbb{R}\times M} \|F_A\|^2 \,\mathrm{Vol}_M < +\infty\,. \tag{74.26}$$

As for ordinary Morse homology, the flow lines admit a free $\mathbb{R}$-action, which should be quotiented out. The resulting moduli space $\overline{\mathcal{M}}(A_-, A_+)$ has the structure of a finite-dimensional (oriented) smooth manifold of dimension $\mathsf{sf}(A_-, A_+) - 1$. The boundary operator is defined as follows:

$$\partial A_- := \sum_{\substack{A_+\in\mathcal{M}^-_{\mathrm{flat}}(P)\\ \mathsf{sf}(A_-,A_+)=1}} \left|\overline{\mathcal{M}}(A_-, A_+)\right| \langle A_+\rangle\,. \tag{74.27}$$

The resulting homology theory is called **instanton Floer homology**.

@@ COMPLETE @@

74.3.3 Seiberg–Witten theory

In this section, M will denote an oriented, Riemannian 4-manifold equipped with a $\mathsf{Spin}^{\mathbb{C}}$-structure $P \to M$. It can be shown that $\mathsf{Spin}^{\mathbb{C}}(4)$ is isomorphic to $\mathsf{SU}(2)\times\mathsf{SU}(2)\times\mathsf{U}(1)/\mathbb{Z}_2$. This group admits two two-dimensional representations

$$\begin{aligned} s_+ &: \mathsf{Spin}^{\mathbb{C}}(4) \to \mathsf{U}(2) : [s^+, s^-, s] \mapsto [s_+, s]\,,\\ s_- &: \mathsf{Spin}^{\mathbb{C}}(4) \to \mathsf{U}(2) : [s^+, s^-, s] \mapsto [s_-, s]\,, \end{aligned} \tag{74.28}$$

and a one-dimensional determinant representation

$$\det : \mathsf{Spin}^{\mathbb{C}}(4) \to \mathsf{U}(1) : [s^+, s^-, s] \mapsto s^2\,. \tag{74.29}$$

These representations in turn induce associated $\mathsf{Spin}^{\mathbb{C}}(4)$-bundles: two spinor bundles $S^\pm$ and one line bundle $\det(P)$. The spinor bundles are a kind of square root of the line bundle (by representation-theoretic arguments):

$$\Lambda^2 S^\pm \cong \det(P)\,. \tag{74.30}$$

Moreover, because the representations are unitary, these bundles admit a (Hermitian) metric.

Any connection on $\det(P)$ induces a connection on an associated $\mathsf{Spin}^{\mathbb{C}}(4)$-bundle after addition by the Levi-Civita connection, since

$$\mathfrak{spin}^{\mathbb{C}}(4) \cong \mathfrak{so}(4) \oplus \mathfrak{u}(1)\,. \tag{74.31}$$

So, choose such a connection A and choose a selfdual two-form μ (this two-form acts as a perturbation parameter). The Seiberg–Witten equations associated to $\mu, \nabla := \nabla^{\mathrm{LC}} + \nabla^A$ and a spinor field $\psi \in \Gamma(S^+)$ are

$$\mathsf{D}\psi = 0\,, \tag{74.32}$$

$$F_A^+ = q(\psi) + i\mu\,, \tag{74.33}$$

where D is the Dirac operator associated to ∇, F_A is the curvature of A and $q : S^+ \to i\Lambda^2 T^*M$ is the quadratic form obtained as the adjoint of the Clifford multiplication extended to an action of $\Lambda^{2,+}_{\mathbb{C}} TM$.

Construction 74.3.2 (Squaring map). The squaring map q deserves some more attention. Through Clifford multiplication, traceless self-adjoint spinor endomorphisms are identified with (imaginary) two-forms. So, the first step of defining q is to map a spinor field to a traceless operator:

$$\psi \mapsto \psi \otimes \psi^\dagger - \frac{1}{2}|\psi|^2\,. \tag{74.34}$$

One can then apply the inverse of the Clifford map to obtain the two-form. For this reason, the second Seiberg–Witten equation is often written as[5]

$$\iota_{C\ell}(F_A^+ - i\mu) = \psi \otimes \psi^\dagger - \frac{1}{2}|\psi|^2\,. \tag{74.33b}$$

Another way is to work with the adjoint of the Clifford map (since the ∂_i are unit vectors, their image is unitary):

$$\begin{aligned}\langle \iota_{C\ell}^\dagger(\psi \otimes \psi^\dagger) \mid \partial_i \wedge \partial_j \rangle &= \langle \psi \otimes \psi^\dagger \mid \iota_{C\ell}(\partial_i \wedge \partial_j) \rangle \\ &= \langle \psi \mid \iota_{C\ell}(\partial_i \wedge \partial_j)\psi \rangle \\ &= \langle \psi \mid \gamma_i\gamma_j \cdot \psi \rangle\,,\end{aligned}$$

where it was used that $\gamma_1\gamma_2 - \gamma_2\gamma_1 = 2\gamma_1\gamma_2$, since $i \neq j$. Since i, j are arbitrary, one obtains:[6]

$$q(\psi) \sim \frac{1}{2}\sum_{i \neq j} \langle \psi \mid \gamma_i\gamma_j \cdot \psi \rangle \mathrm{d}x^i \wedge \mathrm{d}x^j\,. \tag{74.35}$$

A third approach uses the isomorphism $\mathsf{SU}(2) \cong \mathsf{Sp}(1)$ and identifies the $\mathbb{C}^2$-bundles with $\mathbb{H}$-bundles. Imaginary two-forms can then be identified with sections of an $\mathsf{Im}(\mathbb{H})$-subbundle, where $\mathsf{Im}(\mathbb{H}) := \{x \in \mathbb{H} \mid \overline{x} = -x\}$. The squaring map is induced by the following morphism:

$$q : \mathbb{H} \to \mathsf{Im}(\mathbb{H}) : x \mapsto \frac{1}{2}\overline{x}ix\,. \tag{74.36}$$

[5] Another reason is that in dimensions different from 4, the spinor endomorphisms cannot be identified with two-forms and the equations are not equivalent anymore.

[6] Note that a wide variety of signs and prefactor can be found in the literature.

Remark 74.3.3 (Hyper-Kähler manifolds). The study of Seiberg–Witten equations can be vastly generalized exactly due to the relation between $\mathsf{Spin}^{\mathbb{C}}(4)$ and $\mathbb{H}$ that was used to give an alternative definition of the squaring map.

In this context, one looks at hypercomplex or, in particular, hyper-Kähler manifolds. The circle group $\mathsf{U}(1)$ acts on $\mathbb{H}$ by scalar multiplication and preserves the hyper-Kähler structure. The associated moment map (Definition 37.5.2) is exactly the morphism inducing the squaring map:

$$\mu : \mathbb{H} \to \mathsf{Sp}(1) : x \mapsto \frac{1}{2}\bar{x}ix\,. \tag{74.37}$$

Property 74.3.4 (Weitzenböck identity). The Lichnerowicz formula 36.3.25 for the Dirac operator can easily be extended to a formula for twisted Dirac operators on a vector bundle of the form $S \otimes E$. If F is the curvature form of a connection on E, the Weitzenböck identity[7] for the twisted Dirac operator reads[8]

$$\not{D}^2 = \nabla^*\nabla + \frac{1}{4}R + \frac{1}{4}F\,. \tag{74.38}$$

In the case of a proper $\mathsf{Spin}(4)$-structure on M, the half-spinor bundles $S^{\pm}$ can be identified with the tensor product $S \otimes \det^{1/2}(P)$. The Dirac operator on these bundles then satisfy the above Weitzenböck identity. However, if only a $\mathsf{Spin}^{\mathbb{C}}(4)$-structure exists, the Dirac operator cannot be obtained as a twisted operator, although it still satisfies this identity.

Now, consider the solution space $\overline{\mathcal{M}}_{S,\psi,\mu} \subset \mathcal{A}(\det(P)) \times \Gamma(S^+)$ of solutions of the Seiberg–Witten equations (S denotes the choice of $\mathsf{Spin}^{\mathbb{C}}$-structure). As usual, this space admits some gauge symmetries. Any function $g : M \to \mathsf{U}(1)$ acts on the connection ∇^L by ordinary gauge transformations:

$$A \longrightarrow A - 2g^{-1}\mathrm{d}g \qquad\qquad \psi \longrightarrow g\psi\,. \tag{74.39}$$

The factor 2 follows from the fact that a connection one-form A on $\det(P)$ would induce a connection one-form $\frac{1}{2}A$ on $\det^{1/2}(P)$ (if the latter existed). In general, the action of the gauge group is free, except at elements with $\psi = 0$, which have a $\mathsf{U}(1)$ stabilizer. To remove these gauge symmetries, one should pass to a suitable quotient. Either by all of $C^\infty(M, \mathsf{U}(1))$ or by the subset of these functions that fix a fixed basepoint, giving respectively rise to $\mathcal{M}$ and $\mathcal{M}_0$. The topology on these spaces is induced by the C^∞-Fréchet topology.

Property 74.3.5 (Moduli space). The moduli spaces $\mathcal{M}$ and $\mathcal{M}_0$ satisfy the following properties:

[7] The term **Weitzenböck identity** is used for any relation between two symmetric/elliptic second-order partial differential operators (i.e. *generalized Laplacians*) with the same principal symbol.

[8] Again, some authors might use different conventions.

- $\mathcal{M}$ is compact.
- For almost all μ, $\mathcal{M}_0$ is a (smooth) finite-dimensional manifold equipped with a circle action. Moreover, consider the intersection form $Q : H^2(M;\mathbb{Z}) \times H^2(M;\mathbb{Z})$ defined by integrating the wedge product of closed two-forms. This form is quadratic and, hence, there exists a maximal positive subspace of $H^2(M;\mathbb{Z})$. If this subspace has nonzero dimension, $\mathcal{M}$ is also smooth and $\mathcal{M} \to \mathcal{M}_0$ is a principal U(1)-bundle.
- For almost all μ, $\mathcal{M}$ is orientable and its dimension is given by

$$\dim(\mathcal{M}) = \dim(H^1(M)) - 1 - q_+ + \frac{Q\big(c_1(\det(P)), c_1(\det(P))\big) - q_+ + q_-}{4}\,, \tag{74.40}$$

  where (q_+, q_-) is the signature of Q.

74.4 String theory

74.4.1 Worldsheets

Similar to how a relativistic particle is characterized by an action that is proportional to the arc length of its worldline as defined in Eq. (66.9), i.e.

$$S_{\text{point}} = -mc \int_\gamma ds\,, \tag{74.41}$$

the action of the relativistic string is proportional to the area of its worldsheet.

Formula 74.4.1 (Nambu–Goto action).

$$S_{\text{NG}} := -T_0 \int_\Sigma d\Sigma = -\frac{1}{2\pi\alpha'} \int_\Sigma d\Sigma\,, \tag{74.42}$$

where T_0 is the **string tension** and α' is the **slope parameter** (the latter parametrization is a remnant from the work on *Regge trajectories* in nuclear physics).

The relativistic action of a particle could be written in covariant form as a nonlinear σ-model after introducing dynamical gravity on the worldline. In the same way, the Nambu–Goto action can be rewritten as follows.

Formula 74.4.2 (Polyakov action).

$$S_{\text{Polyakov}} := \frac{T_0}{2} \int_\Sigma \sqrt{-h} h^{ab} g_{\mu\nu}(x) \partial_a x^\mu \partial_b x^\nu \, d^2\tau\,. \tag{74.43}$$

Property 74.4.3 (Conformal symmetry). The 2D σ-model defined by the Polyakov action has even more structure. It is not too hard to see that this action is conformally

invariant, i.e. it is invariant under Weyl transformations $h \longrightarrow \Omega h$ on the worldsheet. It follows that the stress-energy tensor

$$T_{ab} := \frac{-2}{\sqrt{-h}} \frac{\delta S}{\delta h_{ab}} \tag{74.44}$$

is traceless.

74.4.2 Branes

Before moving to branes in string theory, an interpretation in CFTs is given. Consider the class of 2D rational CFTs, i.e. those characterized by the *FRS theorem*. Such theories are defined by a choice of modular tensor category (Definition 4.6.6) and an internal special symmetric Frobenius algebra (Definition 28.1.7). To every boundary of a 2D cobordism, the CFT assigns a module of the Frobenius algebra. This is exactly a brane in the CFT.

@@ ADD (Chan–Paton, DBI action, ...) @@

74.4.3 Duality

The first type of duality that one can consider is so-called T-duality. In its simplest form, this relates the string spectrum on two torus bundles. Consider two spacetime manifolds that are (locally) of the form $\mathbb{R}^d \times S^1(r)$ and $\mathbb{R}^d \times S^1(1/r)$. Considering the σ-model action, it can be seen that it is invariant under the transformation $r \longrightarrow 1/r$.

More generally, consider two circle bundles $\pi : P \to M$ and $\pi' : P' \to M$ together with their B-fields, which, through differential cohomology (Section 35.7), are represented by degree-3 integral cohomology classes $\alpha, \alpha' \in H^3(M; \mathbb{Z})$. These two bundles with B-field are said to be T-dual if they satisfy the following conditions:

$$\pi_* \alpha = c_1(P') \qquad \pi'_* \alpha' = c_1(P)\,, \tag{74.45}$$

where c_1 denotes the first Chern class. If this condition is satisfied, there exists an isomorphism between twisted K-theories (Section 41.3):

$$K^\bullet_\alpha(P) \cong K^{\bullet-1}_{\alpha'}(P')\,. \tag{74.46}$$

@@ ADD (S, T, Montonen–Olive, ...) @@

74.4.4 Superstrings

@@ ADD (GSO, 5 theories, ...) @@

74.5 M-theory

@@ COMPLETE (SUGRA, heterotic, Hypothesis H) @@

Chapter 75

Astronomy & Cosmology

The content of this chapter consists of two main (related) topics. The first one will be astronomy, which can mostly be treated by classical mechanics (and some general relativity). For some background, see Chapters 52 and 67. The second part will consist of more quantum-mechanical systems such as black holes and inflation.

Chapter Contents

75.1 Stars 1243
75.2 Galaxies 1243
75.3 Black holes 1244
75.4 Inflation 1244
75.5 Tegmark multiverses 1244

75.1 Stars

@@ COURSES UGENT @@

75.2 Galaxies

@@ COURSES UGENT @@

75.3 Black holes

Formula 75.3.1 (Black hole entropy). The **Bekenstein–Hawking formula** for the entropy of a black hole states that the entropy is proportional to the surface of the horizon:

$$S_{\mathrm{BH}} = \frac{k_B A}{4\ell_p^2}, \tag{75.1}$$

where k_B is the Boltzmann constant and ℓ_p is the Planck length.

Conjecture 75.3.2 (Ryu–Takayanagi formula). Consider a spatial region $\Sigma \subset \mathrm{AdS}_4$ and let S_A denotes the (CFT) entanglement entropy of a subset $A \subset \partial\Sigma$ with its complement.

$$S_A = \frac{\mathrm{Area}(\gamma_A)}{4G}, \tag{75.2}$$

where γ_A is a surface satisfying:

1. $\partial\gamma_A = \partial A$,
2. $[\gamma_A] = [A]$ in $H_2(\Sigma)$, and
3. γ_A is extremal (in terms of area).

The surface γ_A is usually called a **minimal surface**.

@@ ADD (e.g. Thermodynamics, Hawking radiation, ...) @@

75.4 Inflation

@@ ADD @@

75.5 Tegmark multiverses

In contemporary physics, there are various notions of multiverse. *Tegmark* has introduced a hierarchy that orders some of these notions.

1. In most theories of inflation, the local (observable) universe is only a subset of the entire universe since the speed of light is finite and the expansion of space outpaced this limit. The collection of all these local universes (i.e. the initial conditions) constitutes the Level I-multiverse.
2. The foundational equations of the universe (be it Standard Model + Einstein Equations, LQG, string theory, ...) admit various solutions (i.e. physical constants). Nothing, in theory, prevents nature from realizing all these solutions (an

idea that is particularly popular in string theory) by, for example, spontaneous symmetry breaking. This leads to the Level II-multiverse.

3. In the many-worlds interpretation of QM, the idea of physically branching universes leads to an additional level of multiverses. Not only are all (or, at least, various) initial conditions (Level I) and parameter configurations (Level II) realized, so are all possible histories. This collection forms the Level III-multiverse. (Note that the only difference between Level I/II and Level III-universes is the *indexical information* about which branch someone lives on. It does not add any physical content to the multiverse.)

4. On top of these physical distinctions, one can add a further layer of abstraction. Instead of merely realizing all possible parameter configurations and histories, the Level IV-multiverse also combines all possible mathematical equations and/or structures that could describe a universe.

The Level IV-multiverse is based on the following idea.

Axiom 75.1 (Mathematical Universe Hypothesis). The universe is a mathematical object. Moreover, one can strengthen this hypothesis to the **Computable Universe Hypothesis**, which says that the universe, as a mathematical object, is defined by *computable functions* (see Section 76.1).

Part XII

Beyond Math & Physics

Table of Contents

76 Computer Science **1248**
76.1 Computability 1248
76.1.1 Turing machines 1248
76.1.2 Functions 1249
76.2 Quantum computing 1249
76.3 Complexity theory 1249

77 Finance **1251**
77.1 Securities 1251
77.2 Instruments and derivatives 1252
77.3 Markets 1254
77.3.1 Introduction 1254
77.3.2 Arbitrage 1257
77.3.3 Completeness 1258
77.4 Models 1261
77.4.1 Vašíček model 1261
77.4.2 Black–Scholes model 1261

78 Philosophy of science **1263**
78.1 Modal logic 1263
78.1.1 Temporal logic 1263
78.1.2 Possibility & necessity 1263
78.2 Theories 1263
78.2.1 Emergence 1263

79 Epilogue **1264**
79.1 Reading list 1264
79.2 Current topics 1265
79.3 Personal notes 1265

Chapter 76

Computer Science

Although slightly off-topic compared to the other chapters in this compendium, many concepts in (quantum) information theory and quantum computing (Chapter 58) are strongly related to computer science. Moreover, an interest in theoretical computer science grew even more during the PhD years.

Chapter Contents

76.1 Computability **1248**

76.1.1 Turing machines 1248

76.1.2 Functions 1249

76.2 Quantum computing **1249**

76.3 Complexity theory **1249**

76.1 Computability

76.1.1 Turing machines

Definition 76.1.1 (Turing machine). A (**deterministic**) Turing machine consists of 4 components:

- **Tape**: This consists of sequential cells, each containing one symbol from an alphabet $\mathcal{A}$ (Section 15.1).
- **Head**: This can change the symbol of the current cell and change the current cell index.
- **State register**: Stores the state of the Turing machine.

- **Table of instruction**: A finite list of instructions that, given the current state and cell symbol, consisting of three steps:
  1. Erase or modify the symbol of the current cell.
  2. Move the head one cell to the left or right.
  3. Update the current state.

If the instruction table prescribes multiple possible actions for some combinations of state and symbol, the Turing machine is said to be **nondeterministic**.

76.1.2 Functions

Definition 76.1.2 (Recursively enumerable set). A set S of natural numbers is said to be recursively (or **computably**) enumerable if there exists a partial recursive function f such that $\mathsf{dom}(f) = S$.

Definition 76.1.3 (Uniformly recursively enumerable). A sequence $(S_n)_{n\in\mathbb{N}}$ of sets of natural numbers is said to be uniformly recursively enumerable if there exists a sequence $(f_n)_{n\in\mathbb{N}}$ of uniformly partial recursive functions such that $\mathsf{dom}(f_n) = S_n$ for all $n \in \mathbb{N}$.

@@ ADD (e.g. computability, decidability) @@

76.2 Quantum computing

@@ ADD (Shor, Grover, speed ups, ...) @@

76.3 Complexity theory

Definition 76.3.1 (Complexity class). The main complexity classes, together with their (partial) order relation are shown in the top row of Fig. 76.1. These are defined as follows:

- **P**: Problems that can be solved by a deterministic Turing machine in polynomial time.
- **NP**: Problems for which the verification of their solutions lies in P.
- **PSPACE**: Problems that can be solved by a deterministic Turing machine in polynomial space.[1]

[1] The space equivalent of NP does not have to be included since PSPACE = NPSPACE by Savitch's theorem.

Some of the inclusions are straightforward, e.g. $\mathsf{P} \subseteq \mathsf{NP}$ and $\mathsf{P} \subseteq \mathsf{EXP}$. Moreover, the inclusion $\mathsf{P} \subseteq \mathsf{EXP}$ is strict. The inclusion $\mathsf{P} \subseteq \mathsf{PSPACE}$ can be understood by noting that a Turing machine cannot consume space more rapidly than the number of time steps it runs.

Figure 76.1: Complexity order.

[2]EXP should be read as a further partial order consisting of EXPTIME $\subseteq$ EXPSPACE.

Chapter 77

Finance

The content of Chapters 18 to 46 is essential for this chapter. References for this chapter are Elliott and Kopp (2005); Jeanblanc et al. (2009)

Chapter Contents

77.1 Securities **1251**

77.2 Instruments and derivatives **1252**

77.3 Markets **1254**

77.3.1 Introduction 1254

77.3.2 Arbitrage 1257

77.3.3 Completeness 1258

77.4 Models **1261**

77.4.1 Vašíček model 1261

77.4.2 Black–Scholes model 1261

77.1 Securities

Definition 77.1.1 (Security). A tradeable financial asset.

Definition 77.1.2 (Bond). A security for which the issuer (debtor) owes the holder (creditor) a debt and has to provide a cash flow to the holder/creditor.

When the average interest rate is set at 0, i.e. only the **face value** is repaid at time of maturity, the bond is called a **zero-coupon bond** (or **discount bond**).

Definition 77.1.3 (Discounting). Consider an initial asset price S_0 and let $r : \mathbb{R}^+ \to \mathbb{R}$ denote the continuous-time interest rate of the asset. The associated asset price S :

$\mathbb{R}^+ \to \mathbb{R}$ is given by

$$S(t) := S_0 \exp\left(\int_0^t r(s)\,ds\right). \tag{77.1}$$

Any asset satisfying this equation or, equivalently, satisfying the SDE

$$\mathrm{d}S = r_t S_t \mathrm{d}t\,, \tag{77.2}$$

is said to be **riskfree** or **riskless**. For discrete accruals, e.g. annual interest rates, the correct formula is

$$S_n = S_0 \prod_{i=1}^{n} (1 + r_i)\,. \tag{77.3}$$

For fixed interest rates $r \in \mathbb{R}$, this gives the classical formula

$$S_n = S_0 (1 + r)^n\,. \tag{77.4}$$

The proportionality factor is often denoted by B_t, while its (multiplicative) inverse is denoted by $D_t := B_t^{-1}$. The reason for this terminology is that D_t is a **discount factor** and B_t represents the value of a bond with initial value 1.

Definition 77.1.4 (Payoff). Consider an asset with initial price P_0 and interest rate process $(r_t)_{t\in\mathbb{R}^+}$. The payoff at time $t \in \mathbb{R}^+$, i.e. the market value at time t, is given by

$$P_t = P_0 D_t^{-1}\,. \tag{77.5}$$

Again, for fixed (discrete) interest rates, this gives the classical formula

$$P_n = P_0 (1 + r)^n\,. \tag{77.6}$$

Corollary 77.1.5 (Zero-coupon bond). Consider a zero-coupon bond with deterministic interest rate $(r_t)_{t\in\mathbb{R}^+}$ of maturity $T \in \mathbb{R}_+$. Its price at time $t \leq T$ is given by

$$P(t, T) = \exp\left(-\int_t^T r(s)\,ds\right). \tag{77.7}$$

Definition 77.1.6 (At par). A coupon bond is said to be at par if its face value is equal to its initial price. This happens whenever the coupon rate is equal to the interest rate (when expressed in terms of the coupon frequency).

77.2 Instruments and derivatives

Definition 77.2.1 (Contingent claim). A contingent claim (with time of maturity $T \in \mathbb{R}^+$) on a filtered probability space $(\Omega, \Sigma, \mathbb{F}, P)$ is an Σ-measurable function that depends

on the stochastic process $(S_t)_{t\in[0,T]}$. (Often, this function is also required to be square integrable.)

European claims only depend on the asset value S at maturity T, i.e.

$$X = f(S_T) \tag{77.8}$$

for some measurable $f : \mathbb{R}^m \to \mathbb{R}^n$, i.e. they are $\mathbb{F}_T$-measurable functions. Claims that are not European are said to be **American**.

In financial terms: *A (European) contingent claim, with maturity $T \in \mathbb{R}^+$, corresponds to a contract where the seller promises a random payoff at time T.*

Definition 77.2.2 (Forward). A forward with forward price $K \in \mathbb{R}$ for an asset $(S_t)_{t\in\mathbb{R}^+}$ at maturity $T \in \mathbb{R}^+$ has payoff

$$\Phi_T = S_T - K\,. \tag{77.9}$$

I.e. one buys one unit of the asset for a fixed price at time T.

Definition 77.2.3 (IR derivatives). Caps and floors are derivatives for which the buyer receives a payment whenever the interest rate exceeds or stays under a certain threshold.

Definition 77.2.4 (Option). A derivative that allows (but not obliges) the holder to buy/sell an asset at a specified strike price on or before a specified date. Some examples are:

- **Call option**: An option to buy an asset.
- **Put option**: An option to sell an asset.
- **Swaption** (or swap option): An option for a **swap**, i.e. the exchange of assets.

Options are very similar to *futures* and forwards[1], although the latter two oblige the holder to perform the specified transaction.

There are various families of options, depending on the specific ways the payoff is calculated. The main families are the **European** and **American options**. The former, as remarked above, can only be exercised at expiration date, while the latter can be exercised at any moment before the expiration date.

Using Eq. (77.9), the payoff for a (European) call option is

$$\Phi_T = \begin{cases} S_T - K & \text{if } S_T \geq K\,, \\ 0 & \text{otherwise}\,. \end{cases} = \max(S_T - K, 0)\,. \tag{77.10}$$

[1]The main difference between futures and forwards is that the latter are less regulated, they are **over-the-counter** (OTC) contracts.

Remark 77.2.5 (Spots). When the settlement date (also called the **spot date**) of a trade is immediate, i.e. usually two days after the trade, the contract is called a **spot** (contract).

Definition 77.2.6 (Hedge). An investment position that tries to minimize potential losses. An example would be where an asset is acquired for the value $x \in \mathbb{R}^+$. If the value would go down, this would result in a loss. However, if at the same time, the investor also buys a put option for the assets, this loss can be offset.

77.3 Markets

77.3.1 Introduction

The predictable processes (Definition 46.1.10) represent the trade strategies, i.e. the amount of assets a trader has at a point $t \in \mathbb{R}^+$ in time, and the semimartingales (Definition 46.3.24) represent asset prices (a combination of these two gives a **portfolio**). The value (or **wealth**) of a portfolio is given by the stochastic integral (Section 46.4):

$$\mathsf{Val}_t(\pi; S) := \pi \cdot S_t = \int_0^t \pi_s \, dS_s \,. \tag{77.11}$$

An investor is said to be **short** if $\pi_t^i < 0$ for some $i \leq d$ and $t \in \mathbb{R}_+$.

Definition 77.3.1 (Self-financing portfolio). A d-dimensional portfolio (S, π) is said to be self-financing if the total wealth

$$\mathsf{Val}_t(\pi; S) := \sum_{i=1}^d \pi_t^i S_t^i \tag{77.12}$$

satisfies the following condition

$$\mathsf{Val}_t(\pi; S) = V_0 + \sum_{i=1}^d \pi^i \cdot S_t^i \tag{77.13}$$

or, in differential form,

$$\mathsf{dVal}_t(\pi; S) = \sum_{i=1}^d \pi_t^i \mathsf{d}S_t^i \,. \tag{77.14}$$

Essentially, this means that both ways of writing the dot product $\pi \cdot S$, one being the inner product and the other being the stochastic integral, coincide. Practically, this means that the value is completely driven by price fluctuations and that no external funding or withdrawal is allowed.

Formula 77.3.2. For discrete-time portfolios, the self-financing condition can also be rewritten as follows:

$$\sum_{i=1}^d \pi_t^i S_t^i = \sum_{i=1}^d \pi_{t-1}^i S_t^i \tag{77.15}$$

for all $t \in \mathbb{N}_0$.

Property 77.3.3 (Self-financing discount process). If (S, π) is self-financing, so is the discounted portfolio $(D^0 S, \pi)$. Note that discounting happens with respect to the first asset in S, which is assumed to be the 'riskless' asset, i.e. have a nonnegative, deterministic interest rate.

Conversely, if

$$\mathrm{dVal}_t(\pi; S) = r_t \mathrm{Val}_t(\pi; S)\mathrm{d}t + \pi_t(\mathrm{d}S_t - r_t S_t \mathrm{d}t) \tag{77.16}$$

with $\mathrm{Val}_0(\pi; S) = \lambda$ for some $\lambda \in \mathbb{R}^+$, then $\big((1, S), (\mathrm{Val}(\pi; S) - \pi S, \pi)\big)$ is a self-financing portfolio.

Definition 77.3.4 (Replicating portfolio). Two types of replication exist:

- A portfolio (π, S) is said to be statically replicating for a set of assets T when it has the same properties (e.g. cash flow).
- A portfolio (π, S) is said to be dynamically replicating for a set of assets T when it has the same increments ('Greeks').

Definition 77.3.5 (Numéraire). A nonnegative, adapted stochastic process $(\rho_t)_{t \in \mathbb{R}^+}$. The **relative price process** of any asset $(S_t)_{t \in \mathbb{R}^+}$ with respect to ρ is defined as

$$S_t^\rho := \frac{S_t}{\rho_t}. \tag{77.17}$$

Definition 77.3.6 (Equivalent martingale measure). Consider an asset $(S_t)_{t \in \mathbb{R}^+}$ and a numéraire $(\rho_t)_{t \in \mathbb{R}^+}$ on a filtered probability space $(\Omega, \Sigma, \mathbb{F}, P)$. An equivalent martingale measure (EMM) for (S, ρ) is an equivalent measure $Q \sim P$ such that $(S_t^\rho)_{t \in \mathbb{R}^+}$ is an $(\mathbb{F}, Q)$-(local) martingale.

In case where $\rho_t = D_t$ is the discount process of some bond, the EMM is often called the **risk-neutral measure**. This measure turns the discounted process $\{D_t S_t\}_{t \in T}$ into a martingale.

Definition 77.3.7 (Equivalent separating measure). A set of admissible portfolios $\mathcal{X}$ (Definition 46.4.16) on a probability space (Ω, Σ, P) is said to satisfy the equivalent separating measure (ESM) property when there exists an equivalent measure $Q \sim P$ such that

$$\mathrm{E}_Q[S_1] \leq 0 \tag{77.18}$$

for all $S \in \mathcal{X}$.

Property 77.3.8 (Change of numéraire). Consider two numéraires $(N_t)_{t \in \mathbb{R}^+}, (N'_t)_{t \in \mathbb{R}^+}$ and assume that they admit corresponding risk-neutral measures P, Q.

$$\left.\frac{\mathrm{d}Q}{\mathrm{d}P}\right|_t = \frac{N_0}{N_t}\frac{N'_t}{N'_0}. \tag{77.19}$$

Example 77.3.9 (Spot measure). Consider the discounting factor in Definition 77.1.3. The **banking account numéraire** is given by accruing interests with a starting value of 1:

$$N_t = \exp\left(\int_0^t \rho(s)\,ds\right). \tag{77.20}$$

The associated EMM is called the spot measure.

Example 77.3.10 (Forward measure). Consider a zero-coupon bond with maturity $T \in \mathbb{R}_+$ and a risk-neutral measure μ_*. The price of the bond at time $t \leq T$ is given by

$$N_t = \mathsf{E}\left[\exp\left(-\int_t^T \rho_s\,ds\right)\middle|\mathbb{F}_t\right] = P(t,T)\,. \tag{77.21}$$

The associated EMM, after a change of numéraire, is called the T-forward measure:

$$\frac{\mathrm{d}Q_T}{\mathrm{d}P} = \frac{D_T}{\mathsf{E}_{\mu_*}[D_T]}\,. \tag{77.22}$$

If the interest rate $(\rho_t)_{t\in\mathbb{R}^+}$ is deterministic, this implies that the forward measure is simply the risk-free measure.

Now, also consider some stopping tim τ. The risky bond measure is obtained by taking the numéraire of the forward measure, but instead of taking the conditional expectation with respect to the given filtration, the expectation is taken with respect to the indicator function $\mathbb{1}_{\tau>T}$:

$$\mathcal{N}_t = \mathsf{E}\left[\exp\left(-\int_t^T \rho_s\,ds\right)\middle|\mathbb{1}_{\tau>T}\right] = \mathcal{P}(t,T)\,. \tag{77.23}$$

Example 77.3.11. Assume that the stock price $(S_t)_{t\in T}$ is a generalized Brownian motion:

$$\mathrm{d}X_t = \alpha(t)S_t\,\mathrm{d}t + \sigma(t)S_t\,\mathrm{d}W_t\,. \tag{77.24}$$

The discounted process can then be rewritten as

$$D_tS_t = \sigma(t)D_tS_t\big(\theta(t)\,\mathrm{d}t + \mathrm{d}W_t\big)\,, \tag{77.25}$$

where

$$\theta(t) := \frac{\alpha(t) - R(t)}{\sigma(t)} \tag{77.26}$$

is called the **market price of risk** or **Sharpe ratio**. By Girsanov's theorem 46.4.34, an EMM for this process is given by

$$\frac{\mathrm{d}Q}{\mathrm{d}P} = \exp\left(-\int_0^t \theta(s)\,dW_s - \frac{1}{2}\int_0^t \theta^2(s)\,ds\right). \tag{77.27}$$

Definition 77.3.12 (Superhedge). Recall Definition 77.2.6. A superhedge or (x, H)-hedge of a European claim H is an admissible strategy $(\pi_t)_{t\in\mathbb{R}^+}$ such that, at maturity $T \in \mathbb{R}^+$,

$$\mathsf{Val}_T(\pi) \geq H \tag{77.28}$$

with initial value

$$\mathsf{Val}_0(\pi) = x\,. \tag{77.29}$$

A hedge for which $\mathsf{Val}_T(\pi) \geq H$ is called a **minimal hedge**. An example would be a replicating portfolio.

For American claims, a hedge is called **minimal** if there exists a stopping time τ such that

$$\mathsf{Val}_\tau(\pi) \geq H_\tau\,. \tag{77.30}$$

77.3.2 Arbitrage

An *arbitrage* (*opportunity*), to be formally defined below, is the possibility of taking advantage of the difference in price on two markets, thereby obtaining a profit (without the need of having initial assets). To exclude arbitrage opportunities, it will be necessary to put constraints on the allowed portfolios. Consider for example a constant $\omega \in \mathbb{R}^+$. By Dudley's theorem 46.4.42, there exists a process $(\phi_t)_{t\in\mathbb{R}^+}$ such that

$$\omega = \int_0^{+\infty} \phi_s \, dW_s \tag{77.31}$$

for a standard Brownian motion $(W_t)_{t\in\mathbb{R}^+}$. However, this integral exactly represents the final value of the portfolio (W, ϕ) and, hence, an arbitrage opportunity is obtained. Examples like these are called *doubling strategies* (or *martingale strategies*). Since these lead to unbounded portfolios, an admissibility criterion is needed to exclude them. The parameter $\lambda \in \mathbb{R}^+$ in the definition of admissible processes (Definition 46.4.16) can be interpreted as a 'credit line'.

Definition 77.3.13 (Arbitrage). In the language of Chapter 46, an arbitrage (opportunity) can be characterized as follows. Consider a semimartingale $(S_t)_{t\in\mathbb{R}^+}$ representing the asset prices. An arbitrage is an admissible strategy $(H_t)_{t\in\mathbb{R}^+}$ with $H_0 = 0$ a.s. such that

$$H \cdot S_t \geq 0 \text{ a.s.} \qquad \text{and} \qquad \mathsf{Prob}\big(H \cdot S_t > 0\big) > 0\,. \tag{77.32}$$

Definition 77.3.14 (Viability). A market in which no arbitrage opportunities arise is said to be **viable**.

Remark 77.3.15 (Admissibility). One could wonder whether the requirement of being admissible is necessary. When relaxing this requirement and only asking the final payoff value $H \cdot S_T$ to be nonnegative, one obtains so-called **weak arbitrage opportunities**. It can be shown that every weak arbitrage can be modified to give an admissible strategy and, hence, an arbitrage. Restricting to admissible strategies is, accordingly, sufficient. However, if one would consider strategies with negative starting values, an additional condition to rule out arbitrage opportunities is necessary.

No-arbitrage arguments are very common in (mathematical) finance. However, for the purpose of many theorems, the preceding definition is slightly too weak.

Definition 77.3.16 (No free lunch). Consider an Émery-closed convex set $\mathcal{X}$ of semimartingales that start at 0 and are bounded from below. The most common choice here, would be the set of stochastic integrals $(H \cdot S_t)_{t \in \mathbb{R}^+}$ for some semimartingale $(S_t)_{t \in \mathbb{R}^+}$, where H ranges over the admissible strategies. $\mathcal{X}$ is called the set of 'admissible portfolios'. Denote the set of evaluations of $\mathcal{X}$ at the final time ($T = 1$ by convention) by $\mathcal{K}$. Assume, moreover, that, if the 'concatenation property'

$$Z_t := H \cdot S_t + G \cdot T_t \geq -1 \tag{77.33}$$

is satisfied for all nonnegative, bounded predictable strategies with $(GH)_t = 0$, then $(Z_t)_{t \in \mathbb{R}^+} \in \mathcal{X}$.

The **No Arbitrage** (NA) criterium can be restated as saying that $\mathcal{K} \cap L^0_{\geq 0}(\Omega, P) = \{0\}$ or, if $\mathcal{C}$ denotes the linear difference $\left(\mathcal{K} \ominus L^0_{\geq 0}(\Omega, P)\right) \cap L^\infty(\Omega, P)$, that

$$\mathcal{C} \cap L^\infty_{\geq 0}(\Omega, P) = \{0\}\,. \tag{77.34}$$

More generally, the set $\mathcal{X}$ is said to admit **No Free Lunch** (NFL) if

$$\overline{\mathcal{C}}^* \cap L^\infty_{\geq 0}(\Omega, P)\,, \tag{77.35}$$

where $\overline{\mathcal{C}}^*$ denotes the closure in the weak-$*$ topology (Definition 25.1.3). If this condition holds by passing to the norm closure (where convergence is determined by sequence instead of nets), hence in L^∞, the set $\mathcal{X}$ is said to admit **No Free Lunch with Vanishing Risk** (NFLVR).

Another notion is that of **No Unbounded Profit with Bounded Risk** (NUPBR), which says that $\mathcal{K}^1$ is bounded in $L^0(\Omega, P)$, where $\mathcal{K}^1$ is the subset of final asset prices that are generated by portfolios bounded from below by -1.

Property 77.3.17. NFLVR is equivalent to NA and NUPBR combined.

77.3.3 Completeness

Definition 77.3.18 (Completeness). A market is said to be **complete** or to be an **Arrow–Debreu market** if every (European) contingent claim admits a replicating, self-financing

portfolio. This portfolio is also called a **hedging strategy** and assets for which such a strategy exists, are said to be **hedgeable**. The hedging assets are $\big(S, (\mathsf{Val}_t(\pi; S) - \pi_t S_t, \pi)\big)$.

On a finite state space Ω, this corresponds to there being $|\Omega|$ linearly independent assets. These can be, for example, **Arrow–Debreu assets**[2], i.e. assets that pays of one unit of wealth for a given state and zero for all others.

Theorem 77.3.19 (Fundamental theorems of asset pricing: Discrete markets). *The following two statements relate EMMs and no-arbitrage criteria.*

1. *A discrete market on a discrete probability space (Ω, Σ, P) is arbitrage free if and only if there exists a risk-neutral measure Q. Moreover, Q can be taken to have a bounded Radon–Nikodym derivative with respect to P.*

2. *A market on a discrete probability space is complete and, in particular, arbitrage free if and only if there exists a unique risk-neutral measure.*

Theorem 77.3.20 (Fundamental theorems of asset pricing: Continuous markets). *The following two statements relate EMMs and no-arbitrage criteria. Consider a market consisting of asset prices $(S_t)_{t\in\mathbb{R}^+}$.*

1. *Let S be a locally bounded semimartingale. There exists a risk-neutral measure if and only if $\mathcal{X} = \{S\}$ satisfies NFLVR. Moreover, if S is adapted and càdlàg and satisfies NFLVR, then it is a semimartingale.*

2. *A market is complete and, in particular, arbitrage free if and only if there exists a unique risk-neutral measure.*

Property 77.3.21 (Arbitrage pricing[3]). Let $(N_t)_{t\in\mathbb{R}^+}$ be a numéraire as before and consider an asset price $(S_t)_{t\in\mathbb{R}^+}$ in a complete and arbitrage free market. If Q is the unique risk-neutral measure with respect to N, the (relative) price of S satisfies

$$S_s^N = \mathsf{E}_Q[S_t^N \mid \mathbb{F}_s] \tag{77.36}$$

for all $s < t$ and, hence, the relative price process $(S_t^N)_{t\in\mathbb{R}^+}$ is a Q-martingale. In particular, the price of a (bounded) contingent claim H at maturity $t \in \mathbb{R}^+$ is given by:

$$V_t^H = B_t \mathsf{E}_Q[D_T H \mid \mathbb{F}_t]\,, \tag{77.37}$$

A market is arbitrage free if and only if this formula holds for all hedgeable contingent claims. If a contingent claim H is not hedgeable, its set of **arbitrage-free prices**, i.e. all initial values that do not give rise to an arbitrage opportunity when augmenting the assets S with H, is given by the set

$$\Pi(H; S) = \{\mathsf{E}_Q[D_T H] \mid Q \in \mathcal{P}(S) \wedge \mathsf{E}_Q[H] < +\infty\}\,, \tag{77.38}$$

[2]Also called **primitive** or **pure securities**

[3]These formulas hold in any viable market. Completeness is only required to price contingent claims in terms of replicating portfolios.

where $\mathcal{P}(S)$ denotes the set of risk-free measures for S. It is not hard to show that this is either the empty set or an interval.

Corollary 77.3.22 (Put-call parity). Consider (European) put and call options on some underlying asset with costs $P \in \mathbb{R}$ and $C \in \mathbb{R}$, respectively. If $S_0 \in \mathbb{R}$ is the initial asset price, the interest rate is constant and the strike price $K \in \mathbb{R}$ and maturity time $T \in \mathbb{N}$ of the options are the same, the no-arbitrage condition implies that

$$P_t + S_t = C_t + K\beta^{-(T-t)}, \tag{77.39}$$

where $\beta \in \mathbb{R}^+$ is the compounding factor (i.e. $1 + r$ for discrete compounding and e^r for continuous compounding).[4]

Remark 77.3.23 (Equity and debt). This result can also be interpreted in terms of firm values, where it is equivalent to the *Modigliani–Miller theorem*.

Property 77.3.24 (Minimal hedge). In a complete market, the buy and sell prices of European claims coincide and, moreover, the replicating strategy gives a minimal hedge (Definition 77.3.12).

Remark 77.3.25 (Martingale representation). Recall Theorem 46.4.45. Comparing this to the definition of complete markets, it can be seen that market completeness implies that the normalized asset prices $\{S^1/S^0, \dots, S^d/S^0\}$ satisfy the predictable representation property and form a *martingale basis*:

$$H/S_t^0 = H_0 + \sum_{i=1}^{d} \pi^i \cdot \left(\frac{S^i}{S^0}\right)_t. \tag{77.40}$$

Remark 77.3.26 (Adding information). A corollary of Property 46.4.49 is that, if a market S is complete (i.e. satisfies the PRP) and $\mathbb{F}_t \neq \mathbb{G}_t \cap \mathbb{F}_T$ for some $t \leq T$ or $\mathbb{F}_T = \mathbb{G}_T$, for a refinement $\mathbb{F} \hookrightarrow \mathbb{G}$ and maturity $T \in \mathbb{R}^+$, then there are no equivalent martingale laws for S with respect to $\mathbb{G}$.

In financial terms: *Refinement of filtrations corresponds to adding information. By the aforementioned result, this means that, if one adds information without changing the payoff, arbitrage opportunities arise.*

Definition 77.3.27 (Perfect market). A market is said to be perfect if it satisfies the following two conditions:

1. **Frictionless**: Transactions and costs are negligible and and information is perfect.
2. **Liquidity**: Every asset has a price and amounts are unbounded.
3.

@@ COMPLETE (add other conditions (e.g. completeness)) @@

[4] This also holds for risk-free interest rates after replacing the factor $\beta^{-(T-t)}$ by a factor $\beta_{t,T}$ that takes into account the changing interest rate.

77.4 Models

77.4.1 Vašíček model

The *Merton model* models defaults by comparing the 'firm value' to a threshold induced by the probability of default. Assume there are $n \in \mathbb{N}$ obligors and let PD_i denote the probability of default (either modelled or known) for the ith obligor. The firm value for this obligor is given by

$$V_i = \sqrt{\rho_i} Y + \sqrt{1-\rho}\varepsilon_i\,, \tag{77.41}$$

where:

- Y is a standard normally distributed **systematic factor**, which models the common contributions of all obligors[5],
- ρ, which models the asset correlation, and
- ε_i is a standard normally distributed **idiosyncratic factor** (i.i.d. with respect to the other ε_j), which models the intrinsic contributions of obligor i.

The relevance of the distinction between systematic and idiosyncratic contributions is that, conditional on Y, the firm values of the different obligors are independent. Under the Vašíček model, the probability of default i, conditional on the systemic factor $Y = y$, is given by

$$p_i(y) = \Phi\left(\frac{\Phi^{-1}(PD_i) - \sqrt{\rho}y}{\sqrt{1-\rho}}\right). \tag{77.42}$$

In practice, one samples a standard normally distributed random variable for each obligor and compares the value to the threshold in between parentheses.

77.4.2 Black–Scholes model

In this section, all price processes will be assumed to be Itô processes (Definition 46.4.12). Moreover, the filtration on (Ω, Σ, P) will be assumed to be generated by the d-dimensional Brownian motion $(W_t)_{t\in\mathbb{R}^+}$ underlying the price process:

$$\mathrm{d}S^i = \Xi^i(S)\mathrm{d}t + \sum_{j=1}^{d} C^{ij}(S)\mathrm{d}W^j\,. \tag{77.43}$$

One component of the assets in the Black–Scholes model is given by bank account $S^1 \equiv B$, which satisfies

$$\mathrm{d}B = rB\mathrm{d}t \tag{77.44}$$

[5]For example, macro-economic variables.

and $B_0 = 1$. The (constant) parameter $r > 0$ is called the **(risk-free) interest rate**. The second component $S^2 \equiv S$ is given by a risky asset, where the drift and diffusion functions (Ξ and C) are linear in the asset price:

$$\mathrm{d}S = \mu S \mathrm{d}t + \sigma S \mathrm{d}W \,. \tag{77.45}$$

This has the form of the SDE defining the stochastic exponential (Definition 46.4.17) and, hence, Eq. (46.75) gives an explicit expression for the stock price:

$$S_t = S_0 \exp\left(\sigma W_t + \left(\mu - \frac{\sigma^2}{2}\right)t\right) . \tag{77.46}$$

Solutions of the SDE, which represents an Itô process with linear coefficients, are said to follow a **geometric Brownian motion**.

Formula 77.4.1 (Black–Scholes–Merton).

$$\frac{\partial f}{\partial t} + rS\frac{\partial f}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} = rf \,. \tag{77.47}$$

Chapter 78

Philosophy of science

Chapter Contents

78.1 Modal logic **1263**

78.1.1 Temporal logic 1263

78.1.2 Possibility & necessity 1263

78.2 Theories **1263**

78.2.1 Emergence 1263

78.1 Modal logic

78.1.1 Temporal logic

78.1.2 Possibility & necessity

78.2 Theories

78.2.1 Emergence

Chapter 79

Epilogue

79.1 Reading list

Since one of the main goals of this compendium was to keep me motivated to read papers and watch lectures about the various subjects under consideration, I also list here some of the seminal works that I hope to be able to understand in the future or, at least, be a able to skim through.

Author	Title
Arkani-Hamed–Trnka	The Amplituhedron
Brown–Henneaux	Central charges in the canonical realization of ...
Bunke	Differential Cohomology
D'Auria–Fré	Geometric supergravity in $D = 11$ and its hidden supergroup
D'Auria–Fré–Castellani	Supergravity and Superstrings: A Geometric Perspective
Friends	Master and PhD theses
Fuchs–Runkel–Schweigert	TFT construction of RCFT correlators
Grothendieck	Pursuing Stacks
Henneaux–Teitelboim	Quantization of gauge systems
Lurie	Higher Topos Theory (*book*)
Lurie–Hopkins–Freed–Teleman	Topological Quantum Field Theories from Compact Lie Groups
·	On the Classification of Topological Field Theories
Schreiber	Differential cohomology in a cohesive ∞-topos
·	Higher prequantum geometry
·	Local prequantum field theory

79.2 Current topics

Current topics that I want to read up on are:

- $\mathcal{D}$-modules,
- rational CFTs and conformal blocks,
- pAQFT,
- strange correlators,
- celestial/flatspace holography,
- soft and noninvertible symmetries,
- amplituhedron,
- do-calculus and counterfactuals,
- e-values,
- Kramers–Wanier duality,
- ...

79.3 Personal notes

Many of the chapters in this compendium grew out of my personal notes taken during university lectures. As such, it is only fair that I also list the courses and lectures. The last column in the table below also shows to what degree the notes have been integrated.

Lecturer	Course	Relevant chapters	Done
Prof. Hohloch	Capita Selecta Differentiaalmeetkunde	37	Partially
Prof. Hohloch	Integrable Hamiltonian Systems	37	Partially
Prof. Mestdagh	Differentiaalmeetkunde I	30,31,36	Done
Prof. Van Neck	Inleiding tot de theoretische fysica	52	Partially

@@ CHECK NAMES @@

Part XIII

Appendices

Table of Contents

Appendix A *G*-**Structures** **1268**

Appendix B Notes **1271**

B.1 Noether's Theorem and Gauge-Gravity Duality by S. De Haro 1271

B.1.1 Pseudotensors 1271

B.1.2 AdS-CFT duality 1273

B.2 A generalization of Noether's theorem and the information-theoretic approach to the study of symmetric dynamics by R. Spekkens 1273

List of Symbols **1275**

References **1282**

Index **1290**

Appendix A

G-Structures

In the following table, an overview of the most common G-structures on a smooth (simply-connected) manifold M^n is given.

Geometric structure	Structure group	Remarks
Orientation	$\mathsf{SL}(n,\mathbb{R})$	$\mathsf{GL}^+(n,\mathbb{R})$ is sufficient for orientability. The special linear group gives rise to a volume form.
Riemannian metric	$\mathsf{O}(n)$	
Almost-symplectic structure*	$\mathsf{Sp}(n,\mathbb{R})$	Integrability (in the form of a closed form) gives a symplectic manifold.
Almost-complex structure*	$\mathsf{GL}(k,\mathbb{C})$	Integrability (in the sense of Newlander–Nirenberg) gives a complex manifold.
Almost-Hermitian structure*	$\mathsf{U}(k)$	Integrability gives a Kähler manifold.
Calabi–Yau*	$\mathsf{SU}(k)$	
Hyper-Kähler**	$\mathsf{Sp}(k)$	Hyper-Kähler implies Calabi–Yau.
Almost quaternionic**	$(\mathsf{GL}(k,\mathbb{H})\times\mathbb{H}^\times)/\mathbb{R}^\times$	Integrability gives a quaternionic manifold. $k \geq 2$ is required because for $k = 1$ one would obtain that every orientable 4-manifold is quaternionic (amongst other things).
Quaternionic Kähler**	$(\mathsf{Sp}(k)\times\mathsf{Sp}(1))/\mathbb{Z}_2$	These manifolds are not strictly Kähler since the structure group is not a subgroup of $\mathsf{U}(2k)$.

Structures marked with $*$ require the real dimension $n = 2k$ to be even. Structures marked with $**$ require the real dimension $n = 4k$ to be a multiple of 4.

Remark. This table is strongly related to the classification of (*irreducible*, simply-connected and *nonsymmetric*) Riemannian manifolds by *Berger*. (The $\mathsf{SL}(n,\mathbb{R})$-structure is technically not part of the original classification since it is not a subgroup of $\mathsf{O}(n)$ and, hence,

the manifold is not necessarily Riemannian.) A more general classification for manifolds that are not necessarily Riemannian was initiated by *Berger* and finished by others. This extension will only be mentioned here. For references, see Rudolph and Schmidt (2017).

Since not all concepts from this classification were defined throughout the compendium they are explained here:

- **Irreducible**: A Riemannian manifold is said to be irreducible if it is not locally isomorphic to a product of Riemannian manifolds.
- **Symmetric**: A smooth manifold, locally modelled on $V \cong \mathbb{R}^n$, is said to be symmetric (or **locally symmetric**) if the curvature mapping $FM \to \Lambda^2 V^* \otimes \mathfrak{g}$ is covariantly constant.

Remark A.1.1. Although most manifolds from the above list admit an explicit definition, the quaternionic Kähler manifolds are exactly defined by their structure group/holonomy group.

It is also clear that hyper-Kähler manifolds are a specific class of quaternionic Kähler manifolds since $\mathsf{Sp}(k)$ can be embedded in $\mathsf{Sp}(k) \cdot \mathsf{Sp}(1)$. To exclude this class, one can just require the holonomy group to be all of $\mathsf{Sp}(k) \cdot \mathsf{Sp}(1)$. This is equivalent to requiring that quaternionic Kähler manifolds should have a nonvanishing scalar curvature.

This remark is related to the following property.

Property A.1.2 (Einstein). Every quaternionic Kähler manifold is also Einstein (Definition 36.2.11). The hyper-Kähler manifolds are exactly the quaternionic Kähler manifolds with vanishing scalar curvature (which is constant since the manifold is Einstein).

Appendix B

Notes

This chapter contains notes written down during lunch talks, courses and conferences.

B.1 Noether's Theorem and Gauge-Gravity Duality by S. De Haro

Date & location: October 6 2018, London
Conference: *The Philosophy and Physics of Noether's Theorems*

B.1.1 Pseudotensors

Maxwell theory has a Noether stress-energy tensor of the form

$$T^{\mu}{}_{\nu} := F^{\mu\lambda}\partial A_{\lambda} - \frac{1}{4}\delta^{\mu}_{\nu}F^{\lambda\kappa}F_{\lambda\kappa} \tag{B.1}$$

and an associated weak conservation law

$$\partial_{\mu}T^{\mu\nu} \approx 0. \tag{B.2}$$

Through Schwarz's theorem 16.5.11, this tensor can be enlarged to a conserved *Belinfante tensor*

$$\overline{T}^{\mu}{}_{\nu} := T^{\mu}{}_{\nu} + \partial_{\lambda}U^{[\mu\lambda]}{}_{\nu}\,. \tag{B.3}$$

Using such an extension, the standard form

$$T^{\mu}{}_{\nu} := F^{\mu\lambda}F_{\nu\lambda} - \frac{1}{4}\delta^{\mu}_{\nu}F^{\lambda\kappa}F_{\lambda\kappa} + \text{eom} \tag{B.4}$$

can be obtained.

Now, consider the theory coupled to gravity. In this case partial derivatives have to be replaced by covariant derivatives and the (weak) conservation law becomes:

$$\nabla_\mu T^{\mu\nu} \,. \tag{B.5}$$

However, this identity only contains the stress-energy tensor of matter, not of gravity itself. A possible solution, due to *Einstein*, was to construct a stress-energy (pseudo)tensor[1] from the Christoffel symbols, since these are already responsible for the coupling to gravity in the conservation law above:

$$\partial_\mu(\sqrt{g}T^\mu{}_\nu + t^\mu{}_\nu) = 0 \,. \tag{B.6}$$

As above this leads to a **superpotential**:

$$\sqrt{g}T^\mu{}_\nu + t^\mu{}_\nu = \partial_\lambda s^{\mu\lambda}{}_\nu \,, \tag{B.7}$$

where $\partial_\mu\partial_\lambda s^{\mu\lambda}{}_\nu = 0$. All conservation laws will be preserved if $s^{\mu\lambda}{}_\nu = s^{[\mu\lambda]}{}_\nu$.

The main issues with this new 'stress-energy tensor' are:

- it is not a tensor, and
- there are an infinite number of possibilities.

However, the superpotential can be related to boundary conditions and, therefore, has a physical intepretation. Consider the Hamiltonian

$$\overline{H}(n) := \int_\Sigma n^\mu H_\mu + \oint_{\partial\Sigma} B(n) \,, \tag{B.8}$$

where n is the ADM-like shift vector that generates tangential motion along the spacelike hypersurface Σ and $B(n) \sim n^\nu s^{\mu\lambda}{}_\nu$. Noether's theorem implies that H_μ is proportional to the EOM, which implies that $H(n)$ is a boundary term determined by the superpotential. It determines the quasilocal energy.

A different approach is the **Brown–York pseudotensor**. Here, a bounded spacetime region M is considered with two spacelike boundaries, the initial and final slices $\Sigma_\pm$, and a timelike hypersurface N. The matter-coupled gravitational action is given by:

$$S = \frac{1}{2\kappa}\int_M d^4x\sqrt{-g}R + \frac{1}{\kappa}\int_{\Sigma_\pm} d^3x\sqrt{h}K - \frac{1}{\kappa}\int_N d^3x\sqrt{-\gamma}\Theta + S_{\text{matter}} + S_{\text{ref}} \,, \tag{B.9}$$

where S_{ref} is a reference action to regularize the action. Variation with respect to the metric gives

$$\delta S = -\int_M d^4x E^{\mu\nu}\delta g_{\mu\nu} + \frac{1}{2}\int_{\Sigma_\pm} d^3x\sqrt{h}P_{ij}\delta h_{ij} + \frac{1}{2}\int_N d^3x\sqrt{\gamma}\tau^{ij}_{\text{BY}}\delta\gamma_{ij} + \text{matter terms} \,. \tag{B.10}$$

[1] Pseudotensor here just means an object that does not transform tensorially.

The first term gives the Einstein field equations and vanishes on-shell. The other terms define the conjugate momenta. As the notation implied, the object τ^{ij}_{BY} is the (quasilocal) Brown–York stress-energy tensor. For pure gravity this can be written as follows:

$$\tau^{ij}_{\mathrm{BY}} = -\frac{1}{\kappa\sqrt{\gamma}}(\Theta\gamma^{ij} - \Theta^{ij}) + \text{regularizing terms}\,. \tag{B.11}$$

B.1.2 AdS-CFT duality

Now, consider an anti-de Sitter spacetime M. In the CFT description, the stress-energy tensor can be obtained by a functional derivative of the partition function with respect to some fixed background metric $g_{(0)}$:

$$\langle T_{ij}(x)\rangle = \frac{2}{\sqrt{g_{(0)}}}\frac{\delta W[g_{(0)}]}{\delta g^{ij}_{(0)}(x)}\,. \tag{B.12}$$

In the gravitational description the fixed metric rerepsents the asymptotic metric (up to conformal factors). A theorem by *Fefferman–Graham* says that near the boundary, the metric admits a local Poincaré form:

$$ds^2 = \frac{l^2}{r^2}(dr^2 + g_{ij}(r,x)dx^i dx^j)\,, \tag{B.13}$$

with $g_{ij}(0,x) = g_{(0)ij}(x)$. Holographic duality then shows that the qausilocal Brown–York tensor is equal to the holographic stress-energy tensor.

B.2 A generalization of Noether's theorem and the information-theoretic approach to the study of symmetric dynamics by R. Spekkens

Date & location: October 6 2018, London
Conference: *The Philosophy and Physics of Noether's Theorems*

The general idea of this talk is that (quantum) entanglement can be considered from a *resource theory* perspective. In this approach quantum entanglement is a **resource** and LOCC operations are the free operations, i.e. given sufficient entanglement they can be used to perform any quantum operation.

To characterize asymmetry measures, the following principle is adopted.

Axiom B.1 (Curie's principle). *The symmetries of the cause are to be found in the effect*, i.e. the effect is at least as symmetric as the cause.

Now, consider an **asymmetry measure**, i.e. a function $\mu : S(A) \to \mathbb{R}$ such that $\mu(\rho) \geq \mu(\Phi(\rho))$ for all symmetric quantum channels Φ (quantum channels that commute with

symmetry operations). For general dynamics (including open systems), where Noether's theorem does not apply, asymmetry measures give constraints on allowed evolutions. When restricting to closed systems, these measures turn into conserved quantities and for mixed states they are independent of Noether charges.

List of Symbols

The following abbreviations and symbols are used throughout the compendium.

Abbreviations

AIC	Akaike information criterion
ARMA	autoregressive moving-average model
BCH	Baker–Campbell–Hausdorff
BPS	Bogomol'nyi–Prasad–Sommerfield
BPST	Belavin–Polyakov–Schwarz–Tyupkin
BRST	Becchi–Rouet–Stora–Tyutin
CCR	canonical commutation relation
CDF	cumulative distribution function
CFT	conformal field theory
CIS	completely integrable system
CP	completely positive
CPTP	completely positive, trace-preserving
CR	Cauchy–Riemann
dga	differential graded algebra
dgca	differential graded-commutative algebra
EMM	equivalent martingale measure
EPR	Einstein–Podolsky–Rosen
ESM	equivalent separating measure
ETCS	Elementary Theory of the Category of Sets
FIP	finite intersection property
FWHM	full width at half maximum
GA	geometric algebra
GHZ	Greenberger–Horne–Zeilinger

GNS	Gel'fand–Naimark–Segal
HJE	Hamilton–Jacobi equation
HoTT	Homotopy Type Theory
KKT	Karush–Kuhn–Tucker
LIVF	left-invariant vector field
MCG	mapping class group
MPO	matrix-product operator
MPS	matrix-product state
MTC	modular tensor category
NDR	neighbourhood deformation retract
OPE	operator product expansion
OTC	over the counter
OZI	Okubo–Zweig–Iizuka
PAC	probably approximately correct
PDF	probability density function
PID	principal ideal domain
PL	piecewise-linear
PMF	probability mass function
POVM	positive operator-valued measure
PRP	predictable representation property
PVM	projection-valued measure
RKHS	reproducing kernel Hilbert space
SVM	support-vector machine
TDSE	time-dependent Schrödinger equation
TISE	time-independent Schrödinger equation
TQFT	topological quantum field theory
TVS	topological vector space
UFD	unique factorization domain
VC	Vapnik–Chervonenkis
VIF	variance inflation factor
VOA	vertex operator algebra
WKB	Wentzel–Kramers–Brillouin

ZFC	Zermelo–Frenkel set theory with the axiom of choice

Operations

Ad_g	adjoint representation of a Lie group G
ad_X	adjoint representation of a Lie algebra $\mathfrak{g}$
arg	argument of a complex number
$\Box$	d'Alembert operator
$\mathsf{deg}(f)$	degree of a polynomial f
e	identity element of a group
$\Gamma(E)$	set of global sections of a fibre bundle E
Im, $\mathfrak{I}$	imaginary part of a complex number
$\mathsf{Ind}_f(z)$	index of a point $z \in \mathbb{C}$ with respect to a function f
$\hookrightarrow$	injective function
$\cong$	is isomorphic to
$A \multimap B$	linear implication
$N \lhd G$	N is a normal subgroup of G
Par_t^γ	parallel transport map along a curve γ
Re, $\mathfrak{R}$	real part of a complex number
Res	residue of a complex function
$\twoheadrightarrow$	surjective function
$\{\cdot,\cdot\}$	Poisson bracket
$X \pitchfork Y$	transversally intersecting manifolds X, Y
∂X	boundary of a topological space X
$\overline{X}$	closure of a topological space X
$X^\circ, \mathring{X}$	interior of a topological space X
$\sphericalangle(\cdot,\cdot)$	angle between two vectors
$X \times Y$	cartesian product of two sets X, Y
$X + Y$	sum of two vector spaces X, Y
$X \oplus Y$	direct sum of two vector spaces X, Y
$V \otimes W$	tensor product of two vector spaces V, W
$\mathbb{1}_X$	identity morphism on an object X
$\approx$	is approximately equal to

$\hookrightarrow$	is included in
$\cong$	is isomorphic to
$\mapsto$	mapsto

Objects

Ab	category of Abelian groups
$\mathsf{Aut}(X)$	automorphism group of an object X
$\mathcal{B}_0(V,W)$	space of compact bounded operators between two Banach spaces V,W
$\mathcal{B}_1(\mathcal{H})$	space of trace-class operators on a Hilbert space
$\mathcal{B}(V,W)$	space of bounded linear maps between two vector spaces V,W
CartSp	category of Euclidean spaces and 'suitable' morphisms (e.g. linear maps, smooth maps, ...)
$C(X,Y)$	set of continuous functions between two topological spaces X,Y
S'	centralizer of a subset (of a ring)
$C_\bullet$	chain complex
$\mathbf{Ch}(\mathbf{A})$	category of chain complexes with objects in an additive category $\mathbf{A}$
$\mathbf{C}^\infty$, **SmoothSet**	category of smooth sets
$C^\infty_p(M)$	ring of smooth functions $f : M \to \mathbb{R}$ on a neighbourhood of $p \in M$
$C\ell(A,Q)$	Clifford algebra over an algebra A induced by a quadratic form Q
$C^\omega(V)$	set of all analytic functions defined on a set V
$\mathsf{Conf}(M)$	conformal group of a (pseudo-)Riemannian manifold M
$\mathbf{C}^\infty\mathbf{Ring}$, $\mathbf{C}^\infty\mathbf{Alg}$	category of smooth algebras
$S_k(\Gamma)$	space of cusp forms of weight $k \in \mathbb{R}$
Δ_X	diagonal of a set X
Diff	category of smooth manifolds
DiffSp	category of diffeological spaces and smooth maps
$\mathcal{D}_M$	sheaf of differential operators
D^n	standard n-disk
$\mathsf{dom}(f)$	domain of a function f
$\mathsf{End}(X)$	endomorphism monoid of an object X
$\mathcal{E}\mathrm{nd}$	endomorphism operad
$\mathbf{FormalCartSp}_{\mathsf{diff}}$	category of infinitesimally thickened Euclidean spaces

$\mathsf{Frac}(I)$	field of fractions of an integral domain I
$\mathfrak{F}(V)$	space of Fredholm operators on a Banach space V
$\mathbb{G}_a$	additive group (scheme)
$\mathsf{GL}(V)$	general linear group: group of automorphisms of a vector space V
$\mathsf{GL}(n, \mathfrak{K})$	general linear group: group of invertible $n \times n$-matrices over a field $\mathfrak{K}$
Grp	category of groups and group homomorphisms
Grpd	category of groupoids
$\mathsf{Hol}_p(\omega)$	holonomy group at a point p with respect to a principal connection ω
$\mathsf{Hom}_{\mathsf{C}}(V, W)$, $\mathsf{C}(V, W)$	collection of morphisms between two objects V, W in a category C
hTop	homotopy category
$I(S)$	vanishing ideal on an algebraic set S
$I(x)$	rational fractions over an integral domain I
$\mathsf{im}(f)$	image of a function f
$K^0(X)$	K-theory over a (compact Hausdorff) space X
Kan	category of Kan complexes
$K(A)$	Grothendieck completion of a monoid A
$\mathcal{K}_n(A, v)$	Krylov subspace of dimension n generated by a matrix A and a vector v
L^1	space of integrable functions
Law	category of Lawvere theories
$\mathfrak{Lie}$	category of Lie algebras
Lie	category of Lie groups
$\mathfrak{X}^L$	space of left-invariant vector fields on a Lie group
$\mathsf{llp}(I)$	set of morphisms having the left lifting property with respect to I
LX	free loop space on a topological space X
$\mathbf{Man}^p$	category of C^p-manifolds
Meas	• category of measurable spaces and measurable functions, or • category of measure spaces and measure-preserving functions
M^4	four-dimensional Minkowski space
$M_k(\Gamma)$	space of modular forms of weight $k \in \mathbb{R}$
$\mathbb{F}^X$	natural filtration of a stochastic process $(X_t)_{t \in T}$
$N\mathsf{C}$	simplicial nerve of a small category C

$\mathsf{O}(n, \mathfrak{K})$	group of $n \times n$ orthogonal matrices over a field $\mathfrak{K}$
$\mathbf{Open}(X)$	category of open subsets of a topological space X
$P(X), 2^X$	power set of a set X
$\mathsf{Pin}(V)$	pin group of the Clifford algebra $C\ell(V, Q)$
$\mathbf{Psh}(\mathbf{C}), \widehat{\mathbf{C}}$	category of presheaves on a (small) category $\mathbf{C}$
$R((x))$	ring of (formal) Laurent series in x with coefficients in R
$\mathsf{rlp}(I)$	set of morphisms having the right lifting property with respect to I
$R[[x]]$	ring of (formal) power series in x with coefficients in R
S^n	standard n-sphere
$S^n(V)$	space of symmetric rank n tensors over a vector space V
$\mathbf{Sh}(X)$	category of sheaves on a topological space X
$\mathbf{Sh}(\mathbf{C}, J)$	category of J-sheaves on a site $(\mathbf{C}, J)$
$\mathbf{\Delta}$	simplex category
$\mathsf{sing\ supp}(\phi)$	singular support of a distribution ϕ
$\mathsf{SL}_n(\mathfrak{K})$	special linear group: group of all $n \times n$-matrices with unit determinant over a field $\mathfrak{K}$
$W^{m,p}(U)$	Sobolov space in L^p of order m
$\mathbf{Span}(\mathbf{C})$	span category over a category $\mathbf{C}$
$\mathsf{Spec}(R)$	spectrum of a commutative ring R
$\mathbf{sSet}_{\text{Quillen}}$	Quillen's model structure on simplicial sets
$\mathsf{supp}(f)$	support of a function f
$\mathsf{Syl}_p(G)$	set of Sylow p-subgroups of a finite group G
$\mathsf{Sym}(X)$	symmetric group of a set X
S_n	symmetric group of degree n
$\mathsf{Sym}(X)$	symmetric group on a set X
$\mathsf{Sp}(n, \mathfrak{K})$	group of matrices preserving a canonical symplectic form over a field $\mathfrak{K}$
$\mathsf{Sp}(n)$	compact symplectic group
$\mathbb{T}^n$	standard n-torus (n-fold Cartesian product of S^1)
$T_{\leq t}$	set of all elements smaller than (or equal to) $t \in T$ for a partial order T
$\mathsf{TL}_n(\delta)$	Temperley–Lieb algebra with $n-1$ generators and parameter δ
$\mathbf{Top}$	category of topological spaces and continuous functions
$\mathbf{Topos}$	(2-)category of (elementary) topoi and geometric morphisms

$U(\mathfrak{g})$	universal enveloping algebra of a Lie algebra $\mathfrak{g}$
$\mathsf{U}(n, \mathfrak{K})$	group of $n \times n$ unitary matrices over a field $\mathfrak{K}$
$V(I)$	algebraic set corresponding to an ideal I
$\mathbf{Vect}(X)$	category of vector bundles over a manifold X
$\mathbf{Vect}_{\mathfrak{K}}$	category of vector spaces and linear maps over a field $\mathfrak{K}$
Y^X	set of functions between two sets X, Y
$\mathbb{Z}_p$	group of p-adic integers
$\emptyset$	empty set
$\pi_n(X, x_0)$	n^{th} homotopy space over X with basepoint x_0
$[a, b]$	closed interval
$]a, b[$	open interval
$\Lambda^n(V)$	space of antisymmetric rank-n tensors over a vector space V
ΩX	(based) loop space on a topological space X
$\Omega^k(M)$	$C^\infty(M)$-module of differential k-forms on a manifold M
$\rho(A)$	resolvent set of a bounded linear operator A
$\mathfrak{X}(M)$	$C^\infty(M)$-module of vector fields on a manifold M

Units

C	Coulomb
T	Tesla

References

Amari, S. (2016). *Information Geometry and Its Applications*. Springer.

Amini, N. (2014). *Infinite-Dimensional Lie Algebras.* https://people.kth.se/namini/PartIIIEssay.pdf

Anderson, I. M. (1992). *The Variational Bicomplex*.

Arnol'd, V. I. (2013). *Mathematical Methods of Classical Mechanics* (Vol. 60). Springer Science & Business Media.

Atiyah, M. F. (1988). Topological Quantum Field Theory. *Publications Mathématiques de l'IHÉS*, *68*, 175–186.

Augustin, T., Coolen, F. P., De Cooman, G., & Troffaes, M. C. (2014). *Introduction to imprecise probabilities*. John Wiley & Sons.

Baez, J. C., & Crans, A. S. (2003). Higher-dimensional Algebra VI: Lie 2-algebras.

Baez, J. C., & Dolan, J. (1995). Higher-dimensional algebra and topological quantum field theory. *Journal of Mathematical Physics*, *36*(11), 6073–6105.

Baez, J. C., & Hoffnung, A. (2011). Convenient Categories of Smooth Spaces. *Transactions of the American Mathematical Society*, *363*(11), 5789–5825.

Baez, J. C., Hoffnung, A. E., & Rogers, C. (2010). Categorified Symplectic Geometry and the Classical String. *Communications in Mathematical Physics*, *293*, 701–725.

Baez, J. C., & Lauda, A. D. (2003). Higher-dimensional Algebra V: 2-groups.

Baez, J. C., & May, P. (2009). *Towards Higher Categories* (Vol. 152). Springer.

Baez, J. C., & Schreiber, U. (2005). Higher Gauge Theory.

Baez, J. C., Segal, I. E., & Zhou, Z. (2014). *Introduction to Algebraic and Constructive Quantum Field Theory*. Princeton University Press.

Barr, M. (1991). *-Autonomous categories and linear logic. *Mathematical Structures in Computer Science*, *1*(2), 159–178.

Bates, S., & Weinstein, A. (1997). *Lectures on the Geometry of Quantization* (Vol. 8). American Mathematical Soc.

Beggs, E. J., & Majid, S. (2020). *Quantum Riemannian Geometry*. Springer.

Beiglboeck, M., Leonard, C., & Schachermayer, W. (2009). *A General Duality Theorem for the Monge–Kantorovich Transport Problem.* arXiv. https://doi.org/10.48550/arXiv.0911.4347

Belgun, F. (2024). Gauge Theory.

Berezin, F. A., & Marinov, M. S. (1977). Particle spin dynamics as the grassmann variant of classical mechanics. *Annals of Physics*, *104*(2), 336–362.

Berg, M., DeWitt-Morette, C., Gwo, S., & Kramer, E. (2001). The Pin Groups in Physics: C, P And T. *Reviews in Mathematical Physics*, *13*(08), 953–1034.

Berger, J. (2002). *An Infinitesimal Approach to Stochastic Analysis on Abstract Wiener Spaces* (Doctoral dissertation). LMU München.

Blackadar, B. (2013). *Operator Algebras: Theory of C*-Algebras and von Neumann Algebras*. Springer.

Boissonnat, J.-D., Nielsen, F., & Nock, R. (2010). Bregman Voronoi Diagrams. *Discrete & Computational Geometry*, *44*(2), 281–307.

Bott, R., & Tu, L. W. (1995). *Differential Forms in Algebraic Topology*. Springer New York.

Bransden, B. H., & Joachain, C. J. (2000). *Quantum Mechanics*. Prentice Hall.

Brunetti, R., Fredenhagen, K., Imani, P., & Rejzner, K. (2012). *The Locality Axiom in Quantum Field Theory and Tensor Products of {c^*}-algebras.* https://arxiv.org/abs/1206.5484

Brunetti, R., Fredenhagen, K., & Verch, R. (2003). The Generally Covariant Locality Principle – A New Paradigm for Local Quantum Field Theory. *Communications in Mathematical Physics*, *237*(1), 31–68.

Brylinski, J.-L. (1993). *Loop Spaces, Characteristic Classes and Geometric Quantization*. Birkhauser.

Calaque, D., & Strobl, T. (2015). *Mathematical Aspects of Quantum Field Theories*. Springer.

Camosso, S. (2021). Prequantization, Geometric Quantization, Corrected Geometric Quantization. *Journal of Applied Mathematics and Physics*, *09*(09), 2290–2320.

Capinski, M., & Kopp, P. E. (2013). *Measure, Integral and Probability*. Springer Science & Business Media.

Caramello, O. (2018). *Topos-theoretic background.* Oxford University Press.

Caramello, O. (2019). *Lectures on Topos Theory at the University of Insubria.* https://www.oliviacaramello.com/Teaching/Teaching.htm

Cardin, F. (2015). *Elementary Symplectic Topology and Mechanics*. Springer.

Cartier, P. (2001). A mad day's work: from Grothendieck to Connes and Kontsevich The evolution of concepts of space and symmetry. *Bulletin of the American Mathematical Society*, *38*(4), 389–408.

Cattaneo, A. S., & Schätz, F. (2011). Introduction to Supergeometry. *Reviews in Mathematical Physics*, *23*(06), 669–690.

Cheng, E., & Gurski, N. (2007). The Periodic Table of n-categories for Low Dimensions II: Degenerate Tricategories.

Choquet-Bruhat, Y., & DeWitt-Morette, C. (2000). *Analysis, Manifolds and Physics, Part 2*. North-Holland.

Choquet-Bruhat, Y., DeWitt-Morette, C., & Dillard-Bleick, M. (1991). *Analysis, Mani-*

folds and Physics, Part 1: Basics. North-Holland.

Chu, C., Blanchet, J., & Glynn, P. (2019). Probability functional descent: A unifying perspective on gans, variational inference, and reinforcement learning. In *Proceedings of the 36th International Conference on Machine Learning* (pp. 1213–1222). PMLR.

Coleman, S. R., & Mandula, J. E. (1967). All Possible Symmetries of the S-matrix. *Phys. Rev.*, *159*.

Compère, G. (2019). *Advanced Lectures on General Relativity* (Vol. 952). Springer.

Connes, A., & Kreimer, D. (1998). Hopf Algebras, Renormalization and Noncommutative Geometry. *Communications in Mathematical Physics*, *199*(1), 203–242.

Crans, S. E. (1998). Localizations of Transfors.

Debray, A. (2018). *Characteristic Classes.* https://web.ma.utexas.edu/users/a.debray/lecture_notes/u17_characteristic_classes.pdf

De La Hamette, A.-C., & Galley, T. D. (2020). Quantum reference frames for general symmetry groups. *Quantum*, *4*, 367.

Dirac, P. A. M. (1950). Generalized Hamiltonian Dynamics. *Canadian Journal of Mathematics*, *2*, 129–148.

Dudley, R. M., Feldman, J., & Le Cam, L. (1971). On Seminorms and Probabilities, and Abstract Wiener Spaces. *Annals of Mathematics*, *93*(2), 390–408.

Ebrahimi-Fard, K., & Kreimer, D. (2005). The Hopf algebra approach to Feynman diagram calculations. *Journal of Physics A: Mathematical and General*, *38*(50), R385.

Eisert, J., Simon, C., & Plenio, M. B. (2002). On the Quantification of Entanglement in Infinite-dimensional Quantum Systems. *Journal of Physics A: Mathematical and General*, *35*(17), 3911–3923.

Elliott, R. J., & Kopp, P. E. (2005). *Mathematics of Financial Markets*. Springer-Verlag.

Epstein, H., & Glaser, V. (1973). The role of locality in perturbation theory. *Annales de l'institut Henri Poincaré. Section A, Physique Théorique*, *19*(3), 211–295.

Etingof, P., Gelaki, S., Nikshych, D., & Ostrik, V. (2016). *Tensor Categories* (Vol. 205). American Mathematical Soc.

Fewster, C. J., & Rejzner, K. (2020). Algebraic quantum field theory: An introduction. In *Progress and Visions in Quantum Theory in View of Gravity: Bridging Foundations of Physics and Mathematics.* Springer.

Fiorenza, D. (2004). An Introduction to the Batalin-Vilkovisky Formalism.

Fjelstad, J., Fuchs, J., Runkel, I., & Schweigert, C. (2005). *TFT construction of RCFT correlators V: Proof of modular invariance and factorisation.*

Fuchs, J., Runkel, I., & Schweigert, C. (2002). TFT construction of RCFT correlators I: Partition functions. *Nuclear Physics B*, *646*(3), 353–497.

Fuchs, J., Runkel, I., & Schweigert, C. (2004a). TFT construction of RCFT correlators: III: Simple currents. *Nuclear Physics B*, *694*(3), 277–353.

Fuchs, J., Runkel, I., & Schweigert, C. (2004b). TFT construction of RCFT correlators II:

Unoriented world sheets. *Nuclear Physics B*, *678*(3), 511–637.

Fuchs, J., Runkel, I., & Schweigert, C. (2005). TFT construction of RCFT correlators IV:: Structure constants and correlation functions. *Nuclear Physics B*, *715*(3), 539–638.

Fuchs, J., Schweigert, C., Wood, S., & Yang, Y. (2023). *Algebraic structures in two-dimensional conformal field theory* (Tech. Rep. No. 941).

Fulton, W., & Harris, J. (2004). *Representation Theory: A First Course*. Springer.

Gallier, J. (2008). *Clifford Algebras, Clifford Groups, and a Generalization of the Quaternions.*

Gathmann, A. (2019). *Algebraic Geometry.* https://www.mathematik.uni-kl.de/gathmann/class/alggeom-2019/alggeom-2019.pdf

Georgiev, S. (2015). *Theory of Distributions*. Springer.

Girard, J.-Y. (1995). Linear Logic: its syntax and semantics. In *Advances in Linear Logic* (1st ed., pp. 1–42). Cambridge University Press.

Grady, D., & Pavlov, D. (2022). *The geometric cobordism hypothesis.* http://arxiv.org/abs/2111.01095

Greiner, W., & Reinhardt, J. (1996). *Field Quantization*. Springer.

Gross, M. (2011). *Tropical Geometry and Mirror Symmetry* (Vol. 114). American Mathematical Society.

Gui, B. (2023). *Lectures on Vertex Operator Algebras and Conformal Blocks.* arXiv.

Hairer, M. (2023). *An Introduction to Stochastic PDEs.* https://arxiv.org/pdf/0907.4178

Hawkins, E. (1998). *The Correspondence between Geometric Quantization and Formal Deformation Quantization.* arXiv. http://arxiv.org/abs/math/9811049

Hawkins, E. (2000). Geometric Quantization of Vector Bundles and the Correspondence with Deformation Quantization. *Communications in Mathematical Physics*, *215*(2), 409–432.

Henneaux, M., & Teitelboim, C. (1992). *Quantization of Gauge Systems*. Princeton university press.

Hilton, P. J., & Stammbach, U. (1997). *A Course in Homological Algebra*. Springer.

Hitchin, N. (1999). *Lectures on Special Lagrangian Submanifolds.* https://arxiv.org/abs/math/9907034v1

Hoffnung, A. E. (2011). Spans in 2-Categories: A Monoidal Tricategory.

Hovey, M. (2007). *Model Categories* (No. 63). American Mathematical Soc.

Huang, X. (2022). *Locally Cartesian Closed Categories.* https://arxiv.org/abs/2202.04543

Jeanblanc, M., Yor, M., & Chesney, M. (2009). *Mathematical methods for financial markets*. Springer Science & Business Media.

Jeevanjee, N. (2015). *An Introduction to Tensors and Group Theory for Physicists*. Birkhauser.

Johnstone, P. T. (2014). *Topos Theory*. Dover Publications.

Kapranov, M. M., & Voevodsky, V. A. (1994). *2-categories and Zamolodchikov Tetrahedra Equations* (Vol. 56). Amer. Math. Soc., Providence, RI.

Karatzas, I., & Shreve, S. (1991). *Brownian motion and stochastic calculus*. Springer Science & Business Media.

Karoubi, M. (1978). *K-Theory: An Introduction*. Springer.

Kelly, G. M. (1982). *Basic Concepts of Enriched Category Theory* (Vol. 64). CUP Archive.

Kissinger, A., & Uijlen, S. (2019). A categorical semantics for causal structure. *Logical Methods in Computer Science*, *15*.

Kolar, I., Michor, P. W., & Slovak, J. (1999). *Normal Operations in Differential Geometry*. Springer.

Lambrechts, P. (2013). *A Gentle Introduction on Homotopy Limits and Colimits*. https://mysite.science.uottawa.ca/pjacqmin/PrimerHolimHocolim-19sept2013.pdf

Lauda, A. D., & Pfeiffer, H. (2008). Open–closed Strings: Two-dimensional Extended TQFTs and Frobenius Algebras. *Topology and its Applications*, *155*(7), 623–666.

Lawvere, F. W. (1994). Cohesive toposes and Cantor's' lauter einsen'. *Philosophia Mathematica*, *2*(1), 5–15.

Leinster, T. (1998). Basic Bicategories.

Loregian, F. (2021). *(Co)end Calculus*. Cambridge University Press.

Losev, A. (2007). From Berezin Integral to Batalin-Vilkovisky Formalism: A Mathematical Physicist's Point of View.

Lurie, J. (2009). *On the Classification of Topological Field Theories*. http://arxiv.org/abs/0905.0465

Maclagan, D. (2008). *AARMS Tropical Geometry*.

Mac Lane, S. (2013). *Categories for the Working Mathematician* (Vol. 5). Springer.

Mac Lane, S., & Moerdijk, I. (1994). *Sheaves in Geometry and Logic: A First Introduction to Topos Theory*. Springer.

Massey, W. S. (1991). *A Basic Course in Algebraic Topology*. Springer.

Master, J. (2020). Why is Homology so Powerful?

May, P. (2005). A Note on the Splitting Principle. *Topology and Its Applications*, *153*(4), 605–609.

McDuff, D., & Salamon, D. (2017). *Introduction to Symplectic Topology*. Oxford University Press.

Merry, W. J. (2018). *Algebraic Topology*. https://www2.math.ethz.ch/will-merry/files/Merry%20-%20Algebraic%20Topology%20(2018).pdf

Misner, C. W., Thorne, K. S., & Wheeler, J. A. (2017). *Gravitation*. Princeton University Press.

Miti, A. M. (2021). *Homotopy Comomentum Maps in Multisymplectic Geometry*. https://arxiv.org/abs/2105.05645

Miwa, T., Jimbo, M., Jimbo, M., & Date, E. (2000). *Solitons: Differential Equations, Sym-*

metries and Infinite-dimensional Algebras (Vol. 135). Cambridge University Press.

Moretti, V. (2016). Mathematical Foundations of Quantum Mechanics: An Advanced Short Course. *International Journal of Geometric Methods in Modern Physics*, *13*.

Moshayedi, N. (2021). 4-Manifold Topology, Donaldson-Witten Theory, Floer Homology and Higher Gauge Theory Methods in the BV-BFV Formalism.

Mumford, D. (1999). *The Red Book of Varieties and Schemes: Includes the Michigan Lectures (1974) on Curves and Their Jacobians* (Vol. 1358). Springer.

Mustaţă, M. (2011). *Zeta functions in algebraic geometry.* https://websites.umich.edu/~mmustata/zeta_book.pdf (Lecture Notes)

Nash, C., & Sen, S. (2011). *Topology and Geometry for Physicists*. Dover Publications.

Nenna, L. (2020). *Monge and Kantorovich problems: from primal to dual* [Lecture notes]. https://lchizat.github.io/files2020ot/lecture1.pdf

Nuiten, J. (2013). *Cohomological quantization of local prequantum boundary field theory* (Unpublished master's thesis).

Palais, R. (1997). The Symmetries of Solitons. *Bulletin of the American Mathematical Society*, *34*(4), 339–403.

Penrose, R. (1967). Twistor Algebra. *Journal of Mathematical Physics*, *8*(2), 345–366.

Peskin, M. E., & Schroeder, D. V. (1995). *An Introduction to Quantum Field Theory*. Westview Press.

Petersen, P. (2006). *Riemannian Geometry*. Springer.

Rangamani, M., & Takayanagi, T. (2017). *Holographic Entanglement Entropy*. Springer.

Rejzner, K. (2016). Perturbative algebraic quantum field theory. *Math. Phys. Stud., Springer*.

Rezk, C. (2001). A Model for the Homotopy Theory of Homotopy Theory. *Transactions of the American Mathematical Society*, *353*(3), 973–1007.

Riehl, E. (2011). *Homotopy (Limits and) Colimits.* http://www.math.jhu.edu/eriehl/hocolimits.pdf

Riehl, E. (2013). Monoidal Algebraic Model Structures. *Journal of Pure and Applied Algebra*, *217*(6), 1069–1104.

Riehl, E. (2017). *Category Theory in Context*. Dover.

Riehl, E. (2019). *Homotopical Categories: From Model Categories to {$(infty, 1)$}-categories.* https://arxiv.org/abs/1904.00886

Riehl, E., & Verity, D. (2013). The Theory and Practice of Reedy Categories. *Theory and Applications of Categories*, *29*.

Robbin, J., & Salamon, D. (1993). The Maslov index for paths. *Topology*, *32*(4), 827–844.

Roberts, J. E. (1974). *Spontaneously Broken Gauge Symmetries and Superselection Rules* (Tech. Rep.).

Rovelli, C., & Vidotto, F. (2014). *Covariant Loop Quantum Gravity: An Elementary Introduction to Quantum Gravity and Spinfoam Theory*. Cambridge University Press.

Rozenblyum, N., Grantcharov, N. N., & Slipper, A. (2023). *BunG Seminar Notes I:*

Introduction to stacks and Bun G. (Tech. Rep.).

Rudolph, G., & Schmidt, M. (2017). *Differential Geometry and Mathematical Physics: Part II. Fibre Bundles, Topology and Gauge Fields*. Springer.

Salez, J. (2024). *Introduction to stochastic calculus.*

Sanders, R. (2012). *Commutative Spectral Triples & The Spectral Reconstruction Theorem* (Unpublished master's thesis). ((Unpublished master's thesis))

Sandgren, N., & Stoica, P. (2006). *On Moving Average Parameter Estimation* (Tech. Rep. No. 2006-022). Uppsala University.

Schottenloher, M. (2008). *A Mathematical Introduction to Conformal Field Theory* (Vol. 759).

Schreiber, U. (2005). *From Loop Space Mechanics to Nonabelian Strings* (PhD Thesis).

Schreiber, U. (2008). *The nLab.* https://ncatlab.org/nlab

Schuller, F. (2016). *Lectures on the Geometric Anatomy of Theoretical Physics.* https://www.youtube.com/channel/UC6SaWe7xeOp31Vo8cQG1oXw

Segal, I. (1965). Algebraic integration theory. *Bulletin of the American Mathematical Society*, *71*(3.P1), 419–489.

Selinger, P. (2008). *Lecture Notes on Lambda Calculus.*

Shafer, G., & Vovk, V. (2008). A Tutorial on Conformal Prediction. *J. Mach. Learn. Res.*, *9*, 371–421.

Sharpe, R. W. (2000). *Differential Geometry: Cartan's Generalization of Klein's Erlangen Program* (Vol. 166). Springer.

Shewchuk, J. R. (1994). *An Introduction to the Conjugate Gradient Method Without the Agonizing Pain* (Tech. Rep.). Carnegie Mellon University.

Sontz, S. B. (2015a). *Principal Bundles: The Classical Case*. Springer.

Sontz, S. B. (2015b). *Principal Bundles: The Quantum Case*. Springer.

Sorensen, D. (2017). *An Introduction to Characteristic Classes.* http://derekhsorensen.com/docs/sorensen-characteristic-classes.pdf

Stone, A. D. (2005). Einstein's unknown insight and the problem of quantizing chaos. *Physics Today*, *58*(8), 37–43.

Stroock, D. W. (2008). Abstract Wiener space, revisited. *Communications on Stochastic Analysis*, *2*(1), 145–151.

Sutherland, A. (2017). *Modular forms and L-functions.* https://math.mit.edu/classes/18.783/2017/LectureNotes25.pdf (Lecture Notes)

Takens, F. (1979). A Global Version of the Inverse Problem of the Calculus of Variations. *Journal of Differential Geometry*, *14*(4), 543–562.

The Univalent Foundations Program. (2013). *Homotopy Type Theory: Univalent Foundations of Mathematics*. Institute for Advanced Study.

Tiee, C. (2006). *Contravariance, Covariance, Densities, and all That: An Informal Discussion on Tensor Calculus.* https://ccom.ucsd.edu/ctiee/notes/tensors.pdf

Tuybens, B. (2017). *Entanglement Entropy of Gauge Theories* (PhD Thesis). Ghent Uni-

versity.

Tuynman, G. M. (2016). The metaplectic correction in geometric quantization. *Journal of Geometry and Physics*, *106*, 401–426.

Vanhove, R., Bal, M., Williamson, D. J., Bultinck, N., Haegeman, J., & Verstraete, F. (2018). Mapping Topological to Conformal Field Theories through strange Correlators. *Physical Review Letters*, *121*(17), 177203.

Van Proeyen, A., & Freedman, D. (2012). *Supergravity*. Cambridge University Press.

Vidal, G. (2003). Efficient Classical Simulation of Slightly Entangled Quantum Computations. *Phys. Rev. Lett.*, *91*(14).

Vistoli, A. (2004). *Notes on Grothendieck Topologies, Fibered Categories and Descent Theory.* https://arxiv.org/abs/math/0412512

Vovk, V., Gammerman, A., & Shafer, G. (2005). *Algorithmic Learning in a Random World*. Springer.

Weibel, C. A. (1994). *An Introduction to Homological Algebra*. Cambridge University Press.

Wigner, E. (2012). *Group theory: and its application to the quantum mechanics of atomic spectra* (Vol. 5). Elsevier.

Williamson, D. J., Bultinck, N., Mariën, M., Şahinoğlu, M. B., Haegeman, J., & Verstraete, F. (2016). Matrix Product Operators for Symmetry-protected Topological Phases: Gauging and Edge Theories. *Phys. Rev. B*, *94*(20).

Witten, E. (1982). Supersymmetry and Morse theory. *J. Diff. Geom*, *17*(4), 661–692.

Witten, E. (1985). Global Anomalies in String Theory. In *Symposium on Anomalies, Geometry, Topology.*

Witten, E. (1988). Coadjoint Orbits of the Virasoro Group. *Comm. Math. Phys.*, *114*(1), 1–53.

Yeats, K. (2017). *A Combinatorial Perspective on Quantum Field Theory* (Vol. 15). Springer.

You, Y.-Z., Bi, Z., Rasmussen, A., Slagle, K., & Xu, C. (2014). Wave Function and Strange Correlator of Short-Range Entangled States. *Physical Review Letters*, *112*(24), 247202.

Zadeh, L. A. (1965). Fuzzy Sets. *Information and Control*, *8*(3), 338–353.

Şahinoğlu, M. B., Williamson, D. J., Bultinck, N., Mariën, M., Haegeman, J., Schuch, N., & Verstraete, F. (2021). Characterizing Topological Order with Matrix Product Operators. *Annales Henri Poincaré*, *22*, 563–592.

Index

$*$-algebra, *see* algebra, involutive
G_2, 469
Δ-complex, 207
Σ-type, 332
σ
  additivity, 374
  algebra, 45, 839
  finite, 375
ζ-function, 256
∂-lemma, 748
étale
  cohomology, 257
2-
  group, 826
  groupoid, 827
  vector space, 142

A_∞-space, 116
Abel, *see also* commutativity
  identity, 412
Abelian
  category, 132
  group, 59
Abelianization, 60, 1013
absolute value, 72, 251, 502
absorption, 1049
  law, 51
abut, 163
accessible, 303
accumulation point, *see also* limit point
action, 767, 862, 979, 1000, 1218
  Einstein–Hilbert, 1151
  Holst, 1154
  Nambu–Goto, 1240
  Polyakov, 1240
  spooky, 1080
action-angle coordinates, 721, 998
acyclic
  complex, 155
  cover, 222
  morphism, 288
  object, 157
adèle, 252, 253
Adam, 197
adapted
  chart, *see* natural chart
  process, 867
additive
  category, 131
  function, 379
  functor, 131
adherent point, 174
adiabatic
  approximation, 1107
  theorem, 1107
adic, 76
adic numbers, 251
adjoint
  action, 68
  bundle, 655
  cylinder, 344
  Hermitian, 440, 446, 507

operator, 517, 759
representation, 583
adjoint functor theorem, 51, 304
adjugate matrix, 448
adjunction, 91, 126
monadic, 97
strong, 123
admissible, 604
process, 884
solution, 1075, 1077
Ado, 578
advective, 980
affine
algebra, 595
space, 232
Aharonov–Bohm effect, 1073
Aharonov–Casher effect, 1074
Akaike information criterion, 926
AKSZ model, 1031
Alexandrov
compactification, 184
topology, 1090
Alfsen–Shultz, 516
algebra, 114, 455
division, 455
enveloping, 78
Frobenius, 455
involutive, 514
Jordan, 456
of sets, 45
over a monad, 97
over an operad, 115
Poisson, 577
Rota–Baxter, 457
Temperley–Lieb, 456
algebraic, 81
closure, 82
function, 256
set, 232, 237
space, 247
stack, 829
almost everywhere, 374
alternator, 145
amalgamation, 66
Amari–Chentsov tensor, 895
Ambrose–Singer, 672
amplitude, 1045
amplituhedron, 1190
analytic
continuation, 369
function, 356, 368
anchor, 825
ancilla, 1135
angle, 442
annulus, 368
anomaly, 1102, 1166
Ansel–Stricker, 888
anti
-symmetry, 143
anti-
field, 1028
antighost number, 1014
antipode, 535
antisymmetry, 467
arbitrage, 1257
arc
connected, *see* path-connected
length, 555
Archimedean, 72
argument, 366
principle, 372
arity, 115
Arnoldi method, 936
arrow, 143
Arrow–Debreu
asset, 1252
Artin
algebra, 819
ring, 74
stack, 829

Artin–Wedderburn, 79
Arzelà–Ascoli, 230
associahedron, 149
associated
    bundle, 655
    vector bundle, 612
asymptotic
    curve, 561
    expansion, 355
Atiyah–Adams, 787
Atiyah–Guillemin–Sternberg, 724
Atiyah–Hirzebruch, 787
Atiyah–Hitchin–Singer, *see* instanton deformation complex
Atiyah–Jänich, 781
Atiyah–Segal axioms, 1206
Atkinson, 509
atlas, 247, 565, 743
atom, 40, 392
attaching space, 180
Aufhebung, 346
augmentation, 155
autocorrelation, 954
automorphism, 62, 438
    quantum, 1064
autoregressive model, 955

backshift, *see* lag
Baez–Crans 2-vector space, 142
Baker–Campbell–Hausdorff formula, 581, 1196
ball, 226, 496
Banach
    category, 527
    functor, 528
    space, 491
Banach–Alaoglu, 491
Banach–Schauder, 491
band, 503
band gap, 1129
bar construction, 285
Barbero–Immirzi constant, 1154
Bargmann, *see* Segal–Bargmann
barycenter, 201
base change, 264, 265
basic, 661
basis, 67, 78, 174, 435
    dual, 439
    Hamel, 434, 435
    Schauder, 435
Batalin–Fradkin–Vilkovisky, *see* BFV theory
Batalin–Vilkovisky
    integral, 1185
    Laplacian, 1184
    manifold, 1031
bath, 1115
Bayes, 840, 850
    classifier, 943
    risk, 943
Bayesian
    network, 928
Becchi–Rouet–Stora–Tyutin, 1017
Beck's monadicity theorem, 98
being, 346
Bekenstein–Hawking formula, 1244
belief, 969
Belinfante tensor, 1271
Bell state, 1080
Benamou–Brenier, 862
Beppo Levi, 384
Berezin
    integral, 479
Berezinian, 478
Bergner model structure, 299
Bernoulli, *see also* distribution
    number, 678
Berry
    connection, 1108
    phase, 1107
Bertrand, 985

Bessel
correction, 909, 915
function, 416, 419
inequality, 494
beta function, 359
Betti number, 203
BF theory, 1167
BFV theory, 1018
bi-
algebra, 535
holomorphic, 366
module, 78
morphism, 106
product, 130
rational, *see* rational
Bianchi identity, 632, 672, 695
bias, 914
Bichteler–Dellacherie, 883
bicrossed product, 65, 538
Bienaymé, 847
bifurcation, 721
bijection, 44
Birkhoff, *see also* ergodic, 1152
decomposition, 1195
Bishop, *see* finite set
Black–Scholes model, 1261
blade, 470, 526
Bloch
walls, 1132
blocked, 928
blow-up, 239
Boardman condition, 297
Bochner, 408, 948
integral, 395
Laplacian, 694
Bockstein homomorphism, 166
Bogoliubov
R-operation, 1195
Bohr
topos, 1090
Bohr–Sommerfeld condition, 1095, 1099
Bohr–Sommerfeld leaf, 1100
Bohr–Sommerfeld variety, 1099
Bohrification, 1090
Boltzmann
entropy, 1117
transport equation, 991
Bolzano, 352
Boman, 356
bond, 1251
number, 590
Bonferroni correction, 923
Boolean
algebra, 45, 53, 327
subtopos, 276
Booleanization, 276
Borcherds
identity, 1226
Borel
construction, 213
functional calculus, 513
measure, 375, 392
set, 173
space, 381
subalgebra, 588
subgroup, 458
transform, 358
Borel–Lebesgue, 183, 361
Born
approximation, 1110
rule, 1057, 1064
series, 1110
Born rule, 993
Born–Oppenheimer approximation, 1062
bornology, 185
Borsuk–Ulam, 195
Bost–Connes system, 257
Bott

element, 784
periodicity, 527, 783
boundary, 151, 173, 296
condition, 410
manifold, 572
bounded, 225
operator, 505
sequence, 164
space, 183
totally, 229
variation, 352
Bousfield–Kan map, 309
bra-ket notation, 493
bracket, 336, 491
Bragg's law, 1126
braiding, 120
branch, 370, 1193
Bravais lattice, 1125
Bredon, 201
Breit–Wigner, *see* Cauchy distribution
Brenier map, 861
Brillouin function, 1131
Bromwich integral, 407
Brouwer
fixed point theorem, 194
Brouwer–Heyting–Kolmogorov
interpretation, 334
Brown's representability theorem, 211
Brown–York pseudotensor, 1272
Brownian motion, 869, 879
geometric, 1262
Bruhat cell, 238
bubble, 752
bump function, 397
bundle, 274, 602, 604, 829
canonical, 613
determinant, 613
map, 604, 654, 656
Burali–Forti paradox, 49
Butcher group, 1195
C*-algebra, 515
bundle, 1100
Calabi–Yau
category, 138
conjecture, 740
manifold, 740
calculus, 541, 542
calibration, 637, 959
Calkin algebra, 509
call, 1253
Cameron–Martin, 871, 875
space, 874
canonical
class, 678
commutation relation, 1062
coordinates, 986
transformation, 988, 999
Cantor, 50
Cantor–Bernstein–Schröder, 44
cap, 1253
capacitance, 1034
capacity, 708, 967
consonant, 969
Carathéodory
criterion, 376
carrier, 97, 101
Cartan, 575, 734
(connection) form, *see*
Maurer–Cartan
connection, 687, 758
criterion, 585
distribution, 758
geometry, 687
magic formula, 622, 807
matrix, 590, 594, 595
Maurer–Cartan connection, 664
structure equation, 671, 672
subalgebra, 586
Cartan–Dieudonné, 529
Cartan–Killing, *see* Killing form, 591

Cartesian, *see also* Euclidean
closed, 122
morphism, 94
square, 106
Casimir
function, 710
invariant, 597, 1071
Casorati–Weierstrass, 370
category, 86, 320
arrow, 93
autonomous, 135, 136
balanced, 99
comma, 93
compact closed, 135
complete, 104, 106
concrete, 88
dagger, 93, 137
enriched, 124, 291
filtered, 105, 279, 303
fusion, 139
homotopical, 286
indexed, 314
linear, 132
model, *see* model category
of elements, 96, 130
rigid, 135
semicartesian, 863
skeletal, 90
slice, 93
Tannakian, 138
Cauchy
completion, 104
criterion, 229
distribution, 853
fundamental theorem, 992
integral formula, 368
integral theorem, 367
lemma, 992
momentum equation, 992
root test, *see* convergence
sequence, 229
stress tensor, 992
surface, 1150
Cauchy–Goursat, *see* Looman–Menchoff, 367
Cauchy–Riemann
conditions, 366, 736, 746
operator, 746
Cauchy–Schwarz inequality, 386, 492
causal, 1149
path, 929
causality, 864
Granger, 957
caustic, 720
Cayley, 62
Cayley–Hamilton, 449
Cech
cohomology, 221, 646
groupoid, 317
model structure, 320
nerve, 318
cell, 195
complex, 304
cellular complex, 207
center, 61
center of mass, 985
central
limit theorem, 851
moment, 844
central extension
Lie algebra, 583, 600
centralizer, 61, 71, 582
chain, 48, 207
complex, 151
homotopy, 152
rule, 391
chain rule, 393
change of base, 124
Chapman–Kolmogorov equation, 862
character, 408, 518, 1194

characteristic, 421, 762, 776
  θ, 678
  class, 675, 681
  curve, 419
  equation, 412, 449, 452
  function, 845
  polynomial, 449
  system, 629
chart, 564, 743
Chebyshev
  inequality, 844
Cheeger–Simons, *see* differential character
Chen space, 816
Chentsov, 902
Chern
  character, 677, 787
  class, 677
  root, 677
Chern–Simons
  form, 680
Chern–Simons theory, 1174, 1233
Chern–Weil, 676
Chernoff bound, 845
Chevalley–Eilenberg algebra, 598
Chevalley–Serre
  relations, 539, 591
chiral
  basis, *see* Weyl basis
  matrix, 1157
Choi, 520
Choi–Jamiołkowski, 1083
Choquet, 499
  integral, 968
Choquet–Bishop–de Leeuw, 499
Christoffel symbols, 562, 630, 664
claim, 1252
Clairaut, 356
class
  function, 484
classifying
  diagram, 302
  space, 285, 657, 780
  stack, 214
  topos, 213
Clausius
  inequality, 1115
cleavage, 95
Clebsch–Gordan coefficient, 488, 1072
Clifford
  algebra, 524
  bundle, 700
  group, 529
  module, 700, 785
closed, 276
  category, 122
  element, 151
  form, 638
  graph theorem, 507
  operator, 507
closure, 173, 507
  operator, 98
  relation, 1057
cloven, *see* cleavage
clustering, 945
clutching, 605, 606
co-
  adjoint, 584
  algebra, 533
  bordism, 572
  chain, 640
  completion, 104
  cycle, 603
  dense, 110
  differential, 690
  domain, 43
  factor, 448
  fibration, 197
  ideal, 535
  isotrope, 1013

orientable, 730
power, 126
product, 106, 332
root, 589
set, 60
variant, 466
vector, 612
Yoneda, *see* Yoneda, reduction, *see* Yoneda, reduction
cobordism
hypothesis, 1207
cocycle
Lie algebra, 583
Codazzi condition, 895
Codazzi–Mainardi equations, 562
coercive field, 1130
cofibrantly generated, 305
coherence, *see also* MacLane, 1069
maps, 120
cohesion, 809
differential, 811
super-differential, 813
cohomology, 321, 322
L^2, 640
compact support, 641
de Rham, 638
group, 487
Hochschild, 167
Lie algebra, 598
Lie group, 680
non-Abelian, 222
sheaf, 220
singular, 209
twisted, 323, 639, 647
vertical support, 642
Coleman–Mandula, 1231
collapse, 1057, 1064
collection, 44
collider, 928
collinearity, 934
commutant, 71
commutativity, 58, 59
commutator, 437, 584
subgroup, 60
compact, 105, 183, 242
operator, 508
compactification, 184
conformal, 703
competitive, 858
complement, 42, 52, 840
orthogonal, 439
representation, 483
symplectic, 711
vector space, 437, 441
complete
connection, 686
lattice, 51
market, 1258
measure space, 374
set, 494
space, 229
vector field, 615
completion
of rings, 76
profinite, 187
complex conjugate, 365
complexification, 467
complexity
class, 1249
component, 64, 503
composition, 54, 304
of relations, 42
computad, 299
concave, 858
concavity, 860
concordance, 680
concrete, *see also* category, *see also* von Neumann algebra
object, 810, 816
sheaf, 815

site, 815
conditional density, 849
conduction band, 1129
conductivity, 1129
cone, 103, 497
confidence, 920, 959, 960
conformal, 558
element, 1227
field theory, 1220
group, 703
weight, 1222
congruence, 447
subgroup, 254
conjugacy class, 61, 447
conjugate
distribution, 919
gradient, 936, 938
conjunction, 326
connected
component, 283
space, 181, 192
topos, 810
connection, 795
affine, 630, 802
conjugate, 894
Ehresmann, 660
flat, 802
form, 661
invariant, 632
Koszul, 629
principal, 660, 666, 831
pullback, 632, 662
reducible, 674
connective, 153
Connes
complex, 167
differential calculus, 796
reconstruction theorem, 799
Connes–Kreimer renormalization, 1192
conservation law, 772
conservative
force, 975, 981
vector field, 460
conserved quantity, 717
consistency, 914
contact
element, 729
form, 758
structure, 729
context, 328, 1091
continuity, 176, 227, 352, 867
absolute, 352, 383, 389
equation, 860, 862, 990
functor, 104
Hölder, 227
Lipschitz, 227
set, 380
uniform, 227
continuous, 217, 304
contra-
diction, 326
gredient, *see* dual representation
variant, 465, 612
contractible, 189, 336
contraction, 466, 737, 1084, 1179
convective, 980
convergence, 164, 176, 353
bounded, 882
comparison test, 354
dominated, 384, 886
Gauss's test, 355
in distribution, 843
in probability, 843
integral test, 354
monotone, 383
radius, 355
ratio test, 354
root test, 354
sequential, 227
ucp, 869

uniform, 353, 503
weak, 380
converse, 42
convex, 229, 362, 497, 858, 966
conjugate, 491
holomorphically, 733
convolution, 402, 535
theorem, 405
coordinate, 444, 1114
ring, 233, 237
coproduct, 333
copula, 855
core, 114
corner, 572
correlation, 846, 887, *see also* autocorrelation
cosine, 364, 551
cosmos, 123
cost, 857
cotangent, 552
Coulomb, 1033
gauge, 1042
Coulomb's law, 1034
countability axiom, 174
Cousin, 361
covariance, 846, 914
general, 1010
covariant, 612
derivative, 629, 668, 704, 1163
covariation, 878, 879
cover, 44, 269
étale, 246
good, 189, 222, 566
coverage, 269, 316
covering
number, 229
sieve, 269
space, 178
Coxeter group, 593
CP-problem, 1174
Cramer's rule, 444
Cramer–Rao bound, 916
credal set, 967
credibility, 960
credit, 1257
critical
point, 569, 648
cross
product, 468
section, 1109
crossing symmetry, 1158
cumulant, 899
cumulative distribution function, 842
Cuntz algebra, 517
cup product, 638, 783
Curie, 1131
ferromagnetic Curie temperature, 1132
principle, 1273
Curie–Weiss, 1131
curl, 460
current, 1034
presymplectic, 1001
currying, 122
curvature, 556, 632, 671, 681, 687, 795, 802
Gaussian, 561
geodesic, 559
Lie algebra, 146
line, 560
mean, 561
normal, 558
principal, 560
cusp form, 254
cut, 48
cutoff function, 406
CW
approximation theorem, 196
complex, 195
cycle, 62, 151, *see also* limit

cyclic, 984
vector, 518
cyclotron, 1040
cylinder
set, 869, 872
set measure, 872

d'Alembert
operator, 419, 1176
principle, 980
ratio test, *see* convergence
décalage, 475
Darboux, 706
derivative, 663
function, 352
theorem for differentiable functions, 352
dasein, 346
daseinization, 1092
Day convolution, 126
De Donder–Weyl theory, 1003
de Morgan
duality, 136
law, 42
de Rham, 638, 639, *see also* shape
cohomology, *see* cohomology
complex, 638, 826
de Rham–Weil, 220
decategorification, 91
deck transformation, 178, 191
Dedekind, *see* finite set
completeness, 41, 51, 503
cut, 48
domain, 81
deformation, 189, 539
deformation quantization
Fedosov, 1095
Moyal, 1095
degeneracy, 1118
degeneracy map, 282
degree, 807, 1187
functional, 766
of divergence, 1189
of factorization, 295
of field extension, 81
of freedom, 853
of map, 207, 786
of polynomial, 80
of superpartition, 55
of symmetric group, 62
of vector, 474
Dehn twist, 573
Deligne, *see also* tensor product, 134
complex, 684
Deligne–Beilinson product, 684
Deligne–Mumford
compactification, 745
stack, 829
delooping, 657
of groups, 114
of monoidal categories, 312
Delzant polytope, 724
dense, 184, 276
coding, 1080
subtopos, 276
density, 613, 635, 841, 860
operator, 1081
dependent
sum/product, 264
derivation, 567, 579
derivative, *see also* Radon–Nikodym, 401
functorial, 610
Gâteaux, 498
graded, 1212
Michal–Bastiani, 498
Nijenhuis–Lie, 625
of distributions, 399
weak, 399
derived
algebra, 582

category, 287
functor, 156, 287, 294
series, 60
subgroup, 60
descendant, 1226
descent, 270, 318
desirable, 966
determinant, 449, 468, 633, 789
line bundle, *see* bundle, determinant
deterministic, 864
DeWitt
convention, 479, 1026
dg-algebra, 153
Diaconescu, 277
diagonal, 42
functor, 103
diagram, 103, 112
diamagnetism, 1131
diameter, 225
dielectric constant, 1048
diffeology, 816
diffeomorphism, 566
difference, 954
differentiable, 541
differential, 151, 541, 578, 609, 620, 884
character, 681
cohomology, 681
crossed module, 827
operator, 651, 758, 763
differential equation
stochastic, 884
differential form, 619
Lie algebra-valued, 624
vector-valued, 623
differentiation, 355
diffiety, 775
diffraction, 1050, 1125
Dijkgraaf–Witten theory, 150
dimension
categorical, 139
cotangent, 541
mass, 1188
of foliation, 627
of matrix, 443
of variety, 233
of vector space, 436
pivotal, 137
Dini, 183
dipole, 1036
Dirac, 385
adjoint, 1157
algebra, 1155
basis, 1155
bracket, 1009
comb, 400
condition, 1095
conjecture, 1009
correspondence, 1093
delta function, 400, 408
equation, 1156
matrix, *see also* gamma matrix
measure, 392
operator, 701, 741
spinor, 531
Dirac–von Neumann postulates, 1056, 1064
direct
image, 265
limit, 53
sum, 130, 436, 611
system, 53
direct product, 466
of distributions, 402
of groups, 64
direct sum, 582
of groups, 64
directed set, 48
directional derivative, 460
Dirichlet

kernel, 403
problem, 426
series, 256
discounting, 1251
discrete
object, 810
smooth structure, 817
topology, 173
disjoint, 502, 839
disjoint union, 179
disjunction, 326
distance, 225, 385, 800
Chebyshev, 228
supremum, 228
taxicab, 228
distinguishable, 175
distribution, 397, 502, 704, 794, 843, 865
χ^2, 853
Bernoulli, 852
binomial, 852
empirical, 910
exponential, 851, 899
involutive, 801
isonormal, 794
joint, 848
Kolmogorov, 911
marginal, 848
mixture, 900
normal, 794, 850
of k-planes, 627
Pareto, 855
Poisson, 852
runs, 927
Student-t, 853
tempered, 405
tensor, 704
uniform, 850
distributivity, 52
distributor, 88
divergence, 460, 461, 693, 896, 1186, 1190
Bregman, 897
Kullback–Leibler, 854, 897
measure, 896
symmetry, 771
theorem, 463
division
ring, 78
divisor, 60
Dixmier
ideal, 513
trace, 513
Dixmier–Douady class, 681
DL, 878
do-calculus, 929
Dolbeault
cohomology, 748
operator, 736, 741
Dold–Kan correspondence, 286
domain, 43, 351
integral, 71
principal ideal, 75
unique factorization, 72
dominant, 233
Donaldson–Witten theory, 1234
Donsker, 913
Doob
decomposition, 877
martingale, 876
Doob–Meyer, 878
Doob–Ville
inequality, 876
Drinfel'd
morphism, 140
quantum double, 539
dual
connection, 660
map, 439
numbers, 78
object, 135

pair, 439
representation, 482
selfdual, 474
space, 135, 438
duality
electric-magnetic, 1040
duality gap, 941
Dudley, 890
Dugger, 307
Duitstermaat–Heckman localization, 724
Dunkl operator, 800
Durbin method, 957
dynamical
system, 717
Dynkin diagram, 590
dyon, 1173
Dyson series, 1059

Eckmann–Hilton
argument, 58, 111
duality, 194
edge, 1188
Edwards correction, 924
effective
action, *see* faithful
efficient, 916
Ehrenfest, 1061
eigen-
value, 452
vector, 452
Eilenberg–MacLane
space, 193
spectrum, 212
Eilenberg–Moore category, 97
Eilenberg–Steenrod axioms, 210
Eilenberg–Watts, 133
Eilenberg–Zilber, 295
Einstein
energy relation, 1146
field equations, 1151
lcoality, 1208
manifold, 696, 1270
tensor, 696
Einstein–Podolsky–Rosen, *see* Bell state
Eisenstein series, 255
elastic, *see* cohesion, differential
electromagnetism, 1033, 1073
elementary matrix, 450
Elitzur, 1168
elliptic, 722
complex, 651
operator, 427, 651
PDE, 421
Elmendorf, 201
embedding, 88, 178, 265, 570
theorem, *see* Freyd–Mitchell
empty, 965
end, 108
endo-
functor, 87
morphism, 437
morphism operad, 115
energy, 747, 862, 976, 979, 988, 1118, 1146
free, 899
entailment, 268
entanglement, 1080
entangling surface, 1197
entire, 366
entropy
black hole, 1244
free, 1118
metric, 229
Rényi, 1197
relative, 854
Shannon, 854
envelope, 352, 1046
environment, 1114
epimorphism, 99, 107, 316
equality, 329, 334

equalizer, 106
equation of motion
  linearised, 1214
equicontinuity, 230, 500
equilibrium, 1114
equipartition theorem, 1120
equivalence, 336
  adjoint, 92
  Dwyer–Kan, 298, 300
  Morita, 77, 527
  Murray–von Neumann, 523
  of categories, 90
  unitary, 451
equivalent
  measures, 389
equivariant
  cohomology, 212, 639
  map, 68
  sheaf, 213, 219
ergodic, 379, 516
  principle, 1117
  series, 954
  theorem, 385
Erlangen program, 685
error, 913, 922, 923, 943
  function, 851
essential
  infimum, 380
  spectrum, 510
  supremum, 380
etale
  covering, 244
  morphism, 178, 235, 243, 247
  space, 178
Euclidean space, 454
Euler
  characteristic, 204, 572, 613, 744
  class, 645, 679
  formula for normal curvature, 561
  homogeneous function theorem, 357, 807
  integral, 359
  operator, 357, 614, 763, 765
  reflection formula, 359
  sequence, 626
Euler product, 256
Euler–Lagrange
  equations, 980
  operator, 763, 768
Euler–Poincaré formula, 204
evaluation, 82, 135
evanescence, 867
event, 839
evolution, 1057, 1064
evolutionary vector field, 761
Ewald sphere, 1126
ex falso quodlibet, 327, 333
exact
  couple, 160
  element, 151
  form, 638
  functor, 104, 132
  Lagrangian, 713
  sequence, 154
exceptional
  lie algebra, 591
excess, *see* kurtosis
excision, 205, 780, 786
expectation, 844, 1057
  conditional, 847, 850
exponentiable, 828
exponential
  Doléans-Dade, 885
  map, 580
  object, 122
Ext, 158
extension, 101
  algebraic, 82
  finite, 81
  normal, 83

of an ideal, 75
radical, 84
extensionality, 39
exterior
algebra, 471, 542
covariant derivative, 630, 669
derivative, 620, 624
power, 470
product, 526
system, 775
extinction, 1127
extreme
point, 362

F_σ-set, 174
Faá di Bruno formula, 777
face, 202
map, 208, 282
value, 1251
factor, 523, *see also* Levi subalgebra
factorial, 351
factorization, 266
functorial, 93
lemma, 293
weak, 286
Faddeev–Popov
action, 1023
ghost, 1015, 1184
integral, 1183
path integral, 1184
faithful, 88
action, 69
family, 44
Fano plane, 239
Faraday, 1040
Fatou, 383
Fenchel–Moreau, 492
Fenchel–Young inequality, 491
Fermi
function, 1129
golden rule, 1109
level, 1129
fermion, 55
Ferrers diagram, *see* Young diagram
ferrimagnetism, 1133
ferromagentism, 1131
Feynman
diagram, 360, 1188
propagator, 1179, 1180
slash, 1156
fibrant object, 288, 293
fibration, 102, 196, 288, 293, 340
fibre, 43, 602
bundle, 603
category, 94
derivative, 997
intregation, 642
product, 107, 606
fibred
category, 94
manifold, 602
field, 70
extension, 81
finite, 250
global, 253
local, 253
of fractions, 71
of sets, *see* algebra
strength, 1043
theory, 1206
Fierz identity, 1160
filter, 45, 48
filtered set, *see* directed set
filtration, 45
finite, 133, 134
intersection property, 182
set, 41
type, 195, 474, 566
Fisher
information, 900, 916
Fisher–Neyman factorization theorem,

901
fixed point, 200
flag, 458, 675
flat
character, 681
connection, 671
functor, 277, 278
modality, 810
module, 158
morphism, 158, 242
topology, 246
Fletcher–Reeves formula, 938
floor, 1253
flow, 615
focal
set, 969
focus-focus, 722
foliation, 627
symplectic, 710
folium, 518
force, 975
forest, 1192
forgetful, 91
fork, 106, 928
form, 468, 470
formal
completion, 76
set, 820
formally
étale, 76, 243
smooth, 76, 243
formula, 326
forward, 1253
foton, 1199
Fourier
generalized series, 494
series, 403
transform, 404, 408
Fourier–Stieltjes transform, 406
Frölicher space, 814
Frölicher–Nijenhuis bracket, 625
Fréchet, *see also* topology
derivative, 499
manifold, 1222
space, 497
Fréchet–Hoeffding bounds, 856
Fréchet–Urysohn space, *see* sequential space, 175
fractional error, 913
frame, 435, 611
bundle, 655
set theory, 52
Fraunhofer, *see* diffraction
Fredholm
module, 521
operator, 509
free, 69, 101, 224
functor, 101
group, 66
module, 78
product, 66
Frenet formulas, 556
Frenet–Serret frame, 556
Frenkel pair, 1128
frequency, 1045
Freudenthal
spectral theorem, 503
Freudenthal suspension theorem, 194, 196
Freyd–Mitchell embedding theorem, 133
Friedrich, 428
Frobenius, 455
algebra, 138, 534, 1206
integrability theorem, 627, 802
norm, *see* Hilbert–Schmidt
reciprocity, 123
series, 413
transformation, 370
Froda, 393

Fubini, 389
Fubini–Study metric, 1064
Fuchs, 413
Fuchs–Runkel–Schweigert, 1227, 1241
fugacity, 1119
Fukaya category, 752
full, 88
    subcategory, 86
function, 329, 331
    field, 233, 241
functional, 438
    calculus, 512
    form, 765
    graded, 1211
    multilocal, 1211
functor, 87, 124
    category, 89
    finitary, 105
    homotopical, 286
    monadic, 97
fundamental
    category, 297
    class, 636
    form, 557, 558, 739
    group, 190, 202
    groupoid, 191, 809
    solution, 399
fundamental group
    etale, 244
fundamental theorem
    for curves, 556
    for line integrals, 463
    of algebra, 80
    of asset pricing, 1259
    of calculus, 357, 395, 498, 663
    of finitely generated Abelian
        groups, 67
    of Galois theory, 84
    of Riemannian geometry, 693
    of topos theory, 263
    of trigonometry, 552
fuzzy
    relation, 966
    set, 965
FWHM, 910

G_δ-set, 174
G-
    grading, 139
    structure, 673
Gabriel–Ulmer duality, 303
Galerkin condition, 937
Galilei
    group, 598
Galois, *see also* field, finite
    extension, 83
    theory, 79, 248
    topos, 187
gamble, 966
game, 967
gamma
    function, 359
    matrix, 531
Gangbo–McCann, 859
gauge, *see also* integral, Denjoy
    Coulomb, 1165
    derivative, 1025
    field, 796, 1043
    fixing, 1022, 1042, 1164, 1185
    Lorenz, 1165
    principle, 1163
    propagating, 1023
    temporal, 1165
    transformation, 604, 654, 1041, 1043
gauging, 754
Gauss
    distribution, 850, 872, 873
    equation for Riemann curvature,
        562
    formula for surfaces, 562
    integral, 360

laws for electricity and magnetism, 1040
Gauss–Ostrogradsky, 463
Gel'fand
duality, 518
representation, 518
spectrum, 518, 1091
triple, 1076
Gel'fand–Naimark, 519
Gel'fand–Naimark–Segal construction, 519
Gell-Mann–Nishijima formula, 1202
general linear group, 438, 443
generalized
coordinate, 979
force, 980
function, 397
generating
function, 712, 740
generator, 63, 100, 508
generic
point, 184
genus, 678
geodesic, 562, 694, 802
equation, 562
geometric
algebra, 526
point, 240
realization, 284
solution, 1078
gerbe, 830
germ, 175, 216, 217, 566
Gerstenhaber algebra, 144
ghost, 1014, *see* Faddeev–Popov
number, 1015–1017, 1031
Girsanov, 888
Giry monad, 862
Gleason, 793, 1090
Glivenko–Cantelli, 911
class, 912
global
element, 100
sections, 217, 272
Godement product, 311
Goldstone, 1168
boson, 1169, 1170
goodness of fit, 926
graded
algebra, 474
commutativity, 474
object, 45
vector space, 474
gradient, 460, 461
of outer product, 467
grafting, 1192
Gram matrix, 946
Gram–Schmidt process, 442
graph, 1187, 1188
1PI, 1187
Grassmann
algebra, 471
bundle, 625
number, 478
Grassmannian, 458
Lagrangian, 713
gravity, 985
Greek, 1255
Green
function, 426
identity, 463
Green–Liouville method, *see* WKB approximation
Greenberger–Horne–Zeilinger state, 1081
Gribov ambiguity, 1165
Groenewold–Van Hove, 1094
Gromov
nonsqueezing theorem, 708
Grothendieck, 248, *see also* topos
completion, 60

construction, 95
fibration, *see* fibred category
group, 131, 528
topology, 269
universe, 41
group, 58
action, 67, 828
additive, 241
alternating, 63
arithmetic, 254
categorical, 826
cohomology, 164
localic, 187
profinite, 83, 187
scheme, 241
solvable, 60
topological, 173
groupoid, 113, 320
action, 114
path, 829

H-
$*$-algebra, 515
space, 190, 786
Hölder
conjugates, 386
inequality, 386
Haag
duality, 1208
Haag–Kastler axioms, 1207, 1210
Haar, 408
Haar measure, 408
Hadamard, *see also* Wightmann
derivative, 498
lemma, 356
product, 443
Hahn–Banach, 496
Hahn–Jordan, 394
Hahn–Kolmogorov, 377
hairy ball theorem, 614
Hake, 362
half-
edge, 1188
form, 735
Halmos–Savage factorization theorem, 901
ham-sandwich theorem, 195
Hamilton
action, 723
characteristic function, 988
constraint, 1010
equations, 987
function, 717, 987, 1058
principal function, 988
principle, 983
vector field, 706
Hamilton–Jacobi equation, 719, 861, 988
Hamiltonian
vector field, 807
Hardy space, 1101
Hardy–Littlewood maximal function, 387
harmonic
analysis, 408
cylindrical, 418
form, 690
function, 367, 418
wave, 1045
Hasse diagram, 47
Hasse–Weil
ζ-function, 257
Hauptvermutung, 204
Hausdorff
dimension, 857
space, 175
Hauy, 1124
Heaviside
function, 400
step function, 408
Hecke operator, 255

hedge, 1254, 1257
hedging, 1258
Heine–Borel, 183, 230
Heine–Cantor, 230
Heisenberg
  cut, 1066
  group, 1094
  system, 1063
  uncertainty relation, 1057
helicity, 1157, 1158
Hellinger–Toeplitz, 508
Helmholtz
  conditions, 772
  decomposition, 461
  equation, 419
  free energy, 1118
  operator, 765
hemistrict, 145
Hermitian, 440, 454, *see also* manifold
  form, 442
  operator, 508
Hessian, 631
heteromorphism, 88
Heyting
  algebra, 52
Higgs
  field, 656
  mechanism, 1168
Hilbert
  basis, 494
  basis theorem, 79, 232
  module, 521
  Nullstellensatz, 232
  rigged Hilbert space, 1061, 1076
  space, 440, 493, 946
Hilbert–Schmidt
  norm, 443, 515
  operator, 507
Hochschild–Kostant–Rosenberg, 167
Hodge
  decomposition, 691, 748
  Laplacian, 690
  star, 472
Hodge–de Rham operator, 690
Hoeffding
  inequality, 845
Hoeffding–Azuma inequality, 876
holomorphic
  form, 736
  function, 366
  vector bundle, 736
holonomic
  basis, 618
  constraint, 1005
  section, 757
holonomy, 666, 667, 1108
  bundle, 667
homeomorphism, 178
homogeneous
  equation, 411
  function, 357
  space, 70
homology, 152
  cyclic, 167
  Floer, 1236
  Hochschild, 166
  Koszul, 159
  Morse, 650
  reduced, 206
  relative, 205
  simplicial, 202
  singular, 208
  sphere, 206
homotopy, 189, 291, 341
  (co)limits, 308
  (co)product, 307
  category, 287, 292, 297, 298, 300
  diagram, 309
  equivalence, 200, 292, 336
  fibre, *see* mapping fibre, 336

group, 192
hypothesis, 193
level, 336
thin, 828
type, 189, 192
hook length, 55
Hopf
algebra, 535
fibration, 197
invariant, 787
horizontal, 660
curve, 665
elements, 600
form, 661
lift, 664, 668
vector, 660
horn, 297
Householder transformation, 442
hull, 362
Hume-Rothery conditions, 1127
Hurewicz, *see also* fibration, 209
Hurwitz, 455, 469
Huygens–Fresnel principle, 1046
hyper-
bolic operator, 651
charge, 1202
cohomology, 221
hyperbolic, 722
manifold, 692, 1150
normally, 1213
hyperbolic equation, 421
hypotenuse, 551
hypothesis, 922

idèle, 252
ideal, 73, 502
class group, 81
differential, 153
fractional, 75
gas law, 1116
generating set, 75
principal, 75
identity
operator, 1179
theorem, 235, 367
ill-conditioned, 934
image, 43, 263
immersion, 242, 570
conjecture, 692
property, 892
implication, 326
Inönü–Wigner contraction, 598
incompressible, 989
increasing, 868
ind-object, 105
indecomposable, 134
independence, 840, 841, 846, 849
indeterminate, 79
index, 613, 648
Fredholm, 509
Miller, 1124
of a subgroup, 61
of Dirac operator, 701
of map, 372
of Riemannian manifold, 692
Powers, 523
refractive, 1048
set, 44
topological, 702
indicator function, 263, 381
indicial equation, 413
induction, 331, 1038
inductive
limit, 53, 103
inertia, 976
inextensible, 1150
inference rule, 1084
infimum, 47
infinitesimal, 802, 822
thickening, 76
information, 853

initial, 100
    part, 244
injection, 332
injective, 43
    morphism, 102
    MPS, 1136
    object, 102
inner
    product, 440, 472, 526
inner product
    Petersson, 255
instanton, 1170
    't Hooft–Polyakov, 1173
    BPS, 1235
    BPST, 1172
    deformation complex, 1236
    number, 1171
    Yang–Mills, 1235
integer
    algebraic, 81
    profinite, 252
integers, 71
integrability
    complete, 720, 998
    Liouville, 720
integrable, 382, 636, 868, 883
    G-structure, 673
    completely, 776
    complex structure, 733
    distribution, 627
    locally, 879
integral, 663, *see also* Batalin–Vilkovisky
    contour, 367
    curve, 420, 615
    Daniell, 883
    Denjoy, 361
    first, 717
    form, 638
    Gel'fand–Pettis, 498
    Henstock–Kurzweil, 361
    invariant, 629
    Itô, 882
    Lebesgue–Stieltjes, 391, 889
    Luzin, 361
    McShane, 361
    noncommutative, 793
    Perron, 361
    Riemann, 357, 361
    scheme, 241
    signed measure, 394
    stochastic, 882
    subset, 802
    Wiener–Paley, 880
integrating factor, 628
integration
    by parts, 393, 457, 887
integrator, 883
interaction
    picture, 1105
interchange law, 311
interior, 173
    manifold, 572
    product, 622
intermediate value theorem, 181
internal, 111
    hom, 122, 141
intersection, 571
    form, 1233
    number, 571
interstitial, 1128
intertwiner, 68, 482
interval
    category, 93
invariant
    factor, 67
    inner product, 483
    polynomial, 600
    subgroup, *see* normal subgroup
inverse
    image, 265

limit, 54
quasi, 341
system, 53
inverse function theorem, 245, 610
inversion, 529
formula, 846
invertibility, 956
involution, 44, 529
involutive, 627
irreducible, 72
algebraic set, 232
graph, 1187
polynomial, 80
representation, 482
space, 175
irrep, 482
irrotational, 461
Isbell duality, 832
Ising model, 1120
iso-
geny, 575
metry, 558, 579, 692
topy, 189
tropic manifold, 712
tropy group, 68
isolating integrals, 991
isometry, 226, 490, 514, 517
partial, 490
isomorphism theorem, 59
Isserlis, 1180
Itô
isometry, 888
lemma, 892
process, 884

Jacobi
identity, 576
manifold, 710
method, 940
Jacobiator, 144
Jacobson radical, 74
Jeans, 991
Jensen's inequality, 363
jet, 756
comonad, 758, 812
Jimbo–Drinfeld, 539
join, 51, 181, 285
Jones
relations, 456
Jordan, 372
Jordan–Chevalley decomposition, 438
Jordan–Hölder, 133
series, 133
Joyal, 297
model structure, 302
jump, 393, 401, 867

K-theory, 778
Kasparov, 522
operator, 790
twisted, 788
Kähler
differential, 542
identities, 737
manifold, 738
polarization, 717
potential, 739
Kähler–Dirac operator, 741
Künneth formula, 204
Kac–Moody algebra, 594, 596
Kalb–Ramond field, 1231
Kamiltonian, 988
Kan
complex, 297
extension, 109, 127, 130
fibration, 297
Kantorovich, *see also* Monge
distance, 859
duality, 858
monad, 863
Kantorovich–Rubinstein, 859

Kapranov–Voevodsky 2-vector space, 142
Karamata, 363
Karoubi
envelope, 104
Karush–Kuhn–Tucker conditions, 942
Kasparov bimodule, 522
Kelvin–Stokes, 463
Ken Brown, 294
Kepler
problem, 985
kernel, 59, 107, 131, 438, 946, 948
trick, 950
Khinchin's law, 907
Killing
conformal vector, 703
form, 585, 708
tensor, 694
vector, 694, 740
Kirchhoff's laws, 1035
Kirillov–Kostant–Souriau form, 585
Kleene star, 325
Klein
geometry, 685
Klein–Gordon equation, 1176
Kleisli category, 97
Kochen–Specker, 1090
Kock–Lawvere axiom, 803
Kolmogorov, *see also* topology
axioms, 839
continuity theorem, 870
criterion, 907
extension theorem, 870
law, 907
product, 865
randomness, 856
Kolmogorov–Daniell, *see* Kolmogorov extension theorem
Kontsevich quantization, 1095
Koopman–von Neumann
algebra, 994
mechanics, 992
Koszul
complex, 159
connection, 669
formula, 693
resolution, 159
sign, 475
Koszul–Malgrange, 737
Koszul–Tate resolution, 159
Kraus decomposition, 520
Krein–Milman, 499
Kripke–Joyal semantics, 268
Kronecker
product, 466
Krull
dimension, 233
topology, 77, 83
Krull–Azumaya, 79
Krull–Schmidt, 134
Krylov subspace, 937
Kuiper, 781
Kunita–Watanabe, 891
inequality, 887
Kuratowski, *see* finite set, 381
kurtosis, 910

L_∞-algebra, 145
L-
function of modular form, 258
function of variety, 258
Lévy
process, 870
ladder operators, 1070
lag, 954
Lagrange, 59
bracket, 709
derivative, *see* material derivative
function, 719, 760
multiplier, 1005, 1008
multipliers, 940

submanifold, 712
Lagrange–Charpit equations, 421, 776
Lambert–Beer, 1049
language, 325
Laplace
determinant formula, 448
equation, 418
operator, 461, *see also* Laplace–Beltrami
transform, 407
Laplace–Beltrami operator, 694
Laplace–de Rham, *see* Hodge Laplacian
Laplace–Runge–Lenz vector, 986
Laplace–Stieltjes transform, 407
Larmor frequency, 1040
Lasker–Noether, 74
latching object, 295
lattice, 51
bounded, 51
distributive, 52
modular, 52
orthomodular, 52
vector, 502
Laue
conditions, 1125
indices, 1127
Laurent
series, 368
law, 843
of excluded middle, 53
of large numbers, 907, 911
Lawvere
theory, 114
Lawvere–Tierney, 275
lax
monoidal functor, 121
leaf, 627, 802
least squares, 917
Lebesgue
differentiation theorem, 388
integral, 381, 382, 395, 636
measure, 376, 392
outer measure, 376
restricted measure, 377
Lefschetz
operator, 737
Legendre
transformation, 363, 998
Legendre–Fenchel transformation, 491
Leibniz
construction, 290
integral rule, 358, 989
rule, 620
length, 391
Lepage equivalent, 771
leptons, 1198
Leray, 222
level, 265
set, 43
Levi
decomposition, 582
subalgebra, 583
Levi-Civita
connection, *see* Riemann connection
symbol, 468
libration, 986
Lichnerowicz formula, 702
Lie
2-algebra, 145
algebra, 143, 146, 576
algebroid, 825
bracket, 144, 576, 618
completion, 777
derivative, 616, 621, 807
group, 575, 686
groupoid, 825
operad, 145
pseudogroup, 777
third theorem, 578, 828
Lie–Euler operator, 763

Lie–Poisson
bracket, 726
equations, 726
Lie–Trotter formula, 581
lift, 101, 340
light cone
coordinates, 1147
lightlike, 1149
likelihood, 916
estimator, 917
limit, 103, 110, 163, 177, 353
conical, 128
cycle, 995
directed, 105, 303
filtered, 105, 303
ordinal, *see* ordinal
point, 174
theorem, 372
weighted, 128, 129, 314
Lindelöf
space, 182
line
bundle, 610
integral, 462
linear
fit, 918
form, 438
functor, 142
map, 363, 437
Liouville
class, 713
equation, 990
foliation, 720
form, 707
operator, 987
theorem on entire functions, 368
theorem on phase spaces, 990
vector field, 707
Liouville–Arnol'd, 721
Lippman–Schwinger equation, 1109
Lipschitz
group, 529
little
group, 68
local
coefficients, 210
diffeomorphism, 610, 812, 823
functional, 760, 1211
homeomorphism, 178
isomorphism, 272
object, 90, 306
property, 247
representation, 565
ring, 75
system, 220, 671
topos, 809
trivialization, 603
locale, 185
localic
reflection, 274
topos, 275
locality, 1011, 1207, 1210
strong, 1208
localization, 71, 642, 879
Bousfield, 306
essential, 92
Gabriel–Zisman, 287
hammock, 298
locally
connected topos, 809
constant, 181
finite, 375
integrable, 383
logarithm, 370, 885
logic
modal, 343
logos, 297
Looman–Menchoff, 367
loop
algebra, 595

group, 190
space, 190, 338
Lorentz
factor, 1145
force, 1039
metric, 442
Lorenz gauge, 1042
luff, *see* wide subcategory
Lurie, 315
Lusin space, 381
Lyapunov exponent, 995

Möbius transformation, 254
Müger, *see* dimension, categorical
Møller operator, 1106
MacLane
coherence theorem, 119
strictness theorem, 122
Maclaurin expansion, *see* Taylor
Maclaurin–Cauchy integral test, *see* convergence
magma, 58
magnetization, 1038
Majorana
adjoint, 1157
majorization, 353
Malus–Dupin, 1048
manifold, 565, 808, 816, 823
complex, 732, 740
flat, 895
Hermitian, 738
integral, 627, 628
linear Poisson, 1095
statistical, 895
toric, 724
mapping
cone, 198
cylinder, 198
fibre, 197
path space, 197
space, 183, 189, 300
mapping class group, 189
market
perfect, 1260
viable, 1257
marking, 742
Markov
category, 863
inequality, 844
kernel, 862
process, 893
Marsden–Weinstein
reduction, 725
martingale, 875, 879
σ-, 884
representation theorem, 890
transform, *see* integral, stochastic
Maschke, 483
Maslov
covering, 750
cycle, 713
index, 713
Maslov–Hörmander, 712
mass, 841, 1146
reduced, 985
Massieu potential, 1118
master equation, 807, 1185
matched pair, 65
matching
family, 269
object, 295
material
derivative, 980
volume, 990
matrix
adjacency, 1188
coefficient, 484
conjugation, 447
exponentiation, 581
matrix product
operator, 1137

state, 1134
Maurer–Cartan
equation, 147, 663
form, 662
maximum, 47
principle, 422
Maxwell, 1040
Mayer–Vietoris, 211
sequence, 645
McDiarmid inequality, 876
mean, 907
value theorem, 355, 382
measurable, 45, 377, 794
function, 377
operator, 513
progressively, 868
set, 376
measure, 374, 388, 636, 792, 793, 967
dominated, 390
equivalent martingale, 1255
equivalent separating, 1255
Lebesgue–Stieltjes, 391, 394
projection-valued, 511
risk-neutral, 1255
signed, 393
singular, 390
measurement, 1057, 1064
positive-operator valued, 1082
median, 909
mediator, 928
meet, 51
Mellin, 407
memory, 852, 954
Mercer, *see also* kernel, 948
meromorphic, 369
meta
linear group, 734
metaplectic
correction, 1098
group, 727
structure, 727, 735
metric, 225, 490
coefficients, 557
form, 557
Fréchet, 496
maximum, *see* distance, Chebyshev
quantum, 795
Riemannian, 689
metrizable, 226
Meusnier, 559
microcausality, *see also* locality, 1214
Milnor
construction, 658
minimal
coupling, 1073, 1163
polynomial, 82
sector, 1015
minimum, 47
variance bound, 916
Minkowski
inequality, 387
minor, 448, 449
Mitchell–Bénabou language, 267
Mittag-Leffler condition, 187
mixing, 379
mixture, 842
modal
operator, *see also* closure operator, 344
type, 98, 344
modality, 344, *see also* shape, 1089
mode, 909
model, 114, 153
category, 288, 305
geometry, 686
modification, 313, 867
Modigliani–Miller, 1260
modular, *see also* lattice
category, 140
form, 254

group, 254
modularity theorem, 258
module, 97
category, 141
crossed, 70, 827
over a group, 68
over a ring, 77
weight, 587
moduli space
of differential forms, 817
modulus, 366
of continuity, 227
modus ponens, 326
moment, 844, 907
generating function, 844
map, 720, 723, 740
momentum, 976, 1003, 1146
angular, 488
conjugate, 979
monad, 96, 801
codensity, 110
Mondrian, 961
Monge
cone, 776
Monge–Kantorovich problem, 857
monic, 80
monodromy, 683
monoid, 58
monoidal
n-category, 148
category, 118, 312
equivalence, 121
functor, 120
model category, 291
monomorphism, 99, 316
monotone
class, 46
function, 47
Moore
complex, 285
Moore–Aronszajn, 947
Morera, 368
morphism
geometric, 265
induced, 578
logical, 265
of G-sets, 68
of cones, 103
of finite type, 242
of groups, 59
of Lie algebras, 578
of orbifolds, 743
of presheaves, 216
of sites, 278
of varieties, 232, 235
of vector spaces, 437
Morse, *see also* homology
family, 712
function, 648
lemma, 649
Morse–Bott function, 649
Morse–Palais, 649
Morse–Smale pair, 650
moving average, 956
MPO
injectivity, 1138
multi-
plectic manifold, 708
set, 907
vector, 526
verse, 1244
multiplicity, 453
musical isomorphism, 690

n-category, 310
Néel
hyperbola, 1133
temperature, 1132
Naimark dilation theorem, 1083
Nakanishi–Lautrup field, 1015
Nakayama lemma, 79

Nambu, *see also* Goldstone boson
natural, 868
  chart, 609
  pairing, 439
  parameter, 555
  transformation, 88, 108, 125
necessity, 343, 969
negation, 326
negative, 346
neighbourhood, 174, 801, 820
  deformation retract, 206
nerve, 283, 295
net, 48
Neumann
  function, 416
  principle, 1125
neutrino, 1198
Newlander–Nirenberg, 733
Newton
  equation, 996
Newton–Leibniz, 358
Newton–Raphson algorithm, 939
Neyman–Pearson, 923
Nielsen–Schreier, 67
Nijenhuis
  tensor, 733
Nijenhuis–Richardson bracket, 625
nilpotent, 58, 438
no free lunch, 1258
no-cloning, 1080
no-deleting, 1080
Noether, 718, 754, 772, 984
  charge, 1002
  ring, 74
  theorem, 997, 1001
non-
  autonomous, 717
nondegeneracy, 648, 722
norm, 490
  Gross, 874
  measurable, 873
  operator, 505
  spinor, 530
  total variation, 394
normal, 176
  bundle, 611
  closure, 83
  curvature, 558
  element, 514
  equation, 933
  ordering, 1179
  section, 559
  subgroup, 60, 61
  vector, 555, 557
normalized
  complex, 285
normalizer, 61, 581
notation
  braket, 493, 1056
  Dirac, 1056
nothing, 346
Novikov
  condition, 888
  covering, 751
  field, 750
nuclear
  space, 502
nucleus, 1091
null
  set, 374, 636
  vector, 1149
nullity, 438
Nullstellensatz, 84, *see also* Hilbert, 237, 810, 817
numéraire, 1255
number
  cardinal, 50
  field, 81
  ordinal, 49
numerable, 183, 604

observable, 1056, 1064, 1091, 1103
    compatibility, 1056
Ohm, 1036
Oka
    manifold, 734
Oka coherence theorem, 733
Oka–Grauert principle, 733
on-shell, 763, 1018
one-parameter subgroup, 576
one-to-one, *see* injective
onto, *see* surjective
open
    algebra, 1008, 1027
    mapping theorem, 491
    system, 1115
operad, 115
    little k-cubes, 116
operator, 438
    product expansion, 1224
    topology, 505
opposite, 87, 344
    ring, 78
optimum, 942
option, 1253
optional stopping, 876
orbifold, 743
orbit, 68, 114
    category, 201
order, 46
    -preserving, 47
    of a distribution, 398
    of a group, 59
    of a jet, 756, 777
    of finite field, 250
orientable
    manifold, 633, 636, 673, 699
orientation, 472, 633–635, 648
    class, 636, 643
    time, 1149
orthant, 966
orthogonal, 441
    group, 450
    lifting, 102
    measure, *see* measure, singular
orthogonality, 878
orthonormal, 441
oscillation theorem, 416
osculating circle, 559
osculation plane, 561, 562
Ostrowski, 252
outer
    measure, 374, 391
    product, 463
over
    category, 93
OZI rule, 1186

PAC theory, 943
packing, 226
page, 160
pair, 332
    groupoid, 826
pairing, 330
Palais–Smale condition, 648
Palatini action, 1153
Paley–Wiener, 406
par, 1252
para-
    bolic subgroup, 458
    compact, 182
    magnetism, 1131
paradox, 326
parallel, 630
    transport, 666, 668
parallel axis theorem, 978
parallelizable, 541, 610
parallelogram law, 492
parametric family, 842
parametrix, 509
Parceval, 405, 495
parity, 63, 478, 1157

functor, 475
of matrix, 477
put-call, 1260
Parke–Taylor formula, 1190
partition, 44, 54, 55
function, 1118
of unity, 182
pasting law, 107
path, 867, 1115
connected, 181, 202
integral, *see* line integral
lazy, 829
space, 818
Pauli
matrix, 532, 1071
Pauli–Jordan function, 1178
payoff, 1252
PDE, 774
inhomogeneous, 425
Pearson
skewness, *see* skewness
Peierls bracket, 1214
Peiffer identity, 70
pentagonator, 150
percentile, 910
perfect
field, 83
period, 1045
periodic
extension, 404
permanent, 1065
permeability, 1038, 1039
permutation
module, 484
representation, 482
Perron–Frobenius, 453
perturbation, 314, 1105
Peter–Weyl, 484
Pfaffian, 679
system, 420, 628
phase, 1045
phase space, 986, 1026, 1090
reduced, 1012
Picard, 370
group, 224
Picard–Lindelöf, 411
piezoelectricity, 1130
pin, 530
Piron, 1084
pivotal structure, 137
place, 73
Plancherel, 405
Planck, 1122
plausibility, 969
Plebanski form, 1154
plot, 816, 820
Plymen's criterion, 700
Poincaré
dual, 641, 643
duality, 641, 647
groupoid, 191
lemma, 638, 641, 643
map, 995
recurrence theorem, 991
Poincaré–Bendixson, 994
Poincaré–Birkhoff–Witt, 596
Poincaré–Hopf, 613
point, 266
pointed
category, 100
object, 100
tensor category, 139
topological space, 173
type, 338
points-to-pieces, 810
Poisson, *see also* algebra
bracket, 577, 709, 987
manifold, 709, 807
map, 725
reduction, 726

superalgebra, 1014
tensor, 709
vector space, 1095
Poisson algebra
reduced, 1013
Poisson–Vlasov bracket, 727
Polak–Ribière formula, 938
polar, 500
form, 366
polarization, 716, 1047, 1097, 1130
identity, 492
pole, 369
Polish space, 381
poly-
hedron, 202
morphic, 330
nomial, 79, 370
tabloid, 485
polynomial, 337
Pontryagin
class, 679
dual, 408
portfolio, 1254
poset, 46
positive, 516
completely, 516
line bundle, 737
possibility, 343, 969
Postnikov tower, 193
potential, 975, 1034, 1041, 1043
chemical, 1119
presymplectic, 1001
Pouillet, 1035
power, 126, 922, 926
set, 39, 44
Poynting vector, 1039
pre-
image, 43
Lie algebra, 576
measure, 377
order, *see* order
quantum line bundle, 1095
stack, 316, 319
variety, 236
predictable, 868
presentable, 105, 303
presentation, 63, 66, 76
presheaf, 87, 215, 316
separated, 270
topos, 268
price, 1255, 1256, 1259
primary
field, 1222
ideal, 74
primitive, 663
element, 534
principal
axis, 978
bundle, 653, 830
part, 368
symbol, 427
value, 370, 401
prior
conjugate, 919
pro-
category, 105
finite set, 186
functor, 88, 113
group, 187
object, 105
probability, 904, 968
conditional, 840
distribution, 841
simplex, 201
space, 839, 867
product, 42, 106, 107, 330, 331
restricted, 252
space, 228
topology, 179
progressive, *see* measurable

projectable, 712
projection, 330, 503, 515
    finite, 523
    formula, 643
    function, 331, 526
    orthogonal, 441, 904
    stereographic, 184
    theorem, 495, 904
projective
    cover, 134
    group, 487
    limit, 54, 103
    module, 79
    morphism, 102
    object, 102
    space, 236
prolongation, 156, 755, 756, 760
propagation
    of errors, 913
propagator, 1214
proper
    action, 184
    function, 183
    model category, 288
proposition, 267, 326, 335
pseudo-
    complement, 52
    functor, 313
    group, 178, 777
    natural transformation, 313
    scalar, 526, 634
    vector, 526, 635
pseudoholomorphic, 746
Puiseux series, 249
pullback, 106
    of a bundle, 606
    of a differential form, 619
    of a fuzzy set, 966
    of a vector field, 614
    of cohomology, 209
pulling-through condition, 1139
Puppe sequence, 786
pure, 1082
    set, 40
    vector, 470
purification, 1082
pushforward
    of a differential form, 620
    of a measure, 378
    of a vector field, 614
    of homology, 208
pushout, 107
    product, 290
put, 1253
Pythagoras, 493, 552, 904

QR
    decompositon, 451
quantile, 392, 910
quantity, 832
quantization, *see also* deformation, *see also* deformation
    Dirac condition, 1096
    geometric, 1097
    half-form, 1098
    Kähler, 1098
    Schwinger–Zwanziger condition, 1173
quantum
    algebra, 1217
    channel, 1083
    geometry, 795
quark, 1199
quartile, 910
quasi-
    category, 297
    cocommutative, 536
    isomorphism, 152
quaternions, 451
Quillen, 290, 305
    adjunction, 289

bifunctor, 290
model structure, 289
quotient
group, 61
space, 179
stack, 213

R-matrix, 536
Radó–Moise, 565
Rademacher
complexity, 952
variable, 841
radial ordering, 1224
radical, 74, *see also* extension, 582
Radon
measure, 375
space, 376
Radon–Nikodym, 389, 394
Raikov, 853
Raman scattering, 1050
Ramanujan
conjecture, 256
random
variable, 795, 839, 846
range, 909, 910
rank, 224
of a function, 610
of a group, 67
of a Lie algebra, 588
of a linear map, 438
of a matrix, 443
of a vector bundle, 610, 779
Schmidt, 1079
rank-nullity theorem, 438
rate
interest, 1262
rational
conformal field theory, 1227
fractions, 71
function, 233
homotopy equivalence, 198
map, 233
point, 232, 240
space, 198
Rayleigh
scattering, 1050
Rayleigh–Schrödinger theory, 1106
real numbers, 41
realization, 594
reciprocal lattice, 1125
recognition principle, 116
rectifiable curve, 367
recursion, 330
reduced
ring, 71, 76
scheme, 241
reduction, 76, 673, 723, 822
reductive, 686
Reeb vector field, 730
Reedy
category, 295
model structure, 296
reference frame, 1069
quantum, 1067
refinement, 44
reflection, 88
reflective
category, 92
localization, 287
reflexive, 491
refraction, 1048
Regge trajectory, 1240
regression
lasso, 934
ridge, 934
regular, 176, 722
cardinal, 50
curve, 554
function, 234
morphism, 106
ordinal, 49

scheme, 245
sequence, 159
value, 569, 571
regularity, 39
Reidemeister move, 120
Reissner–Nordström metric, 1152
relation, 42, 63
relative
homotopy group, 192
remanent, 1130
renormalizability, 1189
renormalization
BPHZ, 1195
minimal subtraction, 1195
on-shell, 1191, 1195
replete, 86
replicating, 1255
representable, 89, 273
representation, 89
alternating, 482
of a C^*-algebra, 518
of a group, 481
of a Lie algebra, 584
of a Lie group, 583
operator, 487
projective, 486
reservoir, 1115
residual, 935
residue, 371
field, 76
resolution, 155, 289, 299
(co)simplicial, 299
bar, 158, 165
Koszul, 159
Koszul–Tate, 159
of the identity, 135, 1057
standard, *see* bar resolution
resolvent, 510
retract, 99, 189
metric, 226
Reynolds' transport theorem, 989
Ricci
scalar, 696
tensor, 695
Riemann, *see also* integral, 369
ζ-function, 256
ζ-function, 258
cone, 692
connection, 692
curvature, 695
curvature tensor, 562
hypothesis, 257
manifold, 690
sphere, 366
surface, 741
Riemann–Lebesgue lemma, 384
Riesz
representation theorem, 493
space, 502
Riesz–Markov, 493
Riesz–Markov–Kakutani, 398
rig, 248
ring, 70
ringed space, 223
risk, 1255
free, 1251
Robertson, *see* Heisenberg uncertainty relation
Rodrigues's formula, 560
root, 80, 370, 587
system, 593
rotation, 986
rotor, *see* curl, 462
Rund–Trautman identity, 755
Runge–Kutta method, 1195
Ryu–Takayanagi formula, 1244

S-matrix, 140
sample, 867
space, 839
Sanov, 919

Sard, 570
saturated, 305
Sauer's lemma, 951
Savitch, 1249
scalar, 118, 464
  curvature, 696
scale
  factor, 557
scattering
  factor, 1127
Schatten class, 506
Schauder
  fixed point theorem, 194
scheme, 240
  transportation, 857
Schmidt
  decomposition, 1079
Schottky
  defect, 1128
Schouten
  identity, 1160
Schouten–Nijenhuis bracket, 618
Schrödinger
  equation, 1058, 1105
  representation, 1063
Schubert
  cell, 238
  decomposition, 238
Schur
  complement, 478
  functor, 485
  lemma, 134, 483
Schur–Weyl duality, 485
Schwartz
  kernel theorem, 403
Schwartz space, 401
Schwarz, 356
  reflection principle, 369
Schwarzschild metric, 1152
Schwinger–Dyson equation, 1181, 1185
Schwinger–Zwanziger, *see* quantization condition
SDE, 884, 892
secant, 553
secant method, 939
section, 99, 340, 389
  of a (pre)sheaf, 215, 217
  of a bundle, 606
  of a product set, 388
sectional
  curvature, 695
security, 1251
Segal
  category, 300
  condition, 300
  space, 300, 301
Segal–Bargmann representation, 1098
Segal–Kostant–Souriau procedure, 1096
Seiberg–Witten theory, 1237
Selberg, *see* ζ-function
self-adjoint, 93, 440, 508
self-financing, 1254
semi-
  continuity, 177
  direct product, 64, 582
  free, 153
  group, 58
  norm, 496
  spin, 532
semiclassical, 1075
semisimple
  module, 79
  operator, 438, 453
separable, 82, 495
  closure, 83
  holomorphically, 733
  space, 184
  state, 1080
separated, 175, 242, 276

family, 496
separation, 928
axioms, 175
separator, *see* generator
separatrix, 987
sequent, 327
sequential
space, 175
Serre, *see also* fibration
twisting sheaf, 626
Serre–Swan, 611
sesquilinear, 440
set
builder notation, 40
enumerable, 1249
singular, 284
set theory
material, 40
structural, 40
shape, 100, 822
modality, 810
operator, *see* Weingarten map
sharp
modality, 810
Sharpe ratio, 1256
shatter, 950
sheaf, 216, 270, 276, 321
acyclic, 220
coherent, 223
flabby, 219
flasque, *see* flabby
invertible, 224
of sections, 217
soft, 219
sheafification, 217, 272, 277
shear stress, 992
shuffle, 63
sieve, 268
signature, 145, 325, 442, 458, 692, 1233
significance, 922, 958
Simons–Sullivan, 683
model, 788
simple
extension, 81
function, 381, 503
Lie algebra, 582
module, 79
object, 133
ring, 74
simplex, 201, 282
category, 282
infinitesimal, 801
singular, 207
standard, 201
simplicial
complex, 202
localization, 298
object, 282, 295
set, 282
simply-connected, 191
sine, 364, 551
singular
trace, 514
value, 512
singularity
essential, 369
nodal, 745
removable, 369
site, 269, 316
étale, 246
sitting instants, 829
skeleton, 195, 294
skewness, 909, 910
Sklar, 855
Slater
conditions, 941
determinant, 1065
slice, 570
slope, 803
Slutsky, 843

small, 86, 303
    object argument, 305
smash product, 180
smooth, *see also* formal
    algebra, 819
    distribution, 399
    function, 356, 498, 565, 816
    locus, 819
    manifold, 565
    set, 817, 821
    space, 813
Snell's law, 1048
sober, 186
Sobolev
    space, 427
soft
    algebra, 1008, 1027
Sohnius–Lopuszański–Haag, 1232
Sokhotski–Plemelj, 368
solder form, 672
solenoidal, 460
solid, *see* cohesion, super-differential
solvable, 582
source
    form, 770
space, 831
spacelike, 1149
span, 106, 435
Specht module, 485
specification, 39
spectral
    density, 954
    flow, 1236
    measure, 511
    presheaf, 1090
    sequence, 160
    theorem, *see also* Freudenthal, 512, 513
    triple, 796
spectrum, 186, 212, 239, 510
    Thom, 644
sphere
    bundle, 644
    celestial, 1151
spherical structure, 138
spin, 530, 696, 700, 1097
    connection, 1153
    orbitals, 1065
spin-statistics theorem, 1065
spine, 300
spinor, 530, 696
    bundle, 698
Spitzer's identity, 457, 1196
split, 80
    cleavage, 96
    morphism, 99, 106
    product, 64
splitting field, 82, 454
splitting principle, 676
spot, 1254
    measure, 1256
Stäckel potential, 989
stabilization hypothesis, 149
stabilizer, 68
stable
    curve, 745
    group, 779
    isomorphism, 612
    manifold, 650
stack, 316, 319, 321
stalk, 178, 217
    generic, 235
standard
    deviation, 844
    error, 915
    filtration, 867
standardization, 851
star product, 1094
Stasheff
    operad, 116

state, 516, 517, 1056, 1064, 1091
    normal, 518
    pure, 516
stationarity, 953, 956
stationary, 1151
steady-state
    method for boundary conditions, 424
Stein manifold, 733
Steiner, 978
Steno, 1124
step function, 381
Stiefel manifold, 571
Stiefel–Whitney class, 698
Stinespring, 519
Stirling, 359
stochastic, *see also* integral
    map, 862
    matrix, 857, 862
    process, 867
Stokes, 637
Stone, 229, 508
    duality, 186
    space, 186
Stone–Čech compactification, 380
Stone–von Neumann, 1063
stopping time, 869
strange
    correlator, 1227
strange attractor, 995
stress, 992
stress-energy tensor, 1002, 1151
structure
    constants, 577, 586
    factor, 1127
    group, 603
    sheaf, 234, 240, 566
Sturm–Liouville theory, 415
sub-
    topos, 265
subdivergence, 1191
submanifold, 570
submersion, 570
    theorem, 571
subobject, 99
    classifier, 262
successor, 49, 50
Sullivan
    algebra, 199
    model, 200
sum, 436
super-
    algebra, 474
    manifold, 806, 824
    operator, 1083, 1135
    point, 820
    scheme, 819
    space, 820
    symmetry, 1232
    symplectic structure, 1014
    trace, 477
    vector space, 474
support, 43, 399
    singular, 399
support-vector machine, 950
supremum, 47
    norm, 387
surjective, 43, 265, 314
    essentially, 88
susceptibility, 1038
Suslin space, 381
suspension, 180, 475, 789
swap, 1253
Sweedler notation, 534
Sylow, 61
    subgroup, 61
Sylvester's law of inertia, 454
symbol, 426
symmetric
    difference, 42

group, 62
operator, 508
part, 467
symmetrization postulate, 1065
symmetry
Bessel-Hagen, 771
distinguished, 771
dynamical, 718
factor, 1187, 1188
kinematical, 718
Lagrangian, 997
quantum, 1064
varaitional, 754
symplectic
group, 451
manifold, 706
potential, 707
symplectization, 999
symplectomorphism, 708
system, 1114

tailedness, 910
tame, 738, 794
tangent, 364, 551
bundle, 608, 806, 808, 822
cone, 244, 245
plane, 557
space, 244, 567, 609
vector, 555, 567
Tannaka duality, 135, 140, 537
tautological
bundle, 626
form, *see* Liouville
tautology, 326
taxonomy, 960
Tayinama–Shimura, 258
Taylor
expansion, 356, 803
Teichmúller space, 742
temperature, 1118
tenfold way, 527
tensor
algebra, 464
category, 138
density, 634
operator, *see* representation operator
power, 464
type, 464
tensor product, 118
Deligne, 134
exterior, 612
infinite, 865
of functors, 125, 127
of rings, 78
of topological vector spaces, 500
of vector bundles, 612
of vector spaces, 463
projective, 501
representation, 482
spatial, 518
tensorial, 661
term, 267, 325, 328
terminal, 100
tesla, 1038
test
χ^2, 926
Bonferonni–Dunn, 925
Dickey–Fuller, 955
Diebold–Mariano, 958
Friedman, 925
function, *see* bump function
Holm, 926
Kolmogorov–Smirnov, 911
likelihood ratio, 923
Ljung–Box, 955
Mann–Whitney, 924
McNemar, 923
Nemenyi, 925
runs, 927
Shapiro–Wilk, 922

Wilcoxon, 924
tetrad, 1153
Theorema Egregium, 562
thickened space, 819
thin, 681
Thom
class, 643, 648
isomorphism, 643, 647
space, 644
tidal force operator, 695
Tietze extension theorem, 177
Tikhonov regularization, 934
time
slice, 1208, 1210
time ordering, 1059
time series, 953
timelike, 1149
tiny, 104
Todd class, 678
Toeplitz
operator, 1101
quantization, 1101
Tomonaga–Schwinger equation, 1106
Tonelli, 389
topology, 172
σ-weak, 507
Émery, 880
adic, 77
algebraic, 1063
canonical, 271
compact-open, 183
curvaceous, 815
double negation, 276, 1091
final, 177
Fr échet, 175
functional, 815
initial, 176
Kolmogorov, 175
limit, *see* initial
metric, 226
norm, 490, 505
operator, 507
polar, 499
strong, 490, *see also* polar topology
weak, 490
weak-*, 490
topos
Cahiers, 820
cohesive, 810
Dubuc, 821
elementary, 263
Grothendieck, 272
petit & gros, 274
sheaf, 272
spatial, 271, 274
Tor, 158
torque, 978
torsion, 556, 632, 672, 687, 795, 802
group, 59, 67
torsor, 70
total
derivative, 760
vector field, 761
totality, 47
tower law, 61, 81
trace, 137, 138, 443, 453, 495, 517
class, 506
transcendental, 81
function, 256
transducer, 1131
transfer operator, 1135
transfor, 314
transgression, 600, 818
transition map, 564
transitive
action, 69
component, 114
groupoid, 114
transport, 340
transport equation, 1075

transpose, 439
transposition, 63, 528
transversality, 571
triangle
    identities, 91
triangular decomposition, 588
triangulation, 202
triple product, 469, 556
trivial, 605
tropical geometry, 248
tropicalization, 249
Trotter expansion, *see* Lie–Trotter formula
trunk, 1193
truth value, 267
tubular neighbourhood, 611
Turing machine, 1248
twist, 613
twisted, *see* Kac–Moody algebra
twisted conjugation, 529
twistor theory, 1223
Tychonoff
    compactness theorem, 183
    topology, 179
type, 267
    judgement, 328

ultrafilter, 48
ultrametric, 72
umbilical point, 560
uniform
    MPS, 1137
    space, 184
union
    ordered, 48
unit, 71, 91, 135
    type, 330
unitary, 93
unity, 344
univalence, 343
universal
    bundle, 657
    covering space, 191
    enveloping algebra, 596
    morphism, 92, 314
    property, 464
universe, 325
unramified, 76, 243
urelement, 40
Urysohn
    lemma, 177
    metrization theorem, 227
    space, 176
usual conditions, 867

Vašíček model, 1261
vacancy, *see* Schottky defect
vacuum, 1226
    θ, 1172
    Higgs, 1168
    topological, 1171
valence band, 1129
validity, 958
valuation, 72, 73
    ring, 253
Van Holten algorithm, 718
vanish at infinity, 398
Vapnik–Chervonenkis
    dimension, 951
variable, 267, 325
variance, 844, 909
    estimator, 915
    inflation factor, 935
variation, 394
    quadratic, 877, 878
variational
    form, 772
    formula, 764, 770
variety, 232, 236
    Abelian, 248
    affine, 235
vector, 447

bundle, 610, 657
space, 434, 490
vector field, 459, 613, 806
cohomological, 806
fundamental, 659
generalized, 759
gradientlike, 649
left-invariant, 575
symplectic, 708
velocity, 1146
group, 1046
phase, 1045
Verlinde formula, 141
Verma module, 597
version, 867
vertex, 201, 1188
algebra, 1226
operator algebra, 1226
vertical
automorphism, 654
bundle, 658
form, 661
polarization, 716
Vidal, 1136
vielbein, 1153
Vieta, 80
Ville, *see also* Doob
Virasoro algebra, 1225
virial theorem, 1120
virtual
displacement, 980
rank, 779
vector bundle, 779
Vlasov equation, 991
Volterra–Vainberg formula, 773
volume
form, 468, 633, 797
von Mises representation, 905
von Neumann
algebra, 522
ordinal, 49
Voronoi diagram, 946

W-type, 338
walking
arrow, 93
wall, 1115
Wasserstein metric, 859
Watson, 359
wave
carrier, 1046
equation, 419
front, 407, 704
function, 610
number, 1045
weak
equivalence, 192, 286
inverse, 119, 312
type, 387
wedge, 108
product, 470, 623, 624
sum, 180
Weierstrass
extreme value theorem, 352
weight, 587, 1225, 1227
dominant, 589
fundamental, 592
lattice, 592
Weil
algebra, 600, 819
conjectures, 257
integrality, 1095
Weingarten
formulas, 560
map, 559
Weiss, *see* Curie–Weiss
domain, 1132
Weitzenböck identity, 1239
well-
founded, 48
order, 48

pointed, 101
powered, 99
Werner–Simpson formulas, 553
Weyl
basis, 1155
chamber, 589
equation, 1156
group, 588
relations, 1062
spinor, 531
system, 1063
Whitehead, 196, 201, 293
bracket, 192
lemma, 600
tower, 193
Whitney, 565
C^k-topology, 757
embedding theorem, 692
immersion theorem, 692
product formula, 677
sum, 611
Wick, 1180
lemma, 360
wide subcategory, 86
Wien
constant, 1122
displacement law, 1122
Wiener, *see also* integral
measure, 870
process, 871
space (abstract), 875
Wightmann
propagator, 1215
Wigner, 1064
D-function, 1071
friend, 1067
Wigner–Eckart, 488
Wigner–Seitz cell, 1125
Wilson score interval, 921
winding number, 372, 613, 615
Wirtinger derivative, 366
Witt algebra, 1221
Witten, 1174
WKB approximation, 1074
word, 66
length, 475
work, 975
Wronskian, 412

Yang–Baxter, 120
Yang–Mills
field, 663
field strength, 672
Yang–Mills theory, 1166
Yang–Mills–Higgs equations, 1167
Yin & Yang, 345
Yoneda
embedding, 90
extension, 278
lemma, 89
reduction, 109, 127
Young, 1050
diagram, 54
inequality, 403
symmetrizer, 485
tableau, 55
Yukawa interaction, 1200

Zamolodchikov tetrahedron equation, 144
Zappa–Szép product, 65
Zariski
tangent space, 244
topology, 234, 237, 239, 245
zero
object, 100, 130
zig-zag, *see* triangle identity
zilch symmetry, 1027
zone, 1124
Zorn's lemma, 48

Authors

Amari, 903
Atiyah, 724

Baez, 193
Bekenstein, 1244
Bohr, 1066

Cantor, 810
Cartan, 685
Cohen, 692
Csiszár, 897

Daniell, 883
Deligne, 258
Delzant, 724
Dirac, 1005, 1009

Einstein, 1151
Everett, 1066

Fedosov, 1095

Gibbons, 1151
Gleason, 1090
Granger, 958
Green, 1074
Grothendieck, 258
Guillemin, 724

Hawking, 1151, 1244
Hegel, 346, 810
Holst, 1154

Kantorovich, 857
Kleiman, 246
Klein, 685
Kochen, 1090
Kontsevich, 1095
Koopman, 992
Kriegl, 808

Lévy, 870, 879
Lawvere, 345
Lie, 754
Lindeberg, 851
Liouville, 1074
Lyapunov, 851
Lynden-Bell, 991

Martin-Löf, 857
Michor, 808
Milnor, 658
Monge, 857

Parke, 1190
Peierls, 1025

Rényi, 897
Roytenberg, 144
Ryu, 1244

Savitch, 1249
Schwarz, 808
Singer, 1166
Specker, 1090
Sternberg, 724

Takayanagi, 1244

Taylor, 1190
Tegmark, 1244
Tsallis, 897

Vistoli, 246

von Neumann, 49, 992, 1064, 1083

Whitney, 692
Wigner, 1064

York, 1151