Compendium of Mathematics & Physics

Nicolas Dewolf

June 30, 2022

Contents

Chapter 1

Introduction

1.1 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 should be used as lookup 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 short as possible. However, in some cases the major ideas underlying the proofs are given.

1.2 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 a *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 of words in the middle of a 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 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 we might also use upper case 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 we will often adopt specific conventions for the different types of vectors. Vectors in Euclidean space will be denoted by a bold font letter with an arrow above, e.g. $\vec{\boldsymbol{a}}$. 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

Chapter 2

Set Theory

2.1 Axiomatization

2.1.1 ZFC

The following set of axioms and axiom schemata gives a basis for axiomatic set theory that fixes a number of issues in naive set theory where one takes the notion of set for granted. This theory is called **Zermelo-Frenkel** set theory (ZF). When extended with the axiom of choice (see further) it is called ZFC, where the C stand for "choice".

Axiom 2.1 (Power set).

$$\forall x : \exists y : \forall z\big[z \in y \iff \forall w(w \in z \implies w \in x)\big] \tag{2.1}$$

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

Axiom 2.2 (Extensionality).

$$\forall x, y : \forall z\big[z \in x \iff z \in y\big] \implies x = y \tag{2.2}$$

This axiom allows us to compare two sets based on their elements.

Axiom 2.3 (Regularity[1]).

$$\forall x : (\exists z \in x) \implies (\exists a \in x) \wedge \neg(\exists b \in a : b \in x) \tag{2.3}$$

This axiom says that for every non-empty 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 2.4 (Specification).

$$\forall w_1, \ldots, w_n, A : \exists B : \forall x\big(x \in B \iff (x \in A \wedge \varphi(x, w_1, \ldots, w_n, A))\big) \tag{2.4}$$

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.

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

2.1.2 Material set theory

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

Definition 2.1.1 (Pure set). A set U such that for every sequence $x_n \in x_{n-1} \in \cdots \in x_1 \in U$ all the elements x_i are also sets.

Definition 2.1.2 (Urelement[2]). An object that is not a set.

2.1.3 Universes

?? TODO ??

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 performed on its elements, i.e. the sets that one wants to work with, act within this universe.

Definition 2.1.3 (Grothendieck universe). A Grothendieck universe U is a pure set satisfying the following axioms:

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 $\{x, y\} \in U$; and
4. **Unions**: If $I \in U$ and $\{x_i\}_{i\in I} \subset U$, then $\bigcup_{i\in I} x_i \in U$.

2.1.4 Structural set theory

In contrast to material set theory, the fundamental notions in this theory are sets and the relations between them. An element of a set does not have any internal structure and only becomes relevant if one specifies extra structure (or 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 internal structure. Even stronger, it is meaningless to compare two elements if one does not provide relations or extra structure on the sets.

?? COMPLETE ??

2.1.5 ETCS ♣

?? COMPLETE ??

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

Axiom 2.5. The category of sets is a well-pointed (elementary) topos.

2.1.6 Real numbers

Axiom 2.6 (Ordering). The set of real numbers is an ordered field $(\mathbb{R}, +, \cdot, <)$.

Axiom 2.7 (Dedekind completeness). Every non-empty subset of $\mathbb{R}$ that is bounded above has a supremum.

Axiom 2.8. The rational numbers form a subset of the real numbers: $\mathbb{Q} \subset \mathbb{R}$.

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

Remark. There is only one way to extend the field of rational numbers to the field of reals such that it satisfies the previous axioms. This implies that for every two possible constructions, there exists a bijection between the two.

Definition 2.1.4 (Extended real line).

$$\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\} \equiv [-\infty, \infty] \tag{2.5}$$

2.2 Set operations

Definition 2.2.1 (Symmetric difference).

$$A\Delta B := (A\backslash B) \cup (B\backslash A) \tag{2.6}$$

Definition 2.2.2 (Complement). Let Ω be the universe of discours (Section 2.1.3) and let $E \subseteq \Omega$. The complement of E is defined as follows:

$$E^c := \Omega\backslash E. \tag{2.7}$$

Formula 2.2.3 (de Morgan's laws).

$$\left(\bigcup_i A_i\right)^c = \bigcap_i A_i^c \tag{2.8}$$

$$\left(\bigcap_i A_i\right)^c = \bigcup_i A_i^c \tag{2.9}$$

Definition 2.2.4 (Relation). A relation between 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 2.2.5 (Converse relation). Consider a relation $R \subset X \times Y$ between two sets X, Y. The converse relation R^t is defined as follows:

$$R^t := \big\{(y, x) \in Y \times X \,\big|\, (x, y) \in R\big\}. \tag{2.10}$$

Definition 2.2.6 (Composition of relations). Consider two relations $R \subset X \times Y$ and $S \subset Y \times Z$ between three sets X, Y and Z. The composition $S \circ R$ is defined as follows:

$$S \circ R := \big\{(x, z) \in X \times Z \,\big|\, \exists y \in Y : (x, y) \in R \wedge (y, z) \in S\big\}. \tag{2.11}$$

2.3 Functions

2.3.1 (Co)domain

Definition 2.3.1 (Domain). Let $f : X \to Y$ be a function. The set X is called the domain of f.

Notation 2.3.2. The domain of f is denoted by $\mathrm{dom}(f)$.

Definition 2.3.3 (Support). Let $f : X \to \mathbb{R}$ be a function with an arbitrary domain X. The support of f is defined as the set of points where f is nonzero.

Notation 2.3.4. The support of f is denoted by $\mathrm{supp}(f)$.

Notation 2.3.5. Let X, Y be two sets. The set of functions $f : X \to Y$ is denoted by Y^X or $\mathrm{Map}(X, Y)$. (See also Definition 4.6.22 for a generalization.)

Definition 2.3.6 (Codomain). Let $f : X \to Y$ be a function. The set Y is called the codomain of f.

Definition 2.3.7 (Image). Let $f : X \to Y$ be a function. The following subset of Y is called the image of f:

$$\{y \in Y \mid \exists x \in X : f(x) = y\}. \tag{2.12}$$

Notation 2.3.8. The image of a function f is denoted by $\mathrm{im}(f)$.

Remark. Some authors use these two notions interchangeably.

Definition 2.3.9 (Level set). Consider a function $f : X \to \mathbb{R}$. The following set is called the level set of f at $c \in \mathbb{R}$:

$$L_c(f) := f^{-1}(c) \equiv \{x \in X \mid f(x) = c\}. \tag{2.13}$$

For $X = \mathbb{R}^2$ the level sets are called **level curves** and for $X = \mathbb{R}^3$ they are called **level surfaces**.

2.3.2 Functions

Definition 2.3.10 (Injective). A function $f : A \to B$ is said to be injective or **one-to-one** if the following condition is satisfied:

$$\forall a, a' \in A : f(a) = f(a') \implies a = a'. \tag{2.14}$$

Notation 2.3.11 (Injective function).

$$f : A \hookrightarrow B$$

Definition 2.3.12 (Surjective). A function $f : A \to B$ is said to be surjective or **onto** if the following condition is satisfied:

$$\forall b \in B, \exists a \in A : f(a) = b. \tag{2.15}$$

Notation 2.3.13 (Surjective function).

$$f : A \twoheadrightarrow B$$

Definition 2.3.14 (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 2.3.15 (Isomorphic). If two sets X, Y are isomorphic, this is denoted by

$$X \cong Y.$$

Theorem 2.3.16 (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 2.3.17 (Involution). A function $f : A \to A$ such that $f^2 = \mathrm{id}_A$, i.e. f is its own inverse. Every involution is in particular a bijection.

2.4 Collections

Definition 2.4.1 (Power set). Let S be a set. The power set is defined as the set of all subsets of S and is (often) denoted by $P(S)$ or 2^S. The existence of this set is enforced by Axiom 2.1, the *axiom of power set.*

Corollary 2.4.2. All sets are elements of their power set: $S \in P(S)$.

Definition 2.4.3 (Collection). Let A be a set. A collection of elements in A is a subset of A.

Definition 2.4.4 (Family). Let A, I be two sets. A family of elements of A with **index set** I is a function $f : I \to A$. 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 2.4.5 (Helly family). A Helly family of order k is a pair (X, F) with $F \subset P(X)$ such that for every finite $G \subset F$:

$$\bigcap_{V\in G} V = \emptyset \implies \exists H \subseteq G : \left(\bigcap_{V\in H} V = \emptyset\right) \wedge \Big(|H| \leq k\Big). \tag{2.16}$$

A Helly family of order 2 is sometimes said to have the **Helly property**.

Definition 2.4.6 (Diagonal). The diagonal of a set S is defined as follows:

$$\Delta_S := \big\{(a, a) \in S \times S \mid a \in S\big\}. \tag{2.17}$$

Definition 2.4.7 (Cover). A cover of a set S is a collection of sets $\mathcal{F} \subseteq P(S)$ such that

$$\bigcup_{V\in\mathcal{F}} V = S. \tag{2.18}$$

Definition 2.4.8 (Partition). A partition of X is a family of disjoint subsets $(A_i)_{i\in I} \subset P(X)$ such that $\bigcup_{i\in I} A_i = X$.

Definition 2.4.9 (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 2.4.10 (Filter). Let X be a partially ordered set. A family $\mathcal{F} \subseteq P(X)$ is a filter on X if it satisfies the following conditions:

1. **Empty set**: $\emptyset \notin \mathcal{F}$;
2. **Closed under intersections**: $\forall A, B \in \mathcal{F} : A \cap B \in \mathcal{F}$; and
3. **Closed under inclusion**: if $A \in \mathcal{F}$ and $A \subseteq B$, then $B \in \mathcal{F}$.

Definition 2.4.11 (Filtration). Consider a set A together with a collection of subsets F_iA indexed by a totally ordered set I. The collection is said to be a filtration of A if

$$i \leq j \implies F_iA \subseteq F_jA. \tag{2.19}$$

A filtration is said to be **exhaustive** if $\bigcup_i F_iA = A$ and **separated** if $\bigcap_i F_iA = \emptyset$.

Definition 2.4.12 (Associated grading). In the case where one can define quotient objects every filtration $\{F_iA\}_{i\in\mathbb{N}}$ of A defines an associated graded object $\{G_iA := F_iA/F_{i-1}A\}$.

2.5 Algebra of sets

Definition 2.5.1 (Algebra of sets). A collection $\mathcal{F} \subset P(X)$ is a called an algebra over X if it is closed under finite unions, finite intersections and complements. The pair $(X, \mathcal{F})$ is also called a **field of sets**.

Definition 2.5.2 (σ-algebra). A collection $\Sigma \subset P(X)$ is called a σ-algebra over a set X if it satisfies the following axioms:

1. **Total space**: $X \in \Sigma$,
2. **Closed under complements**: $\forall E \in \Sigma : E^c \in \Sigma$, and
3. **Closed under countable unions**: $\forall \{E_i\}_{i=1}^n \subset \Sigma : \bigcup_{i=1}^n E_i \in \Sigma$.

Remark 2.5.3. Axioms (2) and (3) together with de Morgan's laws (2.8) and (2.9) imply that a σ-algebra is also closed under countable intersections.

Corollary 2.5.4 (Algebra of sets). Every algebra of sets is a σ-algebra.

Property 2.5.5 (Intersections). The intersection of a family of σ-algebras is again a σ-algebra.

Definition 2.5.6 (Generated σ-algebras). A σ-algebra $\mathcal{G}$ is said to be generated by a collection of sets $\mathcal{A}$ if

$$\mathcal{G} = \bigcap \{\mathcal{F} \mid \mathcal{F} \text{ is a } \sigma\text{-algebra that contains } \mathcal{A}\}. \tag{2.20}$$

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

Notation 2.5.7. The σ-algebra generated by a collection of sets $\mathcal{A}$ is often denoted by $\mathcal{F}_\mathcal{A}$ or $\sigma(\mathcal{A})$.

Definition 2.5.8 (Product σ-algebras). The product σ-algebra $\mathcal{F}_1 \times \mathcal{F}_2$ on $X_1 \times X_2$ can be defined in the following equivalently ways:

- $\mathcal{F}$ is generated by the collection

$$\mathcal{C} = \{A_1 \times \Omega_2 \mid A_1 \in \mathcal{F}_1\} \cup \{\Omega_1 \times A_2 \mid A_2 \in \mathcal{F}_2\}.$$

- $\mathcal{F}$ is the smallest σ-algebra containing the products $A_1 \times A_2$ for all $A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2$.

Definition 2.5.9 (Monotone class). Let $\mathcal{A}$ be a collection of sets. $\mathcal{A}$ is called a monotone class if it has the following two properties:

1. For every increasing sequence $A_1 \subset A_2 \subset \cdots$:

$$\bigcup_{i=1}^\infty A_i \in \mathcal{A}.$$

2. For every decreasing sequence $A_1 \supset A_2 \supset \cdots$:

$$\bigcap_{i=1}^\infty A_i \in \mathcal{A}.$$

Theorem 2.5.10 (Monotone class theorem). *Let $\mathcal{A}$ be an algebra of sets 2.5.1. If $\mathcal{G}_\mathcal{A}$ is the smallest monotone class containing $\mathcal{A}$ then it coincides with the σ-algebra generated by $\mathcal{A}$.*

2.6 Ordered sets

2.6.1 Posets

Definition 2.6.1 (Preordered set). A set equipped with a reflexive and transitive binary relation.

Definition 2.6.2 (Partially ordered set). A set P equipped with a binary relation $\leq$ is called a partially ordered set (or **poset**) if the following 3 axioms are fulfilled for all elements $a, b, c \in P$:

1. **Reflexivity**: $a \leq a$,
2. **Antisymmetry**: $a \leq b \wedge b \leq a \implies a = b$, and
3. **Transitivity**: $a \leq b \wedge b \leq c \implies a \leq c$.

Equivalently, it is a preordered set for which the binary relation is also antisymmetric.

Definition 2.6.3 (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 2.6.4. Consider the power set of $\{0, 1\}$ with its canonical poset structure. A Hasse diagram of this poset is

$$
\begin{array}{ccccc}
 & & \{0,1\} & & \\
 & \diagup & & \diagdown & \\
\{0\} & & & & \{1\}. \\
 & \diagdown & & \diagup & \\
 & & \emptyset & &
\end{array}
$$

Definition 2.6.5 (Totally ordered set). A poset P with the property that for all $a, b \in P$: $a \leq b$ or $b \leq a$ is called a (nonstrict) totally ordered set. This property is called **totality**.

Definition 2.6.6 (Strict total order). A nonstrict order $\leq$ has an associated strict order $<$ that satisfies $a < b \iff a \leq b \wedge a \neq b$.

Definition 2.6.7 (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 2.6.8. 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 these two.

Definition 2.6.9 (Maximal element). An element m of a poset P such that for every $p \in P : m \leq p \implies m = p$.

Definition 2.6.10 (Chain). A totally ordered subset of a poset.

Theorem 2.6.11 (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 2.6.12 (Directed[4] set). A set X equipped with a preorder $\leq$ with the additional property that every 2-element subset has an upper bound, i.e. for every two elements $a, b \in X$, there exists an element $c \in X$ such that $a \leq c \wedge b \leq c$.

Definition 2.6.13 (Net). A net on a set X is a subset of X indexed by a directed set I.

2.6.2 Bounds

Definition 2.6.14 (Supremum). The supremum $\sup(X)$ of a poset X is the smallest upper bound of X.

Definition 2.6.15 (Infimum). The infimum $\inf(X)$ of a poset X is the greatest lower bound of X.

Definition 2.6.16 (Maximum). If $\sup(X) \in X$, the supremum is called the maximum of X. This is denoted by $\max(X)$.

Definition 2.6.17 (Minimum). If $\inf(X) \in X$, the supremum is called the minimum of X. This is denoted by $\min(X)$.

2.6.3 Ordinals and cardinals ♣

Definition 2.6.18 (Well-ordering). A **well-founded** linear order, i.e. a linear order $<$ such that every nonempty subset has a minimal element.

Definition 2.6.19 (Ordinal number). Consider the class of all well-ordered sets. An ordinal (number) 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 (rank) of all well-ordered sets isomorphic to it.

The most-used such construction is that by *Von Neumann*. For every well-ordered set W there exists an isomorphism $W \to P(W)$ that maps an element to the set of all subsets bounded from above by it. By analogy the Von Neumann ordinals are inductively defined as those well-ordered sets containing all smaller ordinals:

Definition 2.6.20 (Von Neumann ordinal). A set that is strictly well-ordered by membership and such that every element is also a subset.

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

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

[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 2.6.21 (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**.

Remark 2.6.22. The *Burali-Forti paradox* is the statement that the class of all ordinals (and by extension the class of all well-ordered sets) is not a set.

There also exist numbers representing the sizes of sets. These are called **cardinal numbers**. These ‘‘numbers’’ should satisfy the following conditions:

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

Guided by these conditions one could naively use the following definition:

Definition 2.6.23 (Cardinal number). 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 made:

Definition 2.6.24 (Cardinality). The cardinality of a set is the smallest ordinal rank of any well-order on it, i.e. any ordinal number bijective to it.[5] The cardinal numbers inherit a well-ordering from the ordinal numbers.

Remark 2.6.25 (Ordering). The Cantor-Bernstein-Schröder theorem 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 and the well-orderability of all sets, i.e. the well-ordering theorem, is equivalent to the axiom of choice.

Similar to ordinal numbers one can also define successors of cardinal numbers:

Definition 2.6.26 (Successor). Given a cardinal κ, one defines its successor κ^+ as the smallest cardinal larger than κ.

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

Definition 2.6.27 (Regular cardinal). An infinite cardinal κ such that there exist no set of cardinality κ that is the union of less than κ subsets of cardinality less than κ.

2.6.4 Lattices

Definition 2.6.28 (Semilattice). A poset $(P, \leq)$ for which every 2-element subset has a supremum (also called a **join**) in P is called a join-semillatice. Similarly, a poset $(P, \leq)$ for which every 2-element subset has an infimum (also called a **meet**) in P is called a meet-semilattice.

Notation 2.6.29. The join of $\{a, b\}$ is denoted by $a \wedge b$. The meet of $\{a, b\}$ is denoted by $a \vee b$.

Definition 2.6.30 (Lattice). A poset $(P, \leq)$ is called a lattice if it is both a join- and a meet-semilattice.

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

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

Alternative Definition 2.6.31 (Lattice). A lattice is an algebraic structure that admits operations $\wedge, \vee$ and constants $\top, \bot$ that satisfy the following axioms:

1. Both $\wedge$ and $\vee$ are idempotent, commutative and associative.
2. The **absorption laws**:
$$a \vee (a \wedge b) = a \qquad\qquad a \wedge (a \vee b) = a. \tag{2.21}$$
3. $\top$ and $\bot$ are the respective identities of $\wedge$ and $\vee$.

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

$$a \leq b \iff a \wedge b = a.$$

There exists an equivalent relation for the join.

Definition 2.6.32 (Bounded lattice). A lattice $(P, \leq)$ that contains a greatest element (denoted by $\top$ or 1) and a smallest element (denoted by $\bot$ or 0) such that

$$\bot \leq x \leq \top \tag{2.22}$$

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

$$x \wedge \top = x \qquad\qquad x \vee \bot = x. \tag{2.23}$$

Definition 2.6.33 (Frame). A complete lattice[6] $(P, \leq)$ for which the **infinite distributivity law** is satisfied:

$$y \wedge \left(\bigvee_{i \in I} x_i \right) = \bigvee_{i \in I} (y \wedge x_i). \tag{2.24}$$

Definition 2.6.34 (Heyting algebra). A bounded lattice H such that for every two elements $a, b \in H$ there exists a greatest element $x \in H$ for which

$$a \wedge x \leq b. \tag{2.25}$$

This element is denoted by $a \to b$. The **pseudo-complement** $\neg a$ of an element $a \in H$ is then defined as $a \to \bot$.

Definition 2.6.35 (Boolean algebra). A Boolean algebra X is a Heyting algebra in which the **law of excluded middle** holds:

$$\forall x \in X : \neg\neg x = x. \tag{2.26}$$

This can be equivalently stated as

$$\forall x \in X : x \vee \neg x = \top. \tag{2.27}$$

[6]When working with categories this has to be restricted to "all small joins/meets" or, equivalently, the index category should be a set.

2.7 Partitions

2.7.1 Partition

Definition 2.7.1 (Composition). Let $k, n \in \mathbb{N}$. A k-composition of n is a k-tuple $(t_1, \ldots, t_k)$ such that $\sum_{i=1}^{k} t_k = n$.

Definition 2.7.2 (Partition). Let $n \in \mathbb{N}$. A partition of n is an ordered composition of n. Hence multiple different composition can determine the same partition.

Definition 2.7.3 (Young diagram[7]). 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 a part of the partition:

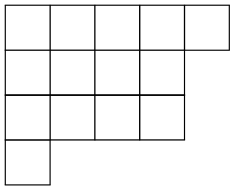


Figure 2.1: A Young diagram representing the partition $(5, 4, 4, 1)$ of 14.

Definition 2.7.4 (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 its main diagonal.

Example 2.7.5. Conjugating Diagram 2.1 gives Diagram 2.2 below. The associated partition is $(4, 3, 3, 3, 1)$.

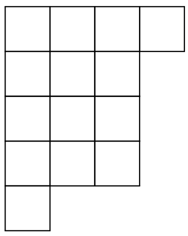


Figure 2.2: A Young diagram representing the partition $(4, 3, 3, 3, 1)$ of 14.

Definition 2.7.6 (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 2.7.7 (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 sum of all elements in $H_{i,j}$.

The number of all possible standard Young tableaux of shape λ (where λ defines a partition of n) is given by the following formula:

$$f^\lambda = \frac{n!}{\prod_{(i,j)\in\lambda} h_{i,j}}. \tag{2.28}$$

Definition 2.7.8 (Young tabloid). A Young tabloid of shape λ is defined as the equivalence class of Young tableaux that are connected by permuting the elements within a row. These are often drawn as in Figure 2.3.

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

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

Figure 2.3: A Young tabloid associated to the Young diagram in Figure 2.1.

2.7.2 Superpartition

For the physical background of the notions introduced in this section, see Chapter 54.

Definition 2.7.9 (Superpartition). Let $m, n \in \mathbb{N}$. A superpartition in the *m-fermion sector* is a sequence of integers of the following form:

$$\Lambda = (\Lambda_1, \dots, \Lambda_m; \Lambda_{m+1}, \dots, \Lambda_n), \tag{2.29}$$

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 themself. 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 2.7.10. A superpartition of degree 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 ended by a circle.

Chapter 3

Algebra

For the later sections on (co)homology theory some terminology from Chapter 5 is used.

3.1 Algebraic structures

Definition 3.1.1 (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 asssociative.

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

Definition 3.1.2 (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 contains an identity element with respect to $\star$.

Definition 3.1.3 (Nilpotent). An element x of a monoid for which there exists an integer $k \in \mathbb{N}$ such that $x^k = e$, where e is the identity element.

Property 3.1.4 (Eckmann-Hilton argument). Let $(M, \circ), (M, \otimes)$ be two monoid structures (or even unital magma structures) on a set M such that

$$(a \circ b) \otimes (c \circ d) = (a \otimes c) \circ (b \otimes d) \tag{3.1}$$

for all $a, b, c, d \in M$. The two monoid structures coincide and are in fact Abelian. (This property admits a vast generalization, see 4.5.1.)

Definition 3.1.5 (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 asssociative.
3. **Unitality**: G has an identity element with respect to $\star$.
4. **Inverses**: Every element in G has an inverse with respect to $\star$.

Notation 3.1.6 (Identity element). The identity element of a general group will be denoted by e. In certain cases, where this makes sense, the identity element will be denoted by 0 or 1 (additive and multiplicative conventions).

Definition 3.1.7 (Abelian group). Let $(G, \star)$ be a group. If $\star$ is commutative, then G is said to be an Abelian or **commutative** group.

Definition 3.1.8 (Morphism). A group **(homo)morphism** $\Phi : (G, \star) \to (H, \cdot)$ is a function $f : G \to H$ such that

$$\Phi(g \star g') = \Phi(g) \cdot \Phi(g') \tag{3.2}$$

for all $g, g' \in G$.

Definition 3.1.9 (Kernel). The kernel of a group morphism $\Phi : G \to H$ is defined as the set

$$\ker(\Phi) := \{g \in G \mid \Phi(g) = e_H\}. \tag{3.3}$$

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

Theorem 3.1.10 (First isomorphism theorem). *Consider a group morphism $\Phi : G \to H$. If Φ is surjective, then $G/\ker(\Phi) \cong H$.*

3.2 Group theory

Definition 3.2.1 (Order of a group). The number of elements in the group. It is denoted by $|G|$ or $\mathrm{ord}(G)$.

Definition 3.2.2 (Order of an element). The order of an element $a \in G$ is the smallest integer $n \in \mathbb{N}$ such that

$$a^n = e, \tag{3.4}$$

where e is the identity element of G.

Definition 3.2.3 (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 3.2.4 (Lagrange). *Let G be a finite group and let H be a subgroup. Then $|H|$ is a divisor of $|G|$.*

Corollary 3.2.5. The order of any element $g \in G$ is a divisor of $|G|$.

Construction 3.2.6 (Grothendieck completion). Let $(A, \boxplus)$ 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

$$(a_1, a_1') \sim (a_2, a_2') \iff \exists c \in A : a_1 \boxplus a_2' \boxplus c = a_1' \boxplus a_2 \boxplus c. \tag{3.5}$$

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 $\alpha \in \Delta_A$. In particular, $[(a, b)]$ is the additive inverse of $[(b, a)]$: $[(a, b)] + [(b, a)] = 0$.

Universal Property 3.2.7. Let $G(A)$ be the Grothendieck completion of A. Every monoid morphism $m : A \to B$, between an Abelian monoid and an Abelian group, factors uniquely through a group morphism $\varphi : G(A) \to B$.

Example 3.2.8 (Integers). The Grothendieck completion of the natural numbers is the additive group of integers $\mathbb{Z}$.

3.2.1 Cosets

Definition 3.2.9 (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

$$gH := \{gh \mid h \in H\}. \tag{3.6}$$

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

Definition 3.2.11 (Index). Consider a group G and a subgroup H. The index $|G:H|$ of H in G is defined as the number of cosets oh H in G. This can also be expressed as

$$|G| = |G:H||H|. \tag{3.7}$$

Alternative Definition 3.2.12 (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 3.2.13. If N is a normal subgroup of G, this is often denoted by $N \triangleleft G$.

Definition 3.2.14 (Quotient group). Let G be a group and let $N \triangleleft$ 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 3.2.15 (Center). The center of a group G is defined as follows:

$$Z(G) := \{z \in G \mid \forall g \in G : zg = gz\}. \tag{3.8}$$

This is a normal subgroup of G.

Definition 3.2.16 (Normalizer). The normalizer of a subgroup $H \subset G$ is defined as follows:

$$N_G(H) := \{g \in G \mid \forall gHg^{-1} = H\} \equiv \{g \in G \mid [G,H] \subseteq H\}. \tag{3.9}$$

Definition 3.2.17 (Centralizer). The centralizer of a subgroup $H \subset G$ is defined as follows:

$$C_G(H) := \{g \in G \mid \forall h \in H : ghg^{-1} = h\} \equiv \{g \in G \mid [G,H] = 0\}. \tag{3.10}$$

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

3.2.2 Sylow theorems

Definition 3.2.18 (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 resect to this property. The set of all Sylow p-subgroups of G is denoted by $\mathrm{Syl}_p(G)$.

Theorem 3.2.19 (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 3.2.20 (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 3.2.21 (Sylow III). *Consider a finite group G together with a prime factor p of $|G|$ of multiplicity n such that $|G| = p^n m$ for some $m \in \mathbb{N}$ and write $n_p := \mathrm{card}(\mathrm{Syl}_p(G))$.*

- $m = |G:H|$ *for every Sylow p-subgroup H of G.*
- $n_p|m$, *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.*

3.2.3 Abelian groups

Definition 3.2.22 (Commutator subgroup[1]). The commutator subgroup $[G, G]$ of G is defined as the group generated by the **commutators**

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

where $g, h \in G$. This group is a normal subgroup of G.

Property 3.2.23 (Abelianization). The Abelianization $G/[G, G]$ is an Abelian group. A group G is Abelian if and only if $[G, G]$ is trivial.

Property 3.2.24 (Abelian quotients). A quotient group G/H is Abelian if and only if $[G, G] \leq H$.

3.2.4 Symmetric groups

Definition 3.2.25 (Symmetric group). The symmetric group S_n (on n elements) is defined as the set of all permutations of $\{1, \ldots, n\}$. The number n is called the **degree** of the symmetric group. The symmetric group $\mathrm{Sym}(X)$ on a finite set X is defined similarly (by first numbering the elements and then acting by S_n)[2].

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

Theorem 3.2.26 (Cayley). *Every group is isomorphic to a subgroup of the permutation group* $\mathrm{Sym}(G)$.

Definition 3.2.27 (Cycle). A k-cycle is a permutation of the form $(a_1\ a_2\ \ldots\ a_k)$ 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 3.2.28 (Cycles are cyclic). Let τ be a k-cycle. Then τ is k-cyclic (hence the name):

$$\tau^k = e. \tag{3.11}$$

Example 3.2.29. Consider the set $\{1, 2, 3, 4, 5, 6\}$. The permutation $\sigma : x \mapsto x + 2 \bmod 6$ can be decomposed as $\sigma = (1\ 3\ 5)(2\ 4\ 6)$.

Definition 3.2.30 (Transposition). A permutation that exchanges two elements but leaves the other ones unchanged.

Property 3.2.31 (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 3.2.32 (Alternating group). The alternating group A_n is the subgroup of S_n containing all even permutations, i.e. those permutations that can be decomposed as an even number of transpositions.

[1] Also called the **derived subgroup**.

[2] Two such choices are related through conjugation by a unique permutation. The resulting groups are isomorphic.

Definition 3.2.33 (Shuffle). A permutation $\sigma \in S_n$ is called a (p,q)-shuffle (where $p+q=n$) if there exist disjoint increasing sequences $I = \{i_1 < \ldots < i_p\}$ and $J = \{j_1 < \ldots < j_q\}$ such that

$$\sigma(x) = \begin{cases} k & x = i_k \\ k+p & x = j_k. \end{cases} \tag{3.12}$$

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 $I = \{i_1 < \ldots < i_p\}$ and $J = \{j_1 < \ldots < j_q\}$ such

$$\tau(k) = \begin{cases} i_k & k \leq p \\ j_{k-p} & k > p. \end{cases} \tag{3.13}$$

3.2.5 Group presentations

Definition 3.2.34 (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 product of the elements g_i. These elements are then called generators of G.

Definition 3.2.35 (Relations). 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 3.2.36 (Complete set of relations). Let G be a group generated by a set of elements S (note that this set does not have be a group itself) 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 3.2.37 (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**.

It is clear that every 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.

Notation 3.2.38. The presentation of a group G is often denoted by $\langle S|R\rangle$, where S is the set of generators and R the set of relations.

3.2.6 Direct products

Definition 3.2.39 (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 $\cdot$ defined as

$$(g_1,h_1)\cdot(g_2,h_2) = (g_1g_2, h_1h_2), \tag{3.14}$$

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 one 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 $(g_1,\ldots,g_n)$ for $g_i \in G_i$, the g_i are called the **components** of g.

Definition 3.2.40 (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 3.2.41. 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 3.2.42 (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 satifies 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 3.2.43 (Normal subgroups). If both H and N in the above definition are normal, the inner semidirect product coincides with the direct product. In particular this includes the case of direct products. For a finite number of groups the direct product is generated by the elements of the groups.

If the subgroups H and N have presentations $\langle S_H|R_H\rangle$ and $\langle S_N|R_N\rangle$, the direct product is given by

$$H \times N = \langle S_H \cup S_N | R_H \cup R_N \cup R_C\rangle, \tag{3.15}$$

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

Definition 3.2.44 (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

$$(g_1, h_1) \cdot (g_2, h_2) = (g_1\varphi(h_1)(g_2), h_1h_2). \tag{3.16}$$

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 3.2.45 (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 3.2.46 (Bicrossed product of groups). Consider a group G with two subgroups $K, H \leq G$ such that every element $g \in G$ can be uniquely decomposed as a product of an element in K and an element in H. This implies that for $k \in K, h \in H$ there exists a unique decomposition of kh of the form

$$kh = (k \cdot h)k^h$$

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. 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
- $k \cdot (hh') = (k \cdot h)(k^h \cdot h')$.

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:

$$(h, k)(h', k') = \big(h(k \cdot h), k^{h'} k'\big). \tag{3.17}$$

3.2.7 Free groups

Definition 3.2.47 (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 3.2.37 can now be restated:

Alternative Definition 3.2.48 (Presentation). A group G is said to have a presentation $\langle S|R\rangle$ if it is isomorphic to the quotient of the free group on S by the normal subgroup generated by R.

Definition 3.2.49 (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 3.2.50 (Cardinality). For nontrivial groups the free product is always infinite.

Alternative Definition 3.2.51 (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 $\langle S_H|R_H\rangle$ respectively, the free product is given by

$$G * H = \langle S_G \cup S_H | R_G \cup R_H\rangle. \tag{3.18}$$

By comparing to (3.15) it can be seen that the free product is a generalization of the direct product.

Definition 3.2.52 (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 $G * H$:

$$\big\{\phi(f)\psi(f)^{-1} = e \,\big|\, f \in F\big\}. \tag{3.19}$$

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

$$G *_F H = (G * H)/N_F, \tag{3.20}$$

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 3.2.53 (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:

$$G=\left\{\sum_{i\in J}a_ig_i\,\middle|\,a_i\in\mathbb{Z},J\subseteq I\text{ is finite}\right\}. \tag{3.21}$$

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 3.2.54 (Nielsen-Schreier). Every subgroup of a free group is free.

Theorem 3.2.55 (Fundamental theorem of finitely generated groups). *Every finitely generated Abelian group G of rank n can either be decomposed as a quotient of two free, finitely generated, Abelian groups*

$$G=F/F' \tag{3.22}$$

or as a direct sum of a free, finitely generated, Abelian group and a torsion group 3.2.3:

$$G=A\oplus T\qquad\textit{with}\qquad T\equiv Z_{h_1}\oplus\cdots\oplus Z_{h_m}. \tag{3.23}$$

*In the second decomposition A has rank $n-m$ and all Z_{h_i} are cyclic groups of order h_i, where h_i is the power a prime. The group T is called the **torsion subgroup**.*

Property 3.2.56 (Uniqueness). The rank $n-m$ and the numbers h_i from previous theorem are unique.

3.3 Group actions

Definition 3.3.1 (Group action). Let G be a group 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 3.3.2. Note that this definition already makes sense for monoids 3.1.2.

Alternative Definition 3.3.3. A group action can equivalently be defined as a group morphism $\rho:G\to\mathrm{Sym}(X)$. 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 $\mathrm{Sym}(X)$ by $\mathrm{Aut}(X)$, i.e. the action of G should respect the structure.

Notation 3.3.4. The action $\rho(g,x)$ is often denoted by $g\cdot x$ or even gx if no confusion can arise.

Definition 3.3.5 (Equivariant map). Let X,Y be two G-spaces. An equivariant map between these sets is a function $f:X\to Y$ satisfying

$$g\cdot f(x)=f(g\cdot x), \tag{3.24}$$

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]

[4] In analogy with the basis of a vector space.
[5] G-spaces together with the G-maps constitute a category.

Example 3.3.6 (G-module). An Abelian group M equipped with a left group action $\varphi : G \to \mathrm{Aut}(M)$, i.e. an action that acts by group morphisms. Equivariant maps of G-modules are also called **intertwining maps** or **intertwiners**.

Definition 3.3.7 (Orbit). The orbit of an element $x \in X$ with respect to the action a group G is defined as the set

$$G \cdot x \equiv \{g \cdot x \mid g \in G\}. \tag{3.25}$$

The relation

$$p \sim q \iff \exists g \in G : p = g \cdot q$$

induces an equivalence relation 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 3.3.8 (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

$$G_x := \{g \in G \mid g \cdot x = x\}. \tag{3.26}$$

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

Theorem 3.3.9 (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:*

$$|G \cdot x||G_x| = |G|. \tag{3.27}$$

Definition 3.3.10 (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 groups of all elements is trivial.

Definition 3.3.11 (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 3.3.12 (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 3.3.13. 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 3.3.14 (Homogeneous space). A set equipped with a transitive group action.

Definition 3.3.15 (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 3.3.16 (Affine space). The n-dimensional affine space $\mathbb{A}^n$ is an $\mathbb{R}^n$-torsor.

Definition 3.3.17 (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 $G \to \mathrm{Aut}(H)$.

These data are required to satisfy two compatibility conditions:

1. **G-equivariance**:
$$t(\alpha(g)h) = gt(h)g^{-1}. \tag{3.28}$$
2. **Peiffer identity**:
$$\alpha(t(h))h' = hh'h^{-1}. \tag{3.29}$$

3.4 Group cohomology

Definition 3.4.1 (Group cohomology). Consider a group G together with a G-module A. Define the k^{th} **chain group** as

$$C^k(G; A) := \big\{\text{all set-theoretic functions from } G^k \text{ to } A\big\}. \tag{3.30}$$

The coboundary map $d^k : C^k(G; A) \to C^{k+1}(G; A)$ is defined as follows:

$$\begin{aligned} d^k f(g_1, \ldots, g_k, g_{k+1}) = g_1 \cdot f(g_2, \ldots, g_k, g_{k+1}) + (-1)^{k+1} f(g_1, \ldots, g_k) \\ + \sum_{i=1}^{k} (-1)^{i+1} f(g_1, \ldots, g_i g_{i+1}, \ldots, g_k, g_{k+1}). \end{aligned} \tag{3.31}$$

The cohomology groups are defined as the following quotient groups:

$$H^k(G; A) := \frac{\ker(d^k)}{\mathrm{im}(d^k)}. \tag{3.32}$$

?? COMPLETE (ADD e.g. classification of extensions) ??

3.5 Hochschild and cyclic cohomology

Definition 3.5.1 (Hochschild homology groups). Let A be an associative algebra and consider an A-bimodule M. For all $n \geq 0$ define

$$HC_n(A;M) := M \otimes A^{\otimes n}. \tag{3.33}$$

The boundary operator $d : HC_n(A;M) \to HC_{n-1}(A;M)$ is defined as follows:

$$\begin{aligned} d_n : m \otimes a_1 \otimes \cdots \otimes a_n \mapsto ma_1 \otimes \cdots \otimes a_n + \sum_{i=1}^{n-1} (-1)^i m \otimes a_1 \otimes \cdots \otimes a_i a_{i+1} \otimes \cdots \otimes a_n \\ + (-1)^n a_1 m \otimes \cdots \otimes a_n. \end{aligned} \tag{3.34}$$

The Hochschild homology is then given by the homology of this chain complex:

$$HH_\bullet(A;M) := \ker(d_n)/\operatorname{im}(d_{n+1}). \tag{3.35}$$

This complex can be turned into a graded-commutative algebra by equipping it with the *shuffle product*.

Remark 3.5.2 (Cohomology). Hochschild cohomology of A can be obtained by dualizing: $HC^n(A) := \mathrm{Hom}(HC_n(A), K)$, where K is the field over which A is defined.

Theorem 3.5.3 (Hochschild-Kostant-Rosenberg). *Let $A = C^\infty(M)$ for M a smooth manifold.*

$$HH_\bullet(A) \cong \Omega^n(M). \tag{3.36}$$

Definition 3.5.4 (Cyclic homology). The cyclic chain complex $CC_\bullet(A)$ is the subcomplex of the Hochschild complex $HC_\bullet(A)$ obtained as the kernel of $1 - \lambda$, where λ is the permutation operator

$$\lambda : HC_n(A) \to HC_n(A) : a_0 \otimes \cdots \otimes a_n \mapsto (-1)^n a_n \otimes a_0 \otimes \cdots \otimes a_{n-1}. \tag{3.37}$$

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 $CH_\bullet(A)$. By dualizing one obtains cyclic cohomology.

Example 3.5.5 (Smooth algebras). Consider the case of $A = C^\infty(M)$, where M is a smooth manifold. Then

$$CH_n(A) \cong Z^n_{\mathrm{dR}}(M) \oplus \bigoplus_{\substack{i=1\\ n-2i\geq 0}} H^{n-2i}_{\mathrm{dR}}(M). \tag{3.38}$$

For this reason cyclic cohomology will be used in noncommutative geometry to generalize de Rham cohomology.

3.6 Rings

Definition 3.6.1 (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 3.6.2 (Field). A ring $(R, +, \cdot)$ for which the monoid $(R\backslash\{1_+\}, \cdot)$ is an Abelian group and $1_+ \neq 1.$.

Definition 3.6.3 (Unit). An invertible element of a ring $(R, +, \cdot)$. The set of units forms a group under multiplication.

Definition 3.6.4 (Integral domain). A commutative ring R in which the product of two nonzero elements is again nonzero.

Definition 3.6.5 (Reduced ring). A ring that contains no nonzero nilpotents.

Construction 3.6.6 (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:

$$(r_1, s_1) \sim (r_2, s_2) \iff \exists t \in S : t(r_1 s_2 - r_2 s_1) = 0. \tag{3.39}$$

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

Remark 3.6.7. The localization of R with respect to the set S can be interpreted as the ring obtained by collapsing S into a single unit of R.

Notation 3.6.8. 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.

Definition 3.6.9 (Valuation). Let k be a field and let Γ be a totally ordered[6], Abelian group. 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. $\nu(ab) = \nu(a) + \nu(b)$; and
3. $\min(\nu(a), \nu(b)) \leq \nu(a + b)$, where the equality holds if $\nu(a) \neq \nu(b)$.

[6] Definition 2.6.5.

3.6.1 Ideals

Definition 3.6.10 (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 3.6.11 (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 2.6.10 of ideals.

Definition 3.6.12 (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 3.6.13 (Simple ring). A ring that has no nontrivial two-sided ideals. (Some authors require the ring to be Artinian.)

Definition 3.6.14 (Unit ideal). A ring considered as an ideal of itself.

Definition 3.6.15 (Proper ideal). An ideal that is not equal to the ring itself.

Definition 3.6.16 (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:

$$ab \in I \implies a \in I \vee b \in I. \tag{3.40}$$

Definition 3.6.17 (Maximal ideal). A proper ideal that is not contained in another proper ideal.

Property 3.6.18. Every maximal ideal is prime.

Definition 3.6.19 (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.

Construction 3.6.20 (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

$$I = \left\{ \sum_{i=1}^{n} l_i x_i r_i \,\middle|\, n \in \mathbb{N}, \forall i \leq n : l_i, r_i \in R \wedge x_i \in X \right\}. \tag{3.41}$$

Left and right ideals are generated in a similar fashion.

Notation 3.6.21. If the ideal I is generated by the elements $\{f_j\}_{j \in J}$ (for some index set J), it is often denoted by

$$I \equiv (f_1, f_2, \ldots). \tag{3.42}$$

Construction 3.6.22 (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 3.6.23 (Principal ideal). An ideal that is generated by a single element.

Definition 3.6.24 (Principal ideal domain). An integral domain 3.6.4 in which every ideal is principal.

Definition 3.6.25 (Local ring). A ring for which a unique, maximal, left ideal exists. This also implies that there exists a unique, maximal, right ideal and that these ideals coincide.

Property 3.6.26 (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 3.6.27 (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 3.6.28 (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.

3.6.2 Modules

Definition 3.6.29 (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 3.6.30 (Categorical perspective ♣). The definition of a ring can be defined more concisely in categorical terms. Recall the definition of a algebra over a monad 4.3.18. 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 3.6.31 (R-module). The above two formulations can be restated similar to that of group modules 3.3.6. Consider the Abelian group $(M, \oplus)$. Its endomorphism set $\mathrm{End}(M, \oplus)$ 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 $R \to \mathrm{End}(M, \oplus)$.

Definition 3.6.32 (Free module). An R-module M is said to be free if it 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 $\sum_{j\in J} r_j x_j$, where $J \subseteq I$ is finite.
2. the set $\{x_i\}_{i\in I}$ is linearly independent in the sense that

$$\sum_{j\in J\subseteq I} r_j x_j = 0 \implies \forall j \in J : r_j = 0. \tag{3.43}$$

Example 3.6.33 (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 3.6.34 (Division rings). For a general R-module the existence of a basis is not guaranteed unless R is a *division ring*. (See Construction 20.1.10 for more information.)

Corollary 3.6.35. Since every field is in particular a division ring, the existence of a basis follows from the above property for R-modules.

Definition 3.6.36 (Projective module). A module P is said to be projective if P can be expressed as

$$P \oplus M = F, \tag{3.44}$$

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

3.6.3 Semisimplicity

Definition 3.6.37 (Simple module). A module over a ring is said to be simple if it 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 3.6.38 (Jacobson radical). A ring is semisimple if and only if it is Artinian and if its Jacobson radical 3.6.19 vanishes.

Theorem 3.6.39 (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).*

3.7 Limits of algebraic structures

Definition 3.7.1 (Direct system). Let $(I, \leq)$ be a directed set 2.6.12 and let $\{A_i\}_{i\in I}$ be a family of algebraic objects (groups, rings, ...). Consider a collection of morphisms $\{f_{ij} : A_i \rightarrow A_j\}_{i,j\in I}$ 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 3.7.2 (Direct limit[7]). Consider a direct system (A_i, f_{ij}) over a directed set I. The direct limit A of this direct system is defined as follows:

$$\varinjlim A_i := \left. \bigsqcup_{i\in I} A_i \middle/ \sim \right. \tag{3.45}$$

where the equivalence relation is given by $x \in A_i \sim y \in A_j \iff \exists k \in I : f_{ik}(x) = f_{jk}(y)$. Informally put: two elements are equivalent if they eventually become the same.

The algebraic operations on A are defined such that the inclusion maps $\phi_i : A_i \rightarrow A$ are morphisms.

Definition 3.7.3 (Inverse system). Let $(I, \leq)$ be a directed set 2.6.12 and let $\{A_i\}_{i\in I}$ be a family of algebraic objects (groups, rings, ...). Consider a collection of morphisms $\{f_{ij} : A_j \rightarrow A_i\}_{i,j\in I}$ 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).

[7] Also called an **inductive limit**.

Definition 3.7.4 (Inverse limit[8]). Consider an inverse system (A_i, f_{ij}) over a directed set I. The inverse limit A of this inverse system is defined as follows:

$$\varprojlim A_k := \left\{ \vec{a} \in \prod_{i\in I} A_i \,\middle|\, a_i = f_{ij}(a_j), \forall i \leq j \right\}. \tag{3.46}$$

For all $i \in I$ there exists a natural projection $\pi_i : \varprojlim A_k \to A_i$.

Remark 3.7.5. 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.

3.8 Galois theory

Definition 3.8.1 (Field extension). Let k be a field. A field extension of k is a field K such that $k \subset K$ and such that the operations of k are the restrictions of those in K.

Notation 3.8.2. A field extension K of k is often denoted by K/k.

Definition 3.8.3 (Degree). If $K + k$ is a field extension, then K can be given the structure of a k-vector space 20.1.1. The dimension of this vector space is called the degree of the extension K. It is often denoted by $[K : k]$.

?? COMPLETE ??

[8] Also called a **projective limit**.

Chapter 4

Category theory

For the general theory of categories, the classical reference is [1]. The main reference for (co)end calculus is [2], while a thorough introduction to the theory of enrichment is given in [3]. For the theory of higher categories and its applications to topology and algebra, the reader is referred to the book by *Baez et al.* [4]. A good starting point for bicategories (and more) is the paper by *Leinster* [5].

4.1 Categories

Definition 4.1.1 (Category). A category $\mathbf{C}$ consists of two collections, the objects $\mathrm{ob}(\mathbf{C})$ and the morphisms $\hom(\mathbf{C})$ or $\mathrm{mor}(\mathbf{C})$, that satisfy the following conditions:

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

Remark 4.1.2. 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 4.1.3 (Subcategory). A subcategory is said to be **full** if for every two objects $x, y \in \mathrm{ob}(\mathbf{S})$:

$$\mathbf{S}(x, y) = \mathbf{C}(x, y). \tag{4.1}$$

A subcategory is said to be **wide** or **lluf** if it contains all objects, i.e. $\mathrm{ob}(\mathbf{S}) = \mathrm{ob}(\mathbf{C})$.

Definition 4.1.4 (Small category). A category $\mathbf{C}$ for which both $\mathrm{ob}(\mathbf{C})$ and $\hom(\mathbf{C})$ are sets. A category $\mathbf{C}$ is said to be locally small if for every two objects $x, y \in \mathrm{ob}(\mathbf{C})$ the collection of morphisms $\mathbf{C}(x, y)$ is a set. A category equivalent to a small category is said to be **essentially small**.

Definition 4.1.5 (Opposite category). Let $\mathbf{C}$ be a category. The opposite category $\mathbf{C}^{op}$ is constructed by reversing all arrows in $\mathbf{C}$, i.e. a morphism in $\mathbf{C}^{op}(x,y)$ is a morphism in $\mathbf{C}(y,x)$.

Property 4.1.6 (Involution). From the definition of the opposite category it readily follows that *op* is an involution:

$$(\mathbf{C}^{op})^{op} = \mathbf{C}. \tag{4.2}$$

4.2 Functors

Definition 4.2.1 (Covariant functor). Let $\mathbf{A}, \mathbf{B}$ be categories. A (covariant) functor is an assignment $F : \mathbf{A} \to \mathbf{B}$ satisfying the following conditions:

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

Property 4.2.2 (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 4.2.3 (Contravariant functor). Let $\mathbf{A}, \mathbf{B}$ be categories. A contravariant functor is an assignment $F : \mathbf{A} \to \mathbf{B}$ satisfying the following conditions:

1. F maps every object $x \in \mathrm{ob}(\mathbf{A})$ to an object $Fx \in \mathrm{ob}(\mathbf{B})$.
2. F maps every morphism $\phi \in \mathbf{A}(x,y)$ to a morphism $F\phi \in \mathbf{B}(Fy, Fx)$.
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 4.2.4 (Endofunctor). A functor of the form $F : \mathbf{C} \to \mathbf{C}$.

Definition 4.2.5 (Presheaf). A contravariant functor $G : \mathbf{C}^{op} \to \mathbf{Set}$. The collection of all presheaves on $\mathbf{C}$ forms a category $\mathbf{Psh}(\mathbf{C})$ (also denoted by $\widehat{\mathbf{C}}$).

Example 4.2.6 (Hom-functor). Let $\mathbf{C}$ be a locally small category. Every object $x \in \mathrm{ob}(\mathbf{C})$ induces a functor $h^x : \mathbf{C} \to \mathbf{Set}$ defined as follows:

- h^x maps every object $y \in \mathrm{ob}(\mathbf{C})$ to the set $\mathbf{C}(x,y)$.
- For all $y, z \in \mathrm{ob}(\mathbf{C})$, 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 4.2.7. The contravariant hom-functor h_x is defined by replacing $\mathbf{C}(x,-)$ with $\mathbf{C}(-,x)$ and replacing postcomposition with precomposition.

Definition 4.2.8 (Faithful functor). A functor $F : \mathbf{A} \to \mathbf{B}$ for which the map

$$\mathbf{A}(x,y) \to \mathbf{B}(Fx, Fy)$$

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

Definition 4.2.9 (Full functor). A functor $F : \mathbf{A} \to \mathbf{B}$ for which the map

$$\mathbf{A}(x, y) \to \mathbf{B}(Fx, Fy)$$

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

Definition 4.2.10 (Embedding). A fully faithful functor.

Definition 4.2.11 (Essentially surjective functor). A functor $F : \mathbf{A} \to \mathbf{B}$ such that for every object $y \in \mathrm{ob}(\mathbf{B})$, there exists an object $x \in \mathrm{ob}(\mathbf{A})$ with $Fx \cong y$.

Definition 4.2.12 (Profunctor[1]). A functor of the form $F : \mathbf{B}^{op} \times \mathbf{A} \to \mathbf{Set}$. Such a functor is often denoted by $F : \mathbf{A} \nrightarrow \mathbf{B}$.[2] Elements of the set $F(x, y)$ are sometimes called **heteromorphisms** (between x and y).

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

4.2.1 Natural transformations

Definition 4.2.13 (Natural transformation). Let $F, G : \mathbf{A} \to \mathbf{B}$ be two functors. A natural transformation $\psi : F \Rightarrow G$[3] consists of a collection of morphisms satisfying the following two conditions:

1. For every object $x \in \mathrm{ob}(\mathbf{A})$, there exists a morphism $\psi_x : Fx \to Gx$ in $\hom(\mathbf{B})$. 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 $f \in \mathbf{A}(x, y)$, 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 4.2.14 (Functor category). Consider two categories $\mathbf{A}, \mathbf{B}$ where $\mathbf{A}$ is small. The functors $F : \mathbf{A} \to \mathbf{B}$ form the objects of a category with the natural transformations as morphisms. This category is denoted by $[\mathbf{A}, \mathbf{B}]$ or $\mathbf{B}^{\mathbf{A}}$ (the latter is a generalization of 2.3.5).

Definition 4.2.15 (Dinatural transformation). Consider two profunctors $F, G : \mathbf{A} \nrightarrow \mathbf{A}$ or, more generally, two functors $F, G : \mathbf{A}^{op} \times \mathbf{A} \to \mathbf{B}$. A dinatural transformation is a family of morphisms

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

that make Diagram 4.1 commute for every morphism $f : y \to x$.

Definition 4.2.16 (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 $x \in \mathrm{ob}(\mathbf{C})$ such that F is naturally isomorphic to h^x. The pair $(x, \psi : F \Rightarrow h^x)$ is called a **representation** of F.

Theorem 4.2.17 (Yoneda lemma). *Let $\mathbf{C}$ be a locally small category and let $F : \mathbf{C} \to \mathbf{Set}$ be a functor. For every object $x \in \mathrm{ob}(\mathbf{C})$, there exists a natural isomorphism*[4]

$$\eta_x : \mathrm{Nat}(h^x, F) \to Fx : \psi \mapsto \psi_x(\mathbb{1}_x). \tag{4.3}$$

[1] Sometimes called a **distributor**.
[2] This is the convention by *Borceux*. Some other authors, such as [6], use the opposite convention.
[3] This notation is in analogy with the general notation for 2-morphisms. See Section 4.9 for more information.
[4] Here, the fact that $\mathrm{Nat}(h^-, -)$ can be seen as a functor $\mathbf{Set}^{\mathbf{C}} \times \mathbf{C} \to \mathbf{Set}$ is used.

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

Figure 4.1: Dinatural transformation.

Corollary 4.2.18 (Yoneda embedding). When F is another hom-functor h^y, the following result is obtained:

$$\mathrm{Nat}(h^x, h^y) \cong \mathbf{C}(y, x). \tag{4.4}$$

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 $x \in \mathrm{ob}(\mathbf{C})$ 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 $G : \mathbf{C}^{op} \to \mathbf{Set}^{\mathbf{C}}$. This way the Yoneda lemma can be seen to give rise to an embedding h^- of $\mathbf{C}^{op}$ in the functor category $\mathbf{Set}^{\mathbf{C}}$.

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

$$\mathcal{Y} := h_- : \mathbf{C} \hookrightarrow \widehat{\mathbf{C}}, \tag{4.5}$$

called the Yoneda embedding.

Definition 4.2.19 (Local object). Consider a collection of morphisms $S \subset \hom(\mathbf{C})$. An object $c \in \mathrm{ob}(\mathbf{C})$ is said to be S-local if the Yoneda embedding $\mathcal{Y}c$ maps morphisms in S to isomorphisms in $\mathbf{Set}$. A morphism $f \in \hom(\mathbf{C})$ is said to be S-local if its image under the Yoneda embedding of every S-local object is an isomorphism in $\mathbf{Set}$.

4.2.2 Equivalences

Definition 4.2.20 (Equivalence of categories). Two categories $\mathbf{A}, \mathbf{B}$ are said to be equivalent if there exist functors $F : \mathbf{A} \to \mathbf{B}$ and $G : \mathbf{B} \to \mathbf{A}$ 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 $\mathbf{A}, \mathbf{B}$ are said to be weakly equivalent if there exist functors $F : \mathbf{A} \to \mathbf{B}$ and $G : \mathbf{B} \to \mathbf{A}$ 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 4.2.21 (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.

Definition 4.2.22 (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 $\mathrm{Decat}:\mathbf{Cat}\rightarrow\mathbf{Set}$.

4.2.3 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.) ??

4.2.4 Adjunctions

Definition 4.2.23 (Hom-set adjunction). Let $F:\mathbf{A}\rightarrow\mathbf{B}$ and $G:\mathbf{B}\rightarrow\mathbf{A}$ be two functors. These functors form a (hom-set) adjunction $F\dashv G$ if the following isomorphism is natural in both x and y:

$$\Phi_{x,y}:\mathbf{B}(Fx,y)\cong\mathbf{A}(x,Gy). \tag{4.6}$$

The functor F (resp. G) is called the left (resp. right) adjoint and the image of a morphism under either of the natural isomoprhisms is called the adjunct of the other morphism.[5].

Notation 4.2.24. An adjunction $F\dashv G$ between categories $\mathbf{A},\mathbf{B}$ is often denoted by

$$\begin{tikzcd} \mathbf{B} \arrow[r, shift right, "G"'] \arrow[r, phantom, "\perp"] & \mathbf{A} \arrow[l, shift right, "F"'] \end{tikzcd}.$$

Definition 4.2.25 (Unit-counit adjunction). Let $F:\mathbf{A}\rightarrow\mathbf{B}$ and $G:\mathbf{B}\rightarrow\mathbf{A}$ be two functors. These functors form a unit-counit adjunction if there exist natural transformations

$$\varepsilon:F\circ G\Rightarrow\mathbb{1}_{\mathbf{B}} \tag{4.7}$$
$$\eta:\mathbb{1}_{\mathbf{A}}\Rightarrow G\circ F \tag{4.8}$$

such that the following compositions are identity morphisms:

$$F\xrightarrow{F\eta}FGF\xrightarrow{\varepsilon F}F \tag{4.9}$$
$$G\xrightarrow{\eta G}GFG\xrightarrow{G\varepsilon}G. \tag{4.10}$$

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.

[5] The terms "adjunct" and "adjoint" are sometimes used interchangeably (cf. French versus Latin).

Property 4.2.26 (Equivalence of the above definitions). Every hom-set adjunction induces a unit-counit adjunction. Let Φ be the natural isomorphism associated to the hom-set adjunction $F \dashv G$. The counit ε_y is obtained as the adjunct $\Phi^{-1}_{Gy,y}(\mathbb{1}_{Gy})$ of the identity morphism on $Gy \in \text{ob}(\mathbf{A})$, and the unit η_x is analogously defined as the adjunct $\Phi_{c,Fc}(\mathbb{1}_{Fx})$ of the identity morphism at $Fx \in \text{ob}(\mathbf{B})$.

Conversely, every unit-counit adjunction induces a hom-set adjunction. Consider a morphism $f : Fx \to y$. The (right) adjunct is defined as the composition

$$\widetilde{f} := Gf \circ \eta_x : x \to (G \circ F)x \to Gy.$$

To construct a (left) adjunct, consider a morphism $\widetilde{g} : x \to Gy$:

$$g := \varepsilon_y \circ F\tilde{g} : Fx \to (F \circ G)y \to y.$$

Definition 4.2.27 (Reflective subcategory). A full subcategory is said to be reflective (resp. coreflective) if the inclusion functor admits a left (resp. right) adjoint.

Property 4.2.28 (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.

4.3 General constructions

Definition 4.3.1 (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 4.3.2. The concept of a dagger structure allows the usual definition of **unitary** and **self-adjoint** morphisms, i.e. morphism satisfying

$$f^\dagger = f^{-1} \qquad \text{or} \qquad f^\dagger = f. \tag{4.11}$$

Definition 4.3.3 (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 $x \in \text{ob}(\mathbf{A}), y \in \text{ob}(\mathbf{B})$ and $\gamma : Fx \to Gy$.
- **Morphisms**: The morphisms $(x, y, \gamma) \to (k, l, \sigma)$ are pairs (f, g) with $f : x \to k \in \text{hom}(\mathbf{A})$ and $g : y \to l \in \text{hom}(\mathbf{B})$ such that $\sigma \circ Ff = Gg \circ \gamma$.
- Composition of morphisms is defined componentwise.

Definition 4.3.4 (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 4.3.5 (Functorial factorization). A *section* (see Definition 4.4.1) of the composition functor

$$\circ : [\mathbf{3}, \mathbf{C}] \to [\mathbf{2}, \mathbf{C}],$$

where $\mathbf{3}$ is the poset $\{0 \to 1 \to 2\}$.

Definition 4.3.6 (Slice category). Let $\mathbf{C}$ be a category and consider an object $x \in \text{ob}(\mathbf{C})$. The slice category $\mathbf{C}_{/x}$ of $\mathbf{C}$ over x is defined as follows:

- **Objects**: The morphisms in $\mathbf{C}$ with codomain x.
- **Morphisms**: The morphisms $f \to g$ are morphisms h in $\mathbf{C}$ such that $g \circ h = f$.

This category is also called the **over-category** of x. By dualizing one obtains the **under-category** of x.

4.3.1 Fibred categories ♣

Definition 4.3.7 (Fibre category). Let $\Pi : \mathbf{A} \to \mathbf{B}$ be a functor. The fibre category (of Π) over $y \in \mathrm{ob}(\mathbf{B})$ is the subcategory of $\mathbf{A}$ consisting of all objects $x \in \mathrm{ob}(\mathbf{A})$ such that $\Pi\, x = y$ and all morphisms $m \in \hom(\mathbf{A})$ such that $\Pi\, m = \mathbb{1}_y$. It will be denoted by $\mathbf{A}_y$.

Morphisms in $\mathbf{A}$ that are mapped to a morphism f in $\mathbf{B}$ are called **f-morphisms** and, in particular (using the identification of objects and their identity morphisms), morphisms in $\mathbf{A}_y$ are called **y-morphisms**. Similarly, **B-categories** are defined as the categories equipped with a (covariant) functor to $\mathbf{B}$. (It is not hard to see that these form a *2-category* under composition of functors that respects the $\mathbf{B}$-category structure.)

Definition 4.3.8 (Cartesian morphism). Consider a $\mathbf{B}$-category $\Pi : \mathbf{A} \to \mathbf{B}$. A morphism f in $\mathbf{A}$ 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 notion of stronger notion. A **strongly Cartesian morphism** is a morphism $f \in \hom(\mathbf{A})$ such that for every morphism $\varphi \in \hom(\mathbf{A})$ 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:

$$
\begin{tikzcd}
\forall x' \arrow[d, "\exists! g"'] \arrow[dr, "\forall\varphi"] & \\
x_1 \arrow[r, "f"'] & x_2
\end{tikzcd}
\qquad \overset{\Pi}{\mapsto} \qquad
\begin{tikzcd}
\Pi\, x' \arrow[d, "\forall\nu"'] \arrow[dr, "\Pi\varphi"] & \\
\Pi\, x_1 \arrow[r, "\Pi f"'] & \Pi\, x_2
\end{tikzcd}
$$

The diagram for (weak) Cartesian morphisms is obtained by identifying the objects $\Pi\, x'$ and $\Pi\, x_1$, 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 4.3.9 (Fibred category). A $\mathbf{B}$-category $\Pi : \mathbf{A} \to \mathbf{B}$ is called a fibred category or **Grothendieck fibration** if the following conditions are satisfied:

1. For each morphism in $\mathbf{B}$ whose codomain lies in the range of Π and each lift of this codomain to $\mathbf{A}$, 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 4.3.10 (Cleavage). Given a $\mathbf{B}$-category $\Pi : \mathbf{A} \to \mathbf{B}$, a cleavage is the choice of a Cartesian g-morphism $f : x \to y$ for every $y \in \mathrm{ob}(\mathbf{A})$ and morphism $g : b \to \Pi\, a'$. A $\mathbf{B}$-category equipped with a cleavage is said to be **cloven**.

It is clear that 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 4.3.11 (Discrete fibration). A functor $F : \mathbf{A} \to \mathbf{B}$ such that for every object $x \in \mathrm{ob}(\mathbf{A})$ and every morphism $f : y \to Fx$ in $\mathbf{B}$ there exists a unique morphism $g : z \to x$ in $\mathbf{A}$ such that $Fg = f$.

Example 4.3.12 (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 4.3.13 (Grothendieck construction ♣). Every fibred category $\Pi : \mathbf{A} \to \mathbf{B}$ defines a *pseudofunctor*[6] $F : \mathbf{B}^{op} \to \mathbf{Cat}$ which sends objects to fibre categories and arrows $f : c \to d$ to the pullback functor $f^* : \mathbf{A}_d \to \mathbf{A}_c$ constructed from a Cartesian lift of f. This pullback functor acts as follows:

- For every object $x \in \mathbf{A}_d$, f^*x is the domain of the Cartesian lift of f through x.
- For every morphism $(\alpha : x \to y) \in \mathbf{A}_d$ 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 $\int : [\mathbf{C}^{op}, \mathbf{Cat}] \to \mathbf{Cat}_{/\mathbf{C}}$ as follows. (These two constructions constitute a 2-equivalence of 2-categories.). Consider a pseudofunctor $F : \mathbf{C}^{op} \to \mathbf{Cat}$. The "bundle" $\int F$ consists of the following data:

- The objects are pairs (x, y) with $x \in \mathrm{ob}(\mathbf{C})$ and $y \in \mathrm{ob}(Fx)$.
- The morphisms $(x, y) \to (x', y')$ are pairs $(f : x \to x', \alpha : y \to Ff(y'))$.

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 4.3.14 (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.

Example 4.3.15 (Category of elements). The Grothendieck construction applied to an ordinary presheaf $F : \mathbf{C}^{op} \to \mathbf{Set}$.

4.3.2 Monads

Definition 4.3.16 (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:

$$\mu \circ T\mu = \mu \circ \mu_T. \tag{4.12}$$

[6] See Definition 4.9.9 towards the end of this chapter.

2. As natural transformations from T to itself:

$$\mu \circ T\eta = \mu \circ \eta_T = \mathbb{1}. \tag{4.13}$$

These conditions say that a monad is a monoid 3.1.2 in the category $\mathbf{End_C}$ of endofunctors on $\mathbf{C}$. Accordingly, η and μ are often called the **unit** and **multiplication** maps.

Example 4.3.17 (Adjunction). Every adjunction $F \dashv G$, with unit ε and counit η, induces a monad of the form $(GF, G\varepsilon F, \eta)$.

Definition 4.3.18 (Algebra over a monad[7]). Consider a monad (T, μ, η) on a category $\mathbf{C}$. An algebra over T 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**.

Definition 4.3.19 (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 4.3.20 (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 $\mathbf{C}_T(x, y) := \mathbf{C}(x, Ty)$.

Definition 4.3.21 (Monadic adjunction). An adjunction between categories $\mathbf{A}$ and $\mathbf{B}$ is said to be monadic if there exists an equivalence between B and the Eilenberg-Moore category of the induced monad.

Definition 4.3.22 (Monadic functor). A functor is said to be monadic if it admits a left adjoint such that the adjunction is monadic.

The following theorem characterizes monadic functors (for more information on some of the concepts, see Section 4.4 further below):

Theorem 4.3.23 (Beck's monadicity theorem). *Consider a functor $F : \mathbf{A} \to \mathbf{B}$. This functor is monadic if and only if the following conditions are satisfied:*

- *F admits a left adjoint.*
- *F reflects isomorphisms, i.e. all morphims in the preimage of an isomorphism are also isomorphisms.*
- *$\mathbf{A}$ has all coequalizers of F-split parallel pairs[8] and F preserves these coequalizers.*

Remark 4.3.24 (Crude monadicity theorem). A sufficient condition for monadicity is obtained by replacing the third condition above by the following weaker statement: "$\mathbf{A}$ has all coequalizers of reflexive pairs and F preserves these coequalizers."

[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 $\mathbf{End_C}$.

[8] These are parallel pairs f, g such that the images Ff, Fg under F admit a split coequalizer.

Definition 4.3.25 (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.

Given a closure operator $T : \mathbf{C} \to \mathbf{C}$, the object Tx is called the closure of $x \in \mathrm{ob}(\mathbf{C})$ and the associated morphism η_x is called the **closing map**. $x \in \mathrm{ob}(\mathbf{C})$ itself is said to be T**-closed** exactly if its closing map is an isomorphism.

An object $x \in \mathrm{ob}(\mathbf{C})$ is called a **modal type** if the unit $\eta_x : x \to Tx$ is an isomorphism.

Remark 4.3.26 (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 the 2-category **Cat**.

In the general setting one can then also define a **module** over a monad. First of all, one can regard any object $x \in \mathrm{ob}(\mathbf{C})$ as a functor from the terminal category 1. One can then replace 1 by any other category in the ordinary definition to obtain a general algebra (or module) over a given monad. 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**.

4.4 Morphisms and diagrams

4.4.1 Morphisms

Definition 4.4.1 (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 4.4.2 (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 $z \in \mathrm{ob}(\mathbf{C})$ 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 4.4.3 (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 $z \in \mathrm{ob}(\mathbf{C})$ 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 $\alpha_1 = \alpha_2$.

Definition 4.4.4 (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 4.4.5 (Balanced category). A category in which every monic epi is an isomorphism.

Definition 4.4.6 (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 4.4.7 (Subobject). Let $\mathbf{C}$ be a category and let $x \in \mathrm{ob}(\mathbf{C})$ 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 in the category $\mathbf{C}$ is an isomorphism class of monos $i : y \hookrightarrow x$ in the slice category $\mathbf{C}/x$.

Definition 4.4.8 (Well-powered category). A category $\mathbf{C}$ is said to be well-powered if for every object $x \in \mathrm{ob}(\mathbf{C})$ the class of subobjects $\mathrm{Sub}(x)$ is small.

4.4.2 Initial and terminal objects

Definition 4.4.9 (Initial object). An object $\emptyset$ such that for every other object x there exists a unique morphism $\iota_x : \emptyset \to x$.

Definition 4.4.10 (Terminal object). An object 1 such that for every other object x there exists a unique morphism $\tau_x : x \to 1$.

Property 4.4.11 (Uniqueness). If an initial (or terminal) object exists, it is unique (up to isomorphisms).

Definition 4.4.12 (Zero object). An object that is both initial and terminal. The zero object is often denoted by 0.

Property 4.4.13 (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 4.4.14 (Pointed category). A category containing a zero object.

Definition 4.4.15 (Global element). Let $\mathbf{C}$ be a category with a terminal object 1. A global element of an object $x \in \mathrm{ob}(\mathbf{C})$ is a morphism $1 \to x$.

Property 4.4.16. Every global element is monic.

Definition 4.4.17 (Pointed object). An object x equipped with a global element $1 \to x$. This morphism is sometimes called the **basepoint**.

Remark 4.4.18. 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 4.4.19 (Generalized element). Let $\mathbf{C}$ be category and consider an object $x \in \mathrm{ob}(\mathbf{C})$. For any object $y \in \mathrm{ob}(\mathbf{C})$, a morphism $y \to x$ is called a generalized element of x. They are also called y**-elements** in x or elements of **shape** y in x.

Definition 4.4.20 (Generator). Let $\mathbf{C}$ be a category. A collection of objects $\mathcal{O} \subset \mathrm{ob}(\mathbf{C})$ 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 $\mathbf{C}$:

$$\forall x, y \in \mathrm{ob}(\mathbf{C}), \forall f, g \in \mathbf{C}(x, y) : \Big(f \neq g \implies \exists o \in \mathcal{O}, \exists h \in \mathbf{C}(o, x) : f \circ h \neq g \circ h\Big). \tag{4.14}$$

Definition 4.4.21 (Well-pointed category). A category for which the terminal object is a generator.

4.4.3 Lifts

Definition 4.4.22 (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 4.4.23 (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 4.4.24 (Injective and projective morphisms). Consider a class of morphisms $I \subseteq \hom(\mathbf{C})$. A morphism $f \in \hom(\mathbf{C})$ 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 $\mathrm{rlp}(I)$ and $\mathrm{llp}(I)$, respectively.

Definition 4.4.25 (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 Figure 4.2.)

$$\begin{tikzcd} x \arrow[r, "g"] \arrow[d, "\forall f \in I"'] & i \\ y \arrow[ur, dashed, "\exists\psi"'] & \end{tikzcd}$$

(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 4.2: 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 4.4.26 (Fibrations and cofibrations). Consider a category $\mathbf{C}$ together with a class $I \subseteq \hom(\mathbf{C})$ of morphisms. A morphism $f \in \hom(\mathbf{C})$ 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.

4.4.4 Limits and colimits

Definition 4.4.27 (Diagram). A diagram in $\mathbf{C}$ with index category $\mathbf{I}$ is a (covariant) functor $D : \mathbf{I} \to \mathbf{C}$.

Definition 4.4.28 (Cone). Let $D : \mathbf{I} \to \mathbf{C}$ be a diagram. A cone from $c \in \mathrm{ob}(\mathbf{C})$ 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 $f : i \to j \in \hom(\mathbf{I})$. This is depicted in Figure 4.3a.

$$\begin{tikzcd} & c \arrow[dl, "\psi_i"'] \arrow[dr, "\psi_{f(i)}"] & \\ Di \arrow[rr, "Df"'] & & Df(i) \end{tikzcd}$$

(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 4.3: Category of cones.

Alternative Definition 4.4.29. The above definition can be reformulated by defining an additional functor $\Delta_x : \mathbf{I} \rightarrow \mathbf{C}$ that maps every element $i \in \text{ob}(\mathbf{I})$ to x and every morphism $g \in \text{hom}(\mathbf{I})$ to $\mathbb{1}_x$, i.e. $\Delta : C \rightarrow [\mathbf{I}, \mathbf{C}]$ 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 4.4.30 (Morphism of cones). Let $D : \mathbf{I} \rightarrow \mathbf{C}$ 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 $f : x \rightarrow y$ such that the diagrams of the form 4.3b commute for all $i \in \text{ob}(\mathbf{I})$. 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 4.4.31 (Limit). Consider a diagram $D : \mathbf{I} \rightarrow \mathbf{C}$. The limit of this diagram, denoted by $\lim D$, is (if it exists) the terminal object of the category $\mathbf{Cone}(D)$.

Remark. In the older literature the name **projective limit** was sometimes used. The dual notion, a **colimit**, is often called an **inductive limit** in the older literature.

This definition leads to the following universal property:

Universal Property 4.4.32. Let $D : \mathbf{I} \rightarrow \mathbf{C}$ be a diagram. For every cone $(x, \psi) \in \mathbf{Cone}(D)$, there exists a unique morphism $f : x \rightarrow \lim D$. This defines a bijection

$$[\mathbf{I}, \mathbf{C}](\Delta_x, D) \cong \mathbf{C}(x, \lim D).$$

If all (small) limits exist, the limit functor $\lim : [\mathbf{I}, \mathbf{C}] \rightarrow \mathbf{C}$ can be defined. The universal property of limits then implies that it is right adjoint to the constant functor Δ.

For diagrams in **Set** one can use the fully faithfulness of the Yoneda embedding to obtain the following expression:

$$\lim D \cong [\mathbf{I}, \mathbf{Set}](\Delta_*, D). \tag{4.15}$$

Remark 4.4.33. In Section 4.7 on enriched category theory, a generalization (the so-called *weighted limits*) of the above construction will be given that is better suited to the enriched setting and allows to express a wide variety of constructions as (weighted) limits.

Example 4.4.34 (Terminal object). The terminal object 1 is the limit of the empty diagram.

Definition 4.4.35 (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 4.4.36 (Presheaf categories). All presheaf categories are both complete and cocomplete.

Definition 4.4.37 (Continuous functor). A functor that preserves all small limits.

Example 4.4.38 (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

$$\mathbf{C}(x, \lim D) \cong \lim \mathbf{C}(x, D). \tag{4.16}$$

In the case where $\mathbf{C}$ is small, one can characterize the Yoneda embedding through a universal property:

Universal Property 4.4.39 (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 4.4.40 (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 4.4.41 (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, 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 4.4.42 (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 4.4.43 (Directed limit). Consider a diagram $D : \mathbf{I} \to \mathbf{C}$. The limit (resp. colimit) of D is said to be codirected (resp. directed) if $\mathbf{I}$ is a downward (resp. upward) directed set 2.6.12.

The following definition is a categorification of the previous one:

Definition 4.4.44 (Filtered limit). Consider a diagram $D : \mathbf{I} \to \mathbf{C}$. The limit (resp. colimit) of D is said to be cofiltered (resp. filtered) if $\mathbf{I}$ is a cofiltered (resp. filtered) category.

Property 4.4.45. A category has all directed colimits if and only if it has all filtered colimits. (A dual statement holds for limits.)

Definition 4.4.46 (Pro-object). A functor $F : \mathbf{I} \to \mathbf{C}$ where $\mathbf{I}$ is a small cofiltered category. The names stems from the fact that one can interpret pro-objects as formal cofiltered (projective) limits.

Definition 4.4.47 (Compact object). An object for which the covariant hom-functor preserves all filtered colimits. These objects are also said to be **finitely presentable**.[11]

[9]Epimorphisms are characterized by a *pushout* (see 4.4.61 further below).

[10]A generalization in the context of enriched categories is given by the *Karoubi envelope*.

[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).

Definition 4.4.48 (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 $\mathbf{C}$.

Definition 4.4.49 (Equalizer). Consider a diagram of the form

$$x \overset{f}{\underset{g}{\rightrightarrows}} y.$$

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

$$e \xrightarrow{\varepsilon} x \overset{f}{\underset{g}{\rightrightarrows}} y \tag{4.17}$$

is universal with respect to (e, ε). By dualizing one obtains **cofork** diagrams $x \rightrightarrows y \to z$ and their universal versions, the **coequalizers**.

Definition 4.4.50 (Split coequalizer). A cofork diagram

$$x \overset{f}{\underset{g}{\rightrightarrows}} y \xrightarrow{\tau} z$$

together with a section φ of f and a section σ of τ such that $\sigma \circ \tau = g \circ \varphi$.

Definition 4.4.51 (Regular morphisms). A mono (resp. epi) is said to be regular if it arises as an equalizer (resp. coequalizer) of two parallel morphisms.

Property 4.4.52 (Regular bimorphism). Both monic regular epimorphisms and epic regular monomorphisms are isomorphisms.

Alternative Definition 4.4.53 (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 4.4.54 (Span). A span in a category C is a diagram of the form 4.4a. 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.

$$\begin{tikzcd} & s \arrow[dl, "f"'] \arrow[dr, "g"] & \\ x & & y \end{tikzcd}$$

(a) Span (category theory).

$$\begin{tikzcd} x \arrow[dr, "f"'] & & y \arrow[dl, "g"] \\ & c & \end{tikzcd}$$

(b) Cospan.

Figure 4.4: (Co)span diagrams.

Definition 4.4.55 (Pullback). The pullback or **fibre product** of two morphisms $f : x \to z$ and $g : y \to z$ is defined as the limit of cospan 4.4b. The full diagram characterizing the pullback, which has the form of a square, is sometimes called a **Cartesian square**.

Notation 4.4.56 (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 Figure 4.5a.

Property 4.4.57 (Product). If a terminal object 1 exists, the pullback $x \times_1 y$ is equal to the product $x \times y$.

$$\begin{tikzcd}
x \times_z y \arrow[r] \arrow[d] \arrow[dr, phantom, "\mathrm{pb}"] & y \arrow[d, "g"] \\
x \arrow[r, "f"'] & z
\end{tikzcd}$$

(a) Pullback square.

$$\begin{tikzcd}
z \arrow[r, "g"] \arrow[d, "f"'] \arrow[dr, phantom, "\mathrm{po}"] & y \arrow[d] \\
x \arrow[r] & x \sqcup_z y
\end{tikzcd}$$

(b) Pushout square.

Figure 4.5: Pullback and pushout diagrams.

Definition 4.4.58 (Kernel pair). Consider a morphism $f : x \to y$. Its kernel pair is defined as the pullback of f along itself.

Definition 4.4.59 (Pushout). The dual notion of a pullback, i.e. the colimit of a span. See Figure 4.5b.

Property 4.4.60. Pullbacks preserve monos and pushouts preserve epis.

Alternative Definition 4.4.61 (Epimorphism). A morphism whose cokernel pair is the identity.

Property 4.4.62 (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 4.4.63 (Wedge). Consider a profunctor $F : \mathbf{C} \nrightarrow \mathbf{C}$. A wedge $e : w \to F$ is an object $w \in \mathrm{ob}(\mathbf{Set})$ 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 4.4.64 (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 $\prod_{x \in \mathrm{ob}(\mathbf{C})} F(x,x)$.

This is equivalent to a definition in terms of equalizers. Consider the two canonical maps

$$\prod_{x \in \mathrm{ob}(\mathbf{C})} \mathbf{C}(x,x) \rightrightarrows \prod_{f : x \to y} \mathbf{C}(x,y).$$

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 $f \in \hom(\mathbf{C})$.

Notation 4.4.65 (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 4.4.66 (Natural transformations). Consider two functors $F, G : \mathbf{A} \to \mathbf{B}$. The map $(x,y) \mapsto \mathbf{B}(Fx, Gy)$ gives a profunctor $H : \mathbf{A} \nrightarrow \mathbf{A}$. By looking at the wedge condition for this profunctor, the following equality for all morphisms $f : x \to y$ can be derived:

$$\tau_y \circ Ff = Gf \circ \tau_x, \tag{4.18}$$

where τ is the wedge projection. Comparing this equality to Definition 4.2.13 gives

$$\mathrm{Nat}(F,G) = \int_{x\in\mathbf{A}} \mathbf{B}(Fx, Gx). \tag{4.19}$$

Property 4.4.67. Using the continuity 4.4.37 of the hom-functor, one can prove the following equality which can be used to turn ends into coends and vice versa:

$$\mathbf{Set}\left(\int^{x\in\mathbf{C}} F(x,x), y\right) = \int_{x\in\mathbf{C}} \mathbf{Set}(F(x,x), y). \tag{4.20}$$

Using the above properties and definitions, one obtains the following two statements, called the **Yoneda reduction** and **co-Yoneda lemma**:

Property 4.4.68 (Ninja Yoneda lemma). Let $F : \mathbf{A} \to \mathbf{B}$ be a covariant functor (similar statements hold for contravariant functors).

$$\int_{x\in\mathbf{A}} \mathbf{Set}(\mathbf{A}(-,x), Fx) \cong F \tag{4.21}$$

$$\int^{x\in\mathbf{A}} \mathbf{A}(x,-) \times Fx \cong F. \tag{4.22}$$

For a generalization to the enriched setting see Definition 4.7.16.

Remark 4.4.69. A common remark at this point is the comparison with the Dirac distribution (17.15):

$$\int \delta(x-y) f(x) = f(y). \tag{4.23}$$

By interpreting the functor F as a function, the representable functors can be seen to behave as Dirac distributions.

Property 4.4.70.

$$\int_{F\in\mathbf{coPsh}(\mathbf{C})} \mathbf{Set}(Fx, Fy) \cong \mathbf{C}(x,y) \tag{4.24}$$

Definition 4.4.71 (Category of elements). Consider a presheaf $F : \mathbf{C}^{op} \to \mathbf{Set}$. Its category of elements $\mathrm{El}(F)$ is defined as the comma category $(\mathcal{Y} \downarrow !_F)$, where $!_F : * \to [\mathbf{C}^{op}, \mathbf{Set}]$ sends the unique object to F itself. Equivalently, it is the category with objects the pairs $(c,x) \in \mathrm{ob}(\mathbf{C}) \times Fc$ and morphisms $f \in \mathbf{C}(c,c')$ such that $c = Ff(c')$.

This category comes equipped with a canonical forgetful functor

$$\mathbf{C}_F : \mathrm{El}(F) \to \mathbf{C} : (c,x) \mapsto c. \tag{4.25}$$

Remark 4.4.72. The category of elements is usually 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).

Definition 4.4.73 (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 $\mathrm{Ran}_G F : \mathbf{C} \to \mathbf{B}$ and natural transformation $\eta : \mathrm{Ran}_G F \circ G \Rightarrow F$:

$$\begin{tikzcd}
\mathbf{C} \arrow[dr, dashed, "\mathrm{Ran}_G F"] & \\
\mathbf{A} \arrow[u, "G"] \arrow[r, "F"'] \arrow[ur, phantom, "\Downarrow\eta"] & \mathbf{B}
\end{tikzcd}$$

The left Kan extension $\mathrm{Lan}_G F$ is obtained by dualizing this construction.

Property 4.4.74 (Complete categories). Complete (resp. cocomplete) categories admit all right (resp. left) Kan extensions.

Definition 4.4.75 (Preservation of Kan extension). A Kan extension $\mathrm{Lan}_G F$ 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**.

Alternative Definition 4.4.76 (Kan extension). The construction above gives a functor Ran_G from the functor category $[\mathbf{A}, \mathbf{B}]$ to the functor category $[\mathbf{C}, \mathbf{B}]$. The right Kan extension Ran_G can be defined as the right adjoint to the pullback functor $G^* : F \mapsto F \circ G$. Similarly, the left Kan extension can be defined as the left adjoint to the pullback functor.

In the spirit of partial adjoints or partial limits, this definition can be used to define **local Kan extensions**. Although the left (or right) Kan extension functors do not have to exists globally, the extension of a single functor could still exist. This local version is defined by the following natural isomorphism (here given for a left extension):

$$[\mathbf{A}, \mathbf{B}](F, G^* -) \cong [\mathbf{C}, \mathbf{B}](\mathrm{Lan}_G F, -). \tag{4.26}$$

Remark 4.4.77. Using this equivalence of hom-spaces, Kan extensions can be generalized from **Cat** to any *2-category*.

Example 4.4.78 (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 4.4.32:

$$\lim F \cong \mathrm{Ran}_{!_C} F. \tag{4.27}$$

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 4.4.79 (Adjoint functors). A functor $F : \mathbf{A} \to \mathbf{B}$ admits a left (resp. right) adjoint if and only if the right (resp. left) Kan extension of the identity functor $\mathbb{1} : \mathbf{A} \to \mathbf{A}$ along F exists. If it exists as an absolute extension, the left adjoint is given exactly by this Kan extension.

Definition 4.4.80 (Codensity monad). Consider a general functor $F : \mathbf{A} \to \mathbf{B}$. If the right Kan extension $\mathrm{Ran}_F F$ 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*.

4.5 Internal structures

Property 4.5.1 (Eckmann-Hilton argument). A monoid internal to **Mon**, the category of monoids is the same as a commutative monoid. (See also Property 3.1.4.)

Definition 4.5.2 (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 $C_0 \in \mathrm{ob}(\mathcal{E})$ of objects;
- an object $C_1 \in \mathrm{ob}(\mathcal{E})$ 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

$$s \circ e = \mathbb{1}_{C_0} \qquad\qquad t \circ e = \mathbb{1}_{C_0};$$

  and

- a composition morphism $c : C_1 \times_{C_0} C_1 \to C_1$ such that the following equations hold:

$$\begin{gathered} s \circ c = s \circ \pi_1 \qquad\qquad t \circ c = t \circ \pi_2 \\ \pi_1 = c \circ (e \times_{C_0} \mathbb{1}) \qquad\qquad c \circ (\mathbb{1} \times_{C_0} e) = \pi_2 \\ c \circ (c \times_{C_0} \mathbb{1}) = c \circ (\mathbb{1} \times_{C_0} c), \end{gathered}$$

  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 4.5.3. 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 4.5.4. The above definition can be reformulated in a very elegant way. An internal category in $\mathcal{E}$ is a monad in the bicategory $\mathbf{Span}(\mathcal{E})$ of spans in $\mathcal{E}$ as shown in Figure 4.6.

Functors between internal categories are not the only relevant morphisms. However, when defining (co)presheafs 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 4.5.5 (Internal diagram). A left module over a monad in $\mathbf{Span}(\mathcal{E})$. The dual notion is better known as an **internal presheaf**.

In fact, this is a specific instance of an even more general concept (for more information on the definitions and applications see [1, 6]):

[12] Codense functors are usually defined in a different way, but one can show that this is an equivalent definition (hence the name).

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

$$\begin{tikzcd} & C_1 \times_{C_0} C_1 \arrow[dl, "s \circ \pi_0"'] \arrow[dr, "t \circ \pi_1"] \arrow[d, "c"] & \\ C_0 & C_1 \arrow[l, "s"] \arrow[r, "t"'] & C_0 \end{tikzcd}$$

Unit gives identity

$$\begin{tikzcd} & C_0 \arrow[dl, "\mathbb{1}_a"'] \arrow[dr, "\mathbb{1}_a"] \arrow[d, "e"] & \\ C_0 & C_1 \arrow[l, "s"] \arrow[r, "t"'] & C_0 \end{tikzcd}$$

Figure 4.6: Internal category as a monad in $\mathbf{Span}(\mathcal{E})$.

Definition 4.5.6 (Internal profunctor). A bimodule between monads in $\mathbf{Span}(\mathcal{E})$. Together with the above definitions this gives rise to an equivalence $\mathbf{Mod}(\mathbf{Span}(\mathcal{E})) \cong \mathbf{Prof}(\mathcal{E})$.

Construction 4.5.7 (Internal Yoneda profunctor). Consider an internal functor $F : \mathbf{A} \to \mathbf{B}$. This functor induces two internal profunctors $F_* : \mathbf{B} \nrightarrow \mathbf{A}$ and $F^* : \mathbf{A} \nrightarrow \mathbf{B}$:

For F_* the object span is defined as (the profunctor F^* is defined similarly)

$$A_0 \xleftarrow{\pi_0} A_0 \times_{B_0} B_1 \xrightarrow{t \circ \pi_1} B_0.$$

The action of $f \in B_1$ is given by postcomposition with f in the second factor, while the action of $g \in A_1$ 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 $\mathbb{1}_\mathbf{C}$ 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 $\mathcal{Y}(\mathbf{C})$.

4.6 Monoidal categories

Definition 4.6.1 (Monoidal category). A category $\mathbf{C}$ equipped with a bifunctor

$$- \otimes - : \mathbf{C} \times \mathbf{C} \to \mathbf{C}$$

called the **tensor product** or **monoidal product**, a distinct object $\mathbf{1}$ called the **unit object**, 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**: $\rho_x : x \otimes \mathbf{1} \cong x$.

$$\begin{tikzcd}
(x \otimes \mathbf{1}) \otimes y \arrow[rr, "\alpha_{x,\mathbf{1},y}"] \arrow[dr, "\rho_x \otimes \mathbb{1}_y"'] & & x \otimes (\mathbf{1} \otimes y) \arrow[dl, "\mathbb{1}_x \otimes \lambda_y"] \\
& x \otimes y &
\end{tikzcd}$$

Figure 4.7: Triangle diagram.

$$\begin{tikzcd}
& ((w \otimes x) \otimes y) \otimes z \arrow[r, "\alpha_{w,x,y} \otimes \mathbb{1}_z"] \arrow[dl, "\alpha_{w \otimes x,y,z}"'] & (w \otimes (x \otimes y)) \otimes z \arrow[dr, "\alpha_{w,x \otimes y,z}"] & \\
(w \otimes x) \otimes (y \otimes z) \arrow[drr, "\alpha_{w,x,y \otimes z}"'] & & & w \otimes ((x \otimes y) \otimes z) \arrow[dl, "\mathbb{1}_w \otimes \alpha_{x,y,z}"] \\
& & w \otimes (x \otimes (y \otimes z)) &
\end{tikzcd}$$

Figure 4.8: Pentagon diagram.

These natural transformations are required make the **triangle** and **pentagon** diagrams 4.7 and 4.8 commute.

A monoidal category for which the associator and the unitors are identity transformations is often said to be **strict**.

Example 4.6.2 (Cartesian category). A monoidal category where the monoidal product is given by the ordinary product 4.4.48.

Definition 4.6.3 (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.

Property 4.6.4. Every scalar $s : \mathbf{1} \to \mathbf{1}$ induces a natural transformation $s : \mathbb{1}_{\mathbf{C}} \Rightarrow \mathbb{1}_{\mathbf{C}}$ with components

$$s_x : x \cong \mathbf{1} \otimes x \xrightarrow{s \otimes \mathbb{1}_x} \mathbf{1} \otimes x \cong x.$$

For every morphism $f \in \hom(\mathbf{C})$, 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 4.6.5 (Weak inverse). Let $(\mathbf{C}, \otimes, \mathbf{1})$ be a monoidal category and consider an object $x \in \mathrm{ob}(\mathbf{C})$. An object $y \in \mathrm{ob}(\mathbf{C})$ is called a weak inverse of x if it satisfies $x \otimes y \cong \mathbf{1}$.

Remark 4.6.6. One can show that the existence of a one-sided weak inverse (as in the definition above) is sufficient to prove that it is in fact a two-sided weak inverse, i.e. $y \otimes x \cong \mathbf{1}$ also holds.

Theorem 4.6.7 (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.*

4.6.1 Braided categories

Definition 4.6.8 (Braided monoidal category). A monoidal category $(\mathbf{C}, \otimes, \mathbf{1})$ equipped with a natural isomorphism

$$\sigma_{x,y} : x \otimes y \cong y \otimes x$$

that makes the two **hexagon** diagrams 4.9a and 4.9b 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 4.9: Hexagon diagram.

Property 4.6.9 (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 $x, y, z \in \mathrm{ob}(\mathbf{C})$:

$$(\sigma_{y,z} \otimes \mathbb{1}_x) \circ (\mathbb{1}_y \otimes \sigma_{x,z}) \circ (\sigma_{x,y} \otimes \mathbb{1}_z) = (\mathbb{1}_z \otimes \sigma_{x,y}) \circ (\sigma_{x,z} \otimes \mathbb{1}_y) \circ (\mathbb{1}_x \otimes \sigma_{y,z}). \tag{4.28}$$

Remark 4.6.10. 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 4.6.11 (Symmetric monoidal category). A braided monoidal category where the braiding σ satisfies

$$\sigma_{x,y} \circ \sigma_{y,x} = \mathbb{1}_{x\otimes y} \,. \tag{4.29}$$

In Chapter 27 the theory of monoidal categories is continued.

4.6.2 Monoidal functors

Definition 4.6.12 (Monoidal functor). Let $(\mathbf{A}, \otimes, \mathbf{1_A}), (\mathbf{B}, \circledast, \mathbf{1_B})$ 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 Diagram 4.10 commute.
2. An isomorphism $\phi : \mathbf{1_B} \to F\mathbf{1_A}$ that makes the two diagrams in Figure 4.11 commute.

$$\begin{tikzcd}
(Fx \circledast Fy) \circledast Fz \arrow[r, "\alpha_{\mathbf{B}}"] \arrow[d, "\psi_{x,y} \circledast \mathbb{1}_{Fz}"'] & Fx \circledast (Fy \circledast Fz) \arrow[d, "\mathbb{1}_{Fx} \circledast \psi_{y,z}"] \\
F(x \otimes y) \circledast Fz \arrow[d, "\psi_{x \otimes y, z}"'] & Fx \circledast F(y \otimes z) \arrow[d, "\psi_{ax, y \otimes z}"] \\
F\Big((x \otimes y) \otimes z\Big) \arrow[r, "F\alpha_{\mathbf{A}}"'] & F\Big(x \otimes (y \otimes z)\Big)
\end{tikzcd}$$

Figure 4.10: Monoidal functor.

$$\begin{tikzcd}
Fx \circledast \mathbf{1_B} \arrow[r, "\mathbb{1}_{Fx} \circledast \phi"] \arrow[d, "\rho_{\mathbf{B}}"'] & Fx \circledast F\mathbf{1_A} \arrow[d, "\psi_{x,\mathbf{1_A}}"] \\
Fx & F(x \otimes \mathbf{1_A}) \arrow[l, "F\rho_{\mathbf{A}}"]
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathbf{1_B} \circledast Fy \arrow[r, "\phi \circledast \mathbb{1}_{Fy}"] \arrow[d, "\lambda_{\mathbf{B}}"'] & F\mathbf{1_A} \circledast Fy \arrow[d, "\psi_{\mathbf{1_A},y}"] \\
Fy & F(\mathbf{1_A} \otimes y) \arrow[l, "F\lambda_{\mathbf{A}}"]
\end{tikzcd}$$

Figure 4.11: Unitality diagrams.

Remark 4.6.13. The maps ψ and ϕ are also called **coherence maps** or **structure morphisms**.

Property 4.6.14 (Canonical unit). For every monoidal functor F there exists a canonical isomorphism $\phi : \mathbf{1_B} \to F\mathbf{1_A}$ defined by the commutative Diagram 4.12.

$$\begin{tikzcd}
\mathbf{1_B} \circledast F\mathbf{1_A} \arrow[r, "\lambda_{\mathbf{B}}"] \arrow[d, "\phi \circledast \mathbb{1}_{F\mathbf{1_A}}"'] & F\mathbf{1_A} \arrow[d, "F\lambda_{\mathbf{A}}"] \\
F\mathbf{1_A} \circledast F\mathbf{1_A} \arrow[r, "\psi_{\mathbf{1_A},\mathbf{1_A}}"'] & F(\mathbf{1_A} \otimes \mathbf{1_A})
\end{tikzcd}$$

Figure 4.12: Canonical unit isomorphism.

Definition 4.6.15 (Lax monoidal functor). A monoidal functor for which the coherence maps are merely morphisms and not isomorphisms.

Definition 4.6.16 (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 Figure 4.13 commute.

Definition 4.6.17 (Monoidal equivalence). An equivalence of monoidal categories consisting of monoidal functors and monoidal natural isomorphisms.

Theorem 4.6.18 (MacLane's strictness theorem). *Every monoidal category is monoidally equivalent to a strict monoidal category.*

4.6.3 Closed categories

Definition 4.6.19 (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

$$
\begin{tikzcd}
 & \mathbf{1_B} \arrow[dl, "\phi_F"'] \arrow[dr, "\phi_G"] & \\
F\mathbf{1_A} \arrow[rr, "\eta_{\mathbf{1_A}}"'] & & G\mathbf{1_A}
\end{tikzcd}
\qquad
\begin{tikzcd}
Fx \circledast Fy \arrow[r, "\psi_{a,b}"] \arrow[d, "\eta_a \circledast \eta_b"'] & F(x \otimes y) \arrow[d, "\eta_{a\otimes b}"] \\
Gx \circledast Gy \arrow[r, "\widetilde{\psi}_{a,b}"'] & G(x \otimes y)
\end{tikzcd}
$$

Figure 4.13: Monoidal natural transformation.

$x \to (y \to z)$. The internal hom-functor $\underline{\mathrm{Hom}}$ is defined by the following natural isomorphism:

$$\mathrm{Hom}(x \otimes y, z) \cong \mathrm{Hom}(x, \underline{\mathrm{Hom}}(y, z)). \tag{4.30}$$

The existence of all internal homs is equivalent to the existence of a right adjoint to the tensor functor.

Notation 4.6.20. The internal hom $\underline{\mathrm{Hom}}(x, y)$ is also often denoted by $[x, y]$. From now on this convention will be followed (unless otherwise specified).

Definition 4.6.21 (Closed monoidal category). A monoidal category is said to be closed monoidal if it has 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**. A locally Cartesian closed category with a terminal object is also Cartesian closed.

Definition 4.6.22 (Exponential object). In the case of Cartesian (monoidal) categories, the internal hom $\underline{\mathrm{Hom}}(x, y)$ 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 confusion with the notation for *2-morphisms*.

Definition 4.6.23 (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 4.6.24 (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

$$L(y \times Rx) \to Ly \times x \tag{4.31}$$

is a natural isomorphism.

Property 4.6.25 (Global elements). The following isomorphism is natural in both $x, y \in \mathrm{ob}(\mathbf{M})$:

$$\mathbf{M}(\mathbf{1}, [x, y]) \cong \mathbf{M}(x, y). \tag{4.32}$$

It is this relation that gives the best explanation for the term "internal hom". One also immediately obtains the following natural isomorphism:

$$\mathbf{M}(x, [\mathbf{1}, y]) \cong \mathbf{M}(x, y). \tag{4.33}$$

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 4.6.26 (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

$$[x \otimes y, z] \cong [x, [y, z]]. \tag{4.34}$$

Furthermore, if $\mathbf{M}$ is also symmetric, there exists an internal isomorphism of the form

$$[x, [y, z]] \cong [y, [x, z]]. \tag{4.35}$$

Definition 4.6.27 (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

$$[Lx, y] \cong [x, Ry]. \tag{4.36}$$

Property 4.6.25 above implies that every strong adjunction is in particular an adjunction in the sense of Section 4.2.4.

4.7 Enriched category theory

The following definition is due to *Bénabou*. It should represent the "ideal place in which to do category theory".

Definition 4.7.1 (Cosmos). A complete and cocomplete closed symmetric monoidal category.

Definition 4.7.2 (Enriched category). Let $(\mathcal{V}, \otimes, \mathbf{1})$ be a monoidal category. A $\mathcal{V}$-enriched category, also called a $\mathcal{V}$-category[13], consists of the following elements:

- a collection of objects $\mathrm{ob}(\mathbf{C})$, and
- for every pair of objects $x, y \in \mathrm{ob}(\mathbf{C})$, an object $\mathbf{C}(x, y) \in \mathrm{ob}(\mathcal{V})$ for which the following morphisms exist:
  1. $\mathrm{id}_x : \mathbf{1} \to \mathbf{C}(x, x)$ 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 4.7.3 (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 4.7.4 (Underlying category). Given a $\mathcal{V}$-enriched category $\mathbf{C}$, the underlying category $\mathbf{C}_0$ is defined as follows:

- **Objects**: $\mathrm{ob}(\mathbf{C})$
- **Morphisms**: $\mathcal{V}(\mathbf{1}, \mathbf{C}(x, y))$,

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 $\mathcal{I}(*, *) \equiv \mathbf{1}$.

Property 4.7.5 ($\mathcal{V}$ 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 4.6.25 then implies that there exists an isomorphism between the underlying category $\widetilde{\mathcal{V}}_0$ and the original category $\mathcal{V}$.

[13]Not to be confused with the notation for fibre categories 4.3.7.

$$\begin{tikzcd}
& \mathbf{A}(x,y) \arrow[dl, "\lambda^{-1}"'] \arrow[dr, "\rho^{-1}"] & \\
\mathbf{1} \otimes \mathbf{A}(x,y) \arrow[d, "\eta_y \otimes F_{x,y}"'] & & \mathbf{A}(x,y) \otimes \mathbf{1} \arrow[d, "G_{x,y} \otimes \eta_x"] \\
\mathbf{B}(Fy,Gy) \otimes \mathbf{B}(Fx,Fy) \arrow[dr, "\circ"'] & & \mathbf{B}(Gx,Gy) \otimes \mathbf{B}(Fx,Gx) \arrow[dl, "\circ"] \\
& \mathbf{B}(Fx,Gy) &
\end{tikzcd}$$

Figure 4.14: $\mathcal{V}$-naturality diagram.

Given two $\mathcal{V}$-enriched categories, one can define suitable functors between them:

Definition 4.7.6 (Enriched functor). A $\mathcal{V}$-enriched functor $F : \mathbf{A} \to \mathbf{B}$ consists of the following data:

- a function $F_0 : \mathrm{ob}(\mathbf{A}) \to \mathrm{ob}(\mathbf{B})$ (as for ordinary functors), and
- for every two objects $x, y \in \mathrm{ob}(\mathbf{A})$, 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 (4.19) using enriched ends, one obtains a definition of enriched natural transformations and, therefore, also a definition of enriched functor categories.:

$$[\mathbf{A}, \mathbf{B}](F, G) := \int_{x \in \mathbf{A}} \mathbf{B}(Fx, Gx). \tag{4.37}$$

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 4.2.13:

Definition 4.7.7 (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 4.14.

The question then becomes how these two definitions are related. The end (4.37) 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, i.e. an element of $\mathcal{V}(\mathbf{1}, [\mathbf{A}, \mathbf{B}](F, G))$, 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.

4.7.1 Enriched constructions

Definition 4.7.8 (Functor tensor product). Consider a covariant functor $G : \mathbf{C} \to \mathcal{V}$ and a contravariant functor $F : \mathbf{C}^{op} \to \mathcal{V}$ 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:

$$F \otimes_{\mathbf{C}} G := \int^{x \in \mathbf{C}} Fx \otimes Gx. \tag{4.38}$$

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 4.7.9 (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:

$$F \otimes_{\text{Day}} G := \iint^{x,y\in\mathbf{C}} \mathbf{C}(x \otimes y, -) \otimes Fx \otimes Gy. \tag{4.39}$$

Property 4.7.10 (Monoidal structure). In the case where $\mathbf{M}$ is a closed symmetric monoidal category, the 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 4.7.11 (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:

$$\mathbf{C}(v \cdot x, y) \cong [v, \mathbf{C}(x, y)], \tag{4.40}$$

where the bracket $[-,-]$ on the right-hand side denotes the internal hom in $\mathcal{V}$. Dually, the power (or cotensor) functor $[-,-] : \mathcal{V} \times \mathbf{C} \to \mathbf{C}$ is defined by the following natural isomorphism:

$$\mathbf{C}(x, [v, y]) \cong [v, \mathbf{C}(x, y)], \tag{4.41}$$

where the bracket $[-,-]$ on the right-hand side again denotes the internal hom in $\mathcal{V}$. If an enriched category admits all (co)powers, it is said to be **(co)powered** (over its enriching category).

Remark 4.7.12. Equation (4.35) 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 4.7.13 (Disjoint unions). Every (co)complete (locally) small category $\mathbf{C}$ admits the structure of a $\mathbf{Set}$-(co)powered category:

$$x^S := \prod_{s\in S} x \tag{4.42}$$

$$S \cdot x := \bigsqcup_{s\in S} x. \tag{4.43}$$

The definition and properties of internal hom-functors and (co)powers can be formalized as follows:

Definition 4.7.14 (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:

- $- \otimes - : \mathbf{A} \times \mathbf{B} \to \mathbf{C}$,
- $\hom_L : \mathbf{A}^{op} \times \mathbf{C} \to \mathbf{B}$, and
- $\hom_R : \mathbf{B}^{op} \times \mathbf{C} \to \mathbf{A}$

admitting the following natural isomorphisms:

$$\mathbf{C}(x \otimes y, z) \cong \mathbf{A}(x, \hom_R(y, z)) \cong \mathbf{B}(y, \hom_L(x, z)). \tag{4.44}$$

It should be noted that fixing any of the variables gives rise to ordinary adjunctions in the sense of Section 4.2.4.

Property 4.7.15 (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 $\mathbf{C}^{op} \times \mathbf{C} \to \mathcal{V}$ is the right adjoint in an enriched two-variable adjunction. The power and copower functors are then given by the other two adjoints.

The following definition constructs Kan extensions in the enriched setting (these can be shown to reduce to 4.4.73 when enriching over $\mathbf{Set}$):

Alternative Definition 4.7.16 (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 $\mathcal{V}$, one can define the left Kan extension of $F : \mathbf{A} \to \mathbf{B}$ along $G : \mathbf{A} \to \mathbf{C}$ as a coend:

$$\mathrm{Lan}_G F := \int^{x \in \mathbf{A}} \mathbf{C}(Gx, -) \cdot Fx. \tag{4.45}$$

If $\mathbf{B}$ is assumed to be powered over $\mathcal{V}$, one can define the right Kan extension as an end:

$$\mathrm{Ran}_G F := \int_{x \in \mathbf{A}} [\mathbf{C}(-, Gx), Fx]. \tag{4.46}$$

Remark 4.7.17. 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 4.4.68.

Property 4.7.18. Kan extensions computed using (co)ends as above are pointwise in the sense of Definition 4.4.75.

Alternative Definition 4.7.19 (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 $F : \mathbf{A}^{op} \to \mathcal{V}$. The tensor product 4.7.8 can be generalized whenever $\mathbf{B}$ is copowered over $\mathcal{V}$:

$$F \otimes_{\mathbf{A}} G := \int^{x \in \mathbf{A}} Fx \cdot Gx. \tag{4.47}$$

4.7.2 Weighted (co)limits

In this section the definition of ordinary limits and, in particular, the defining universal property 4.4.32 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. On 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 $\mathbf{C}(-, D-) : \mathbf{I} \to [\mathbf{C}^{op}, \mathbf{Set}]$. Since presheaf categories are complete (Example 4.4.36), the limit of this diagram exists:

$$\mathbf{Set}(S, \lim \mathbf{C}(x, D-)) \cong [\mathbf{I}, \mathbf{Set}](\Delta_S, \mathbf{C}(x, D-)).$$

By restricting to the terminal set $S = *$, one obtains

$$\lim \mathbf{C}(x, D-) \cong [\mathbf{I}, \mathbf{Set}](\Delta_*, \mathbf{C}(x, D-)).$$

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.

$$[\mathbf{I}, \mathbf{Set}](\Delta_*, \mathbf{C}(x, D-)) \cong \mathbf{C}(x, \lim D). \tag{4.48}$$

?? CLEAN THIS UP (note that continuity and pointwise definition was already mentioned for ordinary limits) ??

Definition 4.7.20 (Weighted limit). This definition 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 4.7.21. A motivation for this construction is the following. As was already pointed out in Remark 4.4.18, 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 4.7.22 (Enriched weighted limits). By expressing the natural transformations as an end as in Equation (4.19) and by using the canonical powering in **Set**, one can express ordinary weighted limits as follows:

$$\lim^W D \cong \int_{i \in \mathbf{I}} [Wi, Di]. \tag{4.49}$$

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:

$$\lim^W D := \int_{i \in \mathbf{I}} [Wi, Di]. \tag{4.50}$$

In a similar way one can define weighted colimits in copowered $\mathcal{V}$-categories as coends:

$$\operatorname{colim}^W D := \int^{i \in \mathbf{I}} Wi \cdot Di. \tag{4.51}$$

Here, the weight functor W is required to be contravariant since colimits (and cocones in general) are natural transformations between contravariant functors.

Property 4.7.23 (Weighted limits are Homs). In the case $\mathbf{C} = \mathcal{V}$, the powering functor 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 4.7.24 (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:

$$\mathrm{colim}^{\mathcal{Y}i} D \cong Di. \tag{4.52}$$

A similar statement for weighted limits can be obtained with the covariant Yoneda embedding.

Alternative Definition 4.7.25 (Weighted (co)limits). The above property can be used to axiomatize small weighted (co)limits in bicomplete categories:

1. **Yoneda**: For every object $i \in \mathrm{ob}(\mathbf{I})$ there exist isomorphisms
$$\lim^{\mathbf{I}(i,-)} D \cong Di \qquad \text{and} \qquad \mathrm{colim}^{\mathbf{I}(-,i)} D \cong Di. \tag{4.53}$$
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 expression 4.7.16):

Property 4.7.26 (Kan extensions). Consider functors $F : \mathbf{A} \to \mathbf{B}$ and $G : \mathbf{A} \to \mathbf{C}$. If for every $x \in \mathrm{ob}(\mathbf{C})$ 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 $\mathrm{Ran}_G F$. The left Kan extension can be obtained as a weighted colimit.

Property 4.7.27 (Category of elements). The weighted (co)limits of a functor (over **Set**) can also be expressed in terms of the category of elements 4.4.71 of the weight:

$$\lim^W F \cong \lim F \circ \mathbf{C}_W, \tag{4.54}$$

where the limit on the right-hand side is a conical limit.

4.8 Abelian categories

4.8.1 Additive and Abelian categories

Definition 4.8.1 (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 4.8.2. Let **A** be a pre-additive category. The following statements are equivalent for an object $x \in \mathrm{ob}(\mathbf{A})$:

- 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 4.4.12.

Property 4.8.3 (Biproducts). In a pre-additive category the following isomorphism holds for all finitely indexed sets $\{x_i\}_{i\in I}$:

$$\prod_{i\in I} x_i \cong \coprod_{i\in I} x_i. \tag{4.55}$$

Finite (co)products in pre-additive categories are often called **direct sums**. In general, if a product and coproduct exist and are equal, one also speaks of a **biproduct**.

Definition 4.8.4 (Additive category). A pre-additive category in which all finite products exist.

When working with additive categories, it is generally assumed that the associated functors are of a specific type:

Definition 4.8.5 (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 as a group morphism on hom-spaces.

Definition 4.8.6 (Grothendieck group). Let $\mathbf{A}$ be an additive category and consider its decategorification 4.2.22. This set carries the structure of an Abelian monoid and, hence, the Grothendieck construction 3.2.6 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 some classical notions from (homological) algebra such as images and kernels:

Definition 4.8.7 (Kernel). Let $f : x \to y$ be a morphism. A[14] kernel of f is a morphism $k : z \to x$ such that:

1. $f \circ k = 0$.
2. **Universal property**: Every morphism $k' : z' \to x$ such that $f \circ k' = 0$ factors uniquely through k.

This implies that a kernel of f could equivalently be defined as the equalizer of f and 0.

Notation 4.8.8 (Kernel). If the kernel of $f : x \to y$ exists, it is denoted by $\ker(f)$.

Definition 4.8.9 (Cokernel). Let $f : x \to y$ be a morphism. A cokernel of f is a morphism $p : y \to z$ such that:

1. $p \circ f = 0$.
2. **Universal property**: Every morphism $p' : y \to z'$ such that $p' \circ f = 0$ factors uniquely through p.

This implies that a cokernel of f could equivalently be defined as the coequalizer of f and 0.

Notation 4.8.10 (Cokernel). If the cokernel of $f : x \to y$ exists, it is denoted by $\mathrm{coker}(f)$.

Remark 4.8.11. The name and notation of the kernel and the cokernel (in the categorical sense) is explained by remarking that $\ker(f)$ represents the functor

$$F : z \mapsto \ker\Big(\mathbf{C}(z, x) \to \mathbf{C}(z, y)\Big),$$

where ker denotes the algebraic kernel 3.1.9, and similarly for he cokernel.

Definition 4.8.12 (Pseudo-Abelian category). An additive category in which every projection/idempotent has a kernel.

[14]Note the word "a". The kernel of a morphism is only determined up to an isomorphism.

Definition 4.8.13 (Pre-Abelian category). An additive category in which every morphism has a kernel and cokernel.

Definition 4.8.14 (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

$$\mathrm{coker}(\ker(f)) \cong \ker(\mathrm{coker}(f)). \tag{4.56}$$

Property 4.8.15 (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.

Example 4.8.16 (k-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, the subscript is often left implicit.)

4.8.2 Exact functors

Definition 4.8.17 (Exact functor). Let $F : \mathbf{A} \to \mathbf{A}'$ be an additive functor between additive categories.

- F is said to be left-exact if it preserves kernels.
- F is said to be right-exact if it preserves cokernels.
- F is said to be exact if it is both left- and right-exact.

Corollary 4.8.18. 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 4.8.19 (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 $\mathbf{Mod}_{R^{op}}$ 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 4.8.20 (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 4.8.21 (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 $R\mathbf{Mod} \to S\mathbf{Mod}$.*

4.8.3 Finiteness

Definition 4.8.22 (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 4.8.23 (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 4.8.24 (Jordan-Hölder series). A filtration

$$0 \longrightarrow x_1 \longrightarrow x_2 \longrightarrow \cdots \longrightarrow x_n = x$$

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 4.8.25 (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 4.8.26 (Krull-Schmidt). *Any object in an Abelian category of finite length admits a unique decomposition as a direct sum of indecomposable objects*[15]*.*

Definition 4.8.27 (Locally finite). A 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 4.8.28 (Finite). A 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*.
3. The set of isomorphism classes of simple objects is finite.

Theorem 4.8.29 (Schur's lemma). *Let* $\mathbf{A}$ *be an Abelian category. For every two simple objects* x, y*, 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*.*

Corollary 4.8.30. If $\mathbf{A}$ is locally finite and k is algebraically closed, then $\mathbf{A}(x, x) \cong k$ for all simple objects x. 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 4.8.20 can be adapted to the finite linear case as follows:

Theorem 4.8.31 (Deligne). *Every finite* k*-linear Abelian category is* k*-linearly equivalent to a category of the form* $A\mathbf{Mod}^{\mathrm{fin}}$ *for* A *a finite-dimensional* k*-algebra.*

Construction 4.8.32 (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 k-algebras A, B. The category $A\mathbf{Mod}^{\mathrm{fin}} \boxtimes B\mathbf{Mod}^{\mathrm{fin}}$ is equivalent to the category $A \otimes_k B\mathbf{Mod}^{\mathrm{fin}}$.

[15] An object is **indecomposable** if it cannot be written as a direct sum of its subobjects.

4.9 Higher category theory ♣

4.9.1 n-categories

Definition 4.9.1 (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 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 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 visible.

Example 4.9.2. The classical examples of a 1-category and 2-category are **Set** and **Cat**, respectively.

Property 4.9.3 (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
$$\beta \circ \alpha : f' \circ f \Rightarrow g' \circ g.$$
- **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
$$\beta \cdot \alpha : f \Rightarrow h.$$

As a consistency condition the horizontal and vertical composition are required to satisfy the following **interchange law**:

$$(\alpha \cdot \beta) \circ (\gamma \cdot \delta) = (\alpha \circ \gamma) \cdot (\beta \circ \delta). \tag{4.57}$$

Definition 4.9.4 ((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 4.9.5 (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 4.2.25 can be generalized to general 2-categories:

Definition 4.9.6 (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.

Remark 4.9.7 (Duals and adjunctions). By looking at the defining relations of duals in a rigid monoidal category (see Section 27.2), 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.

Property 4.9.8 (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 to identify different monoidal structures on this tricategory. (See also Section 27.8.)

4.9.2 n-functors

Definition 4.9.9 (2-functor). A 2-functor $F : \mathbf{A} \to \mathbf{B}$ (often called a **pseudofunctor**) is a morphism between bicategories. It consists of the following data:

- a function $F_0 : \mathrm{ob}(\mathbf{A}) \to \mathrm{ob}(\mathbf{B})$, and
- for every two objects $x, y \in \mathrm{ob}(\mathbf{A})$, a functor $F_{x,y} : \mathbf{A}(x, y) \to \mathbf{B}(Fx, Fy)$.

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 $\hom(\mathbf{A})$, 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 $\hom(\mathbf{A})$ the following identity holds:
$$\gamma_{f\circ g,h} \circ (\gamma_{f,g} \cdot \mathbb{1}_{Fh}) = \gamma_{f,g\circ h} \circ (\mathbb{1}_{Ff} \cdot \gamma_{g,h}). \tag{4.58}$$
2. **Unitor**: For every object $x \in \mathrm{ob}(\mathbf{A})$, a 2-isomorphism $\iota_x : \mathbb{1}_{Fx} \Rightarrow F\mathbb{1}_x$ such that for every morphism $f : x \to y$ in $\hom(\mathbf{A})$ the following identities hold:
$$\iota_y \cdot \mathbb{1}_{Ff} = \gamma_{\mathbb{1}_y,f} \tag{4.59}$$
$$\mathbb{1}_{Ff} \cdot \iota_x = \gamma_{f,\mathbb{1}_x}. \tag{4.60}$$

Note that to be completely formal one should have inserted the unitors and associators of the bicategories $\mathbf{A}, \mathbf{B}$.

Definition 4.9.10 (Lax natural transformation). Consider two 2-functors $F, G : \mathbf{A} \rightarrow \mathbf{B}$ between bicategories. A lax natural transformation $\eta : F \Rightarrow G$ consists of the following data:

1. for every object $x \in \mathrm{ob}(\mathbf{A})$, a 1-morphism $\eta_x : Fx \rightarrow Gx$, and
2. for every 1-morphism $f : x \rightarrow y$ in $\hom(\mathbf{A})$, 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 4.9.11. 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 4.9.12 (Modification). Consider two bicategories $\mathbf{A}, \mathbf{B}$, two 2-functors $F, G : \mathbf{A} \rightarrow \mathbf{B}$ and two parallel (lax) natural transformations $\alpha, \beta : F \Rightarrow G$. A modification $\mathfrak{m} : \alpha \Rrightarrow \beta$ maps every object $x \in \mathrm{ob}(\mathbf{A})$ 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 4.9.13 (Transfor). A k-transfor[16] between two n-categories maps j-morphisms to $(j+k)$-morphisms (in a coherent way).

Example 4.9.14. 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 4.2.11 to higher category theory:

Definition 4.9.15 (n-surjective functor). An ∞-functor $F : \mathbf{A} \rightarrow \mathbf{B}$ is said to be n-surjective if for any two parallel $(n-1)$-morphisms f, g in $\mathbf{A}$ and n-morphism $\alpha : Ff \rightarrow Fg$ in $\mathbf{B}$, there exists an n-morphism $\widetilde{\alpha}$ in $\mathbf{A}$ such that $F\widetilde{\alpha} \cong \alpha$.

Definition 4.9.16 (Indexed category). Consider a category $\mathbf{I}$. An $\mathbf{I}$-indexed category is a pseudofunctor $\mathbf{C} : \mathbf{I}^{op} \rightarrow \mathbf{Cat}$, i.e. a 2-presheaf on $\mathbf{I}$. Indexed functors and natural transformations are defined analogously.

[16]This name was first introduced by *Crans* in [7]. 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.

4.9.3 Higher (co)limits

Definition 4.9.17 (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, one says that $\lim^W F$ is the W-weighted (2-)limit of F if there exists a pseudonatural equivalence

$$\mathbf{C}(x, \lim^W F) \cong [\mathbf{I}, \mathbf{Cat}](W, \mathbf{C}(x, F-)). \tag{4.61}$$

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 ??

4.10 Groupoids

Definition 4.10.1 (Groupoid). A (small) groupoid $\mathcal{G}$ is a (small) category in which all morphisms are invertible.

Example 4.10.2 (Delooping). Consider a group G. Its delooping $\mathbf{B}G$ is defined as the one-object groupoid for which $\mathbf{B}G(*, *) = G$.

Property 4.10.3 (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 3.3.5 are exactly the natural transformations. More generally, all G-sets 3.3.1 can be obtained as functors $\mathbf{B}G \to \mathbf{Set}$.

Definition 4.10.4 (Core). Let $\mathbf{C}$ be a (small) category. The core $\mathrm{Core}(\mathbf{C}) \in \mathbf{Grpd}$ of $\mathbf{C}$ is defined as the maximal subgroupoid of $\mathbf{C}$.

Definition 4.10.5 (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 4.10.6 (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 4.10.7. Every groupoid is a (disjoint) union of its transitive components.

Definition 4.10.8 (Transitive groupoid). A groupoid $\mathcal{G}$ is said to be transitive if for all objects $x \neq y \in \mathrm{ob}(\mathcal{G})$, the set $\mathcal{G}(x, y)$ is not empty.

4.11 Lawvere theories ♣

Definition 4.11.1 (Lawvere theory). Let $\mathbf{F}$ denote the skeleton of **FinSet**. A Lawvere theory consists of a small category $\mathbf{L}$ and a strict (finite) product-preserving *identity-on-objects* functor $\mathcal{L} : \mathbf{F}^{op} \to \mathbf{L}$.

Equivalently, a Lawvere theory is a small category $\mathbf{L}$ with a **generic object** c_0 such that every object $c \in \mathrm{ob}(\mathbf{L})$ is a finite power of c_0.

Property 4.11.2. Lawvere theories $(\mathbf{L}, \mathcal{L})$ form a category **Law**. Morphisms between Lawvere theories are (finite) product-preserving functors.

Definition 4.11.3 (Model). A model or **algebra** over a Lawvere theory $\mathbf{L}$ is a (finite) product-preserving functor $A : \mathbf{L} \to \mathbf{Set}$.

?? COMPLETE ??

4.12 Operad theory ♣

4.12.1 Operads

Definition 4.12.1 (Plain operad[17]). 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 $n, k_1, \ldots, k_n$ there exists a composition map

$$\begin{aligned}\circ : P(n) \times P(k_1) \times \cdots \times P(k_n) &\to P(k_1 + \cdots + k_n)\\ : (\psi, \theta_1, \ldots, \theta_n) &\mapsto \psi \circ (\theta_1, \ldots, \theta_n)\end{aligned} \tag{4.62}$$

that satisfies two additional axioms:

- **identity**:

$$\theta \circ (\mathbb{1}, \ldots, \mathbb{1}) = \mathbb{1} \circ \theta = \theta, \tag{4.63}$$

and

- **associativity**:

$$\begin{aligned}&\psi \circ \Big(\theta_1 \circ (\theta_{1,1}, \ldots, \theta_{1,k_1}), \ldots, \theta_n \circ (\theta_{n,1}, \ldots, \theta_{n,k_n})\Big)\\ &\qquad = \Big(\psi \circ (\theta_1, \ldots, \theta_n)\Big) \circ (\theta_{1,1}, \ldots, \theta_{1,k_1}, \theta_{2,1}, \ldots, \theta_{n,k_n}).\end{aligned} \tag{4.64}$$

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 4.12.2. Plain operads can be defined in any monoidal category. In the same way symmetric operad can be defined in any symmetric monoidal category.

Example 4.12.3 (Endomorphism operad). Consider a vector space V. For every $n \in \mathbb{N}$, one can define the endomorphism algebra $\mathrm{End}(V^{\otimes n}, V)$. The endomorphism operad $\mathcal{E}\mathrm{nd}(V)$ is defined as $\{\mathrm{End}(V^{\otimes n}, V)\}_{n\in\mathbb{N}}$.

Definition 4.12.4 (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 4.12.5 (Categorical O-algebra). An O-algebra in the category **Cat**.

[17] Also called a **nonsymmetric operad** or **non-Σ operad**.

4.12.2 Algebraic topology

Definition 4.12.6 (Stasheff operad). A topological operad $\mathcal{K}$ such that $\mathcal{K}(n)$ is given by the n^{th} *Stasheff polytope/associahedron*. Composition is given by the inclusion of faces.

Definition 4.12.7 (A_∞-space). An algebra over the Stasheff operad. This induces the structure of a multiplication that is associative up to a coherent homotopy.

Definition 4.12.8 (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 4.12.9 (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.

?? COMPLETE ??

Chapter 5

Homological algebra

References for this chapter are [44, 45].

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 $\mathrm{im}(\partial_k)$ are called **boundaries** or **exact elements** and elements of $\ker(\partial_k)$ 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 a collection 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 can define the category $\mathbf{Ch}(\mathbf{A})$ 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 its homology groups $H_n(C_\bullet)$. Since $\partial^2 = 0$, the kernel $\ker(\partial_k)$ is a subgroup of the image $\mathrm{im}(\partial_{k+1})$ and it is even a normal subgroup. This way one can define the quotient group:

$$H_k(C_\bullet) := \frac{\ker(\partial_k)}{\mathrm{im}(\partial_{k+1})}. \tag{5.3}$$

The kernel in this definition, i.e. the group of k-cycles, is denoted by $Z_k(C_\bullet)$. The image in this definition, i.e. 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.

[1] A **cochain complex** is constructed similarly with an ascending order $\partial_k : C_k \to C_{k+1}$.

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 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.

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. $d\partial + \partial d = 0$, and
2. $d^2 = [\partial, D]$ for some other chain endomorphism D, i.e. d^2 is ∂-exact in $\mathrm{End}(C_\bullet)$.

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 $H_\bullet(d|H_\bullet(C_\bullet))$.

The following definitions use the language of Chapter 27:

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, 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 (Minimal model). Let $(C_\bullet, \partial_\bullet)$ be a (cohomological) DGCA of finite type. A model for $C_\bullet$ is a quasi-isomorphism $\rho : (A_\bullet, d_\bullet) \to (C_\bullet, \partial_\bullet)$ from a semifree DGCA $(A_\bullet, d_\bullet)$. This model is said to be **minimal** if $A_\bullet$ is freely generated in degrees ≥ 2 and satisfies $dA \subseteq \Lambda^{\geq 2} A$.

Remark 5.1.15 (Model structure on DGCAs ♣). By Property 8.1.75, the (minimal) models 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 $k \in \mathbb{N}$:

$$\mathrm{im}(\Phi_k) = \ker(\Phi_{k+1}). \tag{5.6}$$

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 5.1.1.

Definition 5.2.2 (Short exact sequence). A short exact sequence is 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 small examples, one can derive some important constraints for certain exact sequences. Consider the sequence

$$0 \longrightarrow C \xrightarrow{\Phi} D.$$

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 4.4.52).

5.3 Resolutions

Consider some Abelian category $\mathbf{A}$ and let $\mathbf{Ch}(\mathbf{A})$ denote the category of chain complexes with objects in $\mathbf{A}$.

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

$$\cdots \longrightarrow C_1 \longrightarrow C_0 \stackrel{\varepsilon}{\longrightarrow} X \longrightarrow 0. \tag{5.8}$$

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 (Figures 4.2a and 4.2b), 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. 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:*

- $\deg(s) - r(s) = 1$, *and*
- $s = \delta + d + \sum_{i=1}^{\infty} s_{(i)}$ *where* $r(s_{(i)}) = i$ *and* $\deg(s_{(i)}) = i + 1$.

Moreover, any differential that satisfies these properties has the same homology as d:

$$H_\bullet(s) \cong H_\bullet(d|H_\bullet(C_\bullet)) \equiv H_\bullet(d|H_0(C_\bullet)). \tag{5.9}$$

5.4 Derived functors

Given an additive functor 4.8.5, one can define its **prolongation** on 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 $P_\bullet \xrightarrow{\varepsilon} X \to 0$. Apply the prolongation to this resolution and construct the homology of the resulting chain complex:
>
> $$L_iF(X) := H_i(FP_\bullet). \tag{5.10}$$
>
> 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 remarking that every projective object P admits a projective resolution of the form

$$\cdots \longrightarrow 0 \longrightarrow 0 \longrightarrow P \longrightarrow P \longrightarrow 0.$$

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 $\mathbf{A}$:

$$0 \longrightarrow A \longrightarrow B \longrightarrow C \longrightarrow 0. \tag{5.11}$$

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:

$$0 \longrightarrow A_\bullet \longrightarrow B_\bullet \longrightarrow C_\bullet \longrightarrow 0. \tag{5.12}$$

Since F is additive and the above sequence is exact, the induced complex is also exact, i.e. the sequence

$$0 \longrightarrow FA_\bullet \longrightarrow FB_\bullet \longrightarrow FC_\bullet \longrightarrow 0 \tag{5.13}$$

is exact and so the *zig-zag lemma* is applicable. This theorem gives the following long exact sequence in homology:

$$\cdots \longrightarrow H_i(FB_\bullet) \longrightarrow H_i(FC_\bullet) \longrightarrow H_{i-1}(FA_\bullet) \longrightarrow H_{i-1}(FB_\bullet) \longrightarrow \cdots . \tag{5.14}$$

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:

$$L_1F = 0 \implies L_iF = 0 \qquad \forall i \geq 1, \tag{5.15}$$

and, more generally:

$$L_iF = 0 \implies L_jF = 0 \qquad \forall j \geq i. \tag{5.16}$$

5.4.1 Modules

Consider the tensor and hom-bifunctors $-\otimes-$ and $\mathrm{Hom}(-,-)$ in the category **Mod** 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 $\mathrm{Tor}^R_n(-,A)$ 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 $\mathrm{Ext}^n_R(-,A)$ are defined as the right derived functors of the hom-functor $\mathrm{Hom}_R(-,A)$.

Definition 5.4.12 (Flat module). An R-module M such that the induced tensor functor

$$-\otimes_R M : R\mathbf{Mod} \to R\mathbf{Mod} \tag{5.17}$$

is exact. By Property 5.4.4 this implies that a flat module is $\otimes$-acyclic 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.

Definition 5.4.13 (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 defines the Koszul complex $K(s)$ as follows:

$$0 \longrightarrow \Lambda^r M \longrightarrow \Lambda^{r-1}M \longrightarrow \cdots \longrightarrow M \xrightarrow{s} R \longrightarrow 0, \tag{5.18}$$

where the exterior powers $\Lambda^k M$ are defined as in Section 21.4.4, i.e. they are the free modules spanned by totally antisymmetric k-tuples in M. The differentials are defined as

$$d_k(m_1\wedge\ldots\wedge m_k) := \sum_{i=1}^k (-1)^{k+1} s(m_i) m_1\wedge\ldots\wedge\widehat{m_i}\wedge\ldots\wedge m_k, \tag{5.19}$$

where the caret $\widehat{\cdot}$ means that this element is omitted. It is clear that $d_1 = s$. The homology of this complex is called the **Koszul homology** of s.

Example 5.4.14. 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

$$s : R^n \to R : (r_1,\ldots,r_n) \mapsto r_1x_1 + \cdots + r_nx_n. \tag{5.20}$$

The associated Koszul complex is denoted by $K(x_1,\ldots,x_n)$.

Property 5.4.15 (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:

$$H_{i\geq 1}(K(x_1, \ldots, x_n)) = 0, \tag{5.21}$$

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 5.4.16 (Koszul-Tate resolution). Consider a commutative ring R with an ideal I. For any element $x \in I$, one can construct the polynomial algebra $R[t]$ on a formal generator t and extend the differential by $\partial t := x$. 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 kill the higher homology of I. If[2] $I \equiv (x_1, \ldots, x_k)$, one can consider the Koszul complex $(X^0, d^0) := (K(x_1, \ldots, x_k), d)$. 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 $H_1(X^0, d^0)$. Now, consider the Koszul complex $(X^1, d^1) := (K(x_1, \ldots, x_k, x'_1, \ldots, x'_l), d')$ induced by the morphism

$$s' : R^{k+l} \to R : (r_1, \ldots r_{k+l}) \mapsto r_1x_1 + \cdots + r_kx_k + r_{k+1}x'_1 + \cdots + r_{k+l}x'_l, \tag{5.22}$$

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 the graded-commutative algebra 27.1.8). It should be clear that $H_0(X^1, d^1) \cong R/I$ and $H_1(X^1, d^1) = 0$. The direct limit of this construction is called the **Koszul-Tate resolution** of (R, I).

5.4.2 Group cohomology

In this section an important application of derived functors is given. In fact this was one of the motivating applications. 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. However, the literature on these applications often starts with an ad hoc construction based on maps from a group to a module (see Definition 3.4.1).

For simplicity only finite groups and Abelian coefficients groups will be considered. Every G-module 3.3.6 can be regarded as a module over the group ring $\mathbb{Z}[G]$, i.e. there exists an equivalences 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**. For groups 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:

$$H^\bullet(G; A) := \mathrm{Ext}^\bullet_{\mathbb{Z}[G]}(\mathbb{Z}, A) \tag{5.23}$$

$$H_\bullet(G; A) := \mathrm{Tor}^{\mathbb{Z}[G]}_\bullet(A, \mathbb{Z}), \tag{5.24}$$

where $\mathbb{Z}$ carries the trivial G-module structure. To explicitly calculate the (co)homology groups, one has to find an acyclic resolution of $\mathbb{Z}$:

[2] If R is Noetherian, this is always possible.

Construction 5.4.17 (Normalized bar resolution). Let P'_k be a free rank-k G-module. The boundary maps are defined as follows:

$$\partial_k(g_1,\ldots,g_k) = g_1(g_2,\ldots,g_k) + \sum_{i=1}^{k}(-1)^i(g_1,\ldots,g_ig_{i+1},\ldots,g_k) + (-1)^k(g_1,\ldots,g_{k-1}). \quad (5.25)$$

To obtain the normalized bar[3] resolution (in inhomogeneous form), one has to quotient out the submodule of P'_n generated by tuples $(g_1,\ldots,g_n)$ 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 $H^k(G;A) = H^k(\mathrm{Hom}_{\mathbb{Z}[G]}(P_\bullet, A))$, 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:

$$S = \{(g_1,\ldots,g_k) \mid \forall i \leq k : g_i \in G\}.$$

Since the module is free over $\mathbb{Z}[G]$, one can write every element as a formal linear combination of elements of the form

$$g_0(g_1,\ldots,g_k).$$

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:

$$\varphi(g_1,\ldots,g_k) := (e, g_1, g_1g_2, \ldots, g_1g_2\cdots g_k). \quad (5.26)$$

It is not hard to show that this morphism is in fact an isomorphism (of G-modules) and that

$$H^k(G;A) = H^k(\mathrm{Hom}_{\mathbb{Z}[G]}(\mathbb{Z}[G^{k+1}], A)). \quad (5.27)$$

By a little more algebra, it can also be shown that this hom-set is isomorphic to the (set-theoretic) mapping space $\mathrm{Map}(G^k, A)$. 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:

Construction 5.4.18 (Group cohomology). Let G be a finite group and let A be a G-module. Denote by C^k the free Abelian group generated by the set-theoretic functions $f : G^k \to A$ with the property that if any of their arguments is the identity, the result is 0. The boundary maps ∂^k, induced by the maps defined in Equation (5.25), are given by:

$$(\partial^k f)(g_1,\ldots,g_{k+1}) = g_1 \cdot f(g_2,\ldots,g_{k+1}) + \sum_{i=1}^{k}(-1)^i f(\ldots,g_ig_{i+1},\ldots) + (-1)^{k+1}f(g_1,\ldots,g_k). \quad (5.28)$$

This is exactly the relation used to obtain group cohomology in Definition 3.4.1.

Property 5.4.19 (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.

[3] One of the possible explanations for this name is that the formal generating elements are often written as $[g_1|g_2|\ldots|g_k]$.

Property 5.4.20 (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

$$H^\bullet(G; C) \to H^{\bullet+1}(G; A) \tag{5.29}$$

is called the Bockstein homomorphism.

5.5 Spectral sequences

Remark. In this section the homological convention is adopted, i.e. differentials lower the degree.

Definition 5.5.1 (Spectral sequence). Consider a collection $\{(E_i, d_i)\}_{i\in\mathbb{N}}$ of differential objects. This collection is called a spectral sequence if it satisfies

$$H(E_i, d_i) \cong E_{i+1} \tag{5.30}$$

for every $i \in \mathbb{N}$. A morphism of spectral sequences is a collection of morphisms $(\varphi_i)_{i\in\mathbb{N}}$ satisfying

1. $\varphi_i \circ d_i = d'_i \circ \varphi_i$, and
2. $\varphi_{i+1} = H(\varphi_i)$.

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.

$$\begin{tikzcd}
A \arrow[rr, "\alpha"] & & A \arrow[dl, "\beta"] \\
& B \arrow[ul, "\gamma"] &
\end{tikzcd}$$

(a) Exact couple.

$$\begin{tikzcd}
A' \arrow[rrrr, "\alpha'"] & & & & A' \arrow[dddll, "\beta'"] \\
& A \arrow[ul, "f"'] \arrow[rr, "\alpha"] & & A \arrow[ur, "f"] \arrow[dl, "\beta"] & \\
& & B \arrow[ul, "\gamma"] \arrow[d, "g"] & & \\
& & B \arrow[uuull, "\gamma'"] & &
\end{tikzcd}$$

(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:

$$E_0 := B \tag{5.31}$$
$$d_0 := \beta \circ \gamma \tag{5.32}$$

$$\vdots$$

$$E_n := \frac{\gamma^{-1}(\alpha^n(A))}{\beta(\alpha^{-n}(0))} \tag{5.33}$$
$$d_n := \beta \circ \alpha^{-n} \circ \gamma \tag{5.34}$$

It is not hard to see that $E_{n+1} = H(E_n, d_n)$, 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, 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 then 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

$$E_\infty := \frac{\cap_i Z(E_i)}{\cup_i B(E_i)}, \tag{5.35}$$

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 $\{F_pC\}_{p\in\mathbb{N}}$. Definition 2.4.11 of a filtration immediately gives a short exact sequence for every $p \in \mathbb{N}$:

$$0 \longrightarrow F_{p-1}C \longrightarrow F_pC \longrightarrow F_pC/F_{p-1}C \longrightarrow 0. \tag{5.36}$$

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:

$$\begin{tikzcd} D \arrow[rr, "\alpha"] & & D \arrow[dl, "\beta"] \\ & E \arrow[ul, "\gamma"] & \end{tikzcd} \tag{5.37}$$

where $D_p = H(F_pC)$ and $E_p = H(F_pC/F_{p-1}C)$. 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

$$\mathrm{Gr} : C \mapsto \Big\{G_pC := F_pC/F_{p-1}C\Big\}_{p\in\mathbb{N}}.$$

On the other hand, one could of course also construct the composition $\mathrm{Gr} \circ H$ that first maps a differential object to its homology object and then builds the graded object associated to the filtration

$$F_pH(C) := \mathrm{im}(H(F_pC) \to H(C)).$$

Some straightforward questions can arise at this point: "*How are the functors* $H \circ \mathrm{Gr}$ *and* $\mathrm{Gr} \circ H$ *related?*", "*Do they coincide?*", ... The latter question is easy to answer: "*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. $(C_\bullet, d_\bullet) = (\{C_p\}_{p\in\mathbb{Z}}, d)$ such that $dC_p \subseteq C_p$. 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. $dF_iC_j \subseteq F_iC_{j-1}$.

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

$$d_0 : \frac{F_pC_q}{F_{p-1}C_q} \to \frac{F_pC_{q-1}}{F_{p-1}C_{q-1}} \tag{5.38}$$

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 $dx = 0 \mod F_{p-1}C_{q-1}$ (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 $E_1 = H(E_0, d_0)$ is given by

$$E_1^{p,q} := \frac{\{x \in F_pC_q \mid dx \in F_{p-1}C_{q-1}\}}{F_{p-1}C_q + dF_pC_{q+1}}. \tag{5.39}$$

The first term in the denominator was already explained above. The second term comes from the $\mathrm{im}(d_0)$-part in the definition of $H(E_0, d_0)$. One might suspect that some data is missing since the relevant map $d_0^{p,q+1}$ 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:

$$E_r^{p,q} := \frac{\{x \in F_pC_q \mid dx \in F_{p-r}C_{q-1}\}}{F_{p-1}C_q + dF_{p+r-1}C_{q+1}}. \tag{5.40}$$

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
$$Z^r_{p,q} := \{c \in G_pC_q \mid dc = 0 \bmod F_{p-r}C_{q-1}\}$$
are called r-**almost** (p,q)-**cycles**.

- The elements of
$$B^r_{p,q} := dF_{p+r-1}C_{q+1}$$
are called r-**almost** (p,q)-**boundaries**.

It is then easy to see that the page E^r satisfies

$$E^r_{p,q} = Z^r_{p,q}/B^r_{p,q}, \tag{5.41}$$

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:

$$B^0_{p,q} \hookrightarrow \cdots \hookrightarrow B^\infty_{p,q} \hookrightarrow Z^\infty_{p,q} \hookrightarrow \cdots \hookrightarrow Z^0_{p,q}. \tag{5.42}$$

5.5.3 Convergence

Definition 5.5.3 (Limit term). Consider a spectral sequence $\{E^r_{p,q}\}$. 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 $r \geq r(p,q)$:

$$E^r_{p,q} \cong E^{r(p,q)}_{p,q}, \tag{5.43}$$

the object $E^\infty := \{E^{r(p,q)}_{p,q}\}$ 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

$$E^\infty_{p,q} \cong G_pH_q \qquad\qquad \forall p,q \in \mathbb{Z}. \tag{5.44}$$

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:

$$E^r_{p,q} \Rightarrow H_\bullet(C). \tag{5.45}$$

?? COMPLETE ??

Chapter 6

Logic and Type theory

The main reference for this chapter is [35]. For a formal introduction to λ-calculus see [116].

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 (in particular the chapter on category theory 4). 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.

6.1 Logic

6.1.1 Languages

Definition 6.1.1 (Language). An **alphabet** is a set of symbols. A **word** in the language is a string of symbols in the alfabet.

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 6.1.2 (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 $\mathrm{ar} : A \to \mathbb{N}$ 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 6.1.3 (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 $\mathrm{ap}_f : U^{\mathrm{ar}(f)} \to U$. In particular, for each constant c, an element $u_c \in U$.
3. For each relation symbol $\in$, a set $R_\in \subseteq U^{\mathrm{ar}(\in)}$.

Definition 6.1.4 (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 $x_1, \ldots, x_n$, $f(x_1, \ldots, x_n)$ is also a term.

Definition 6.1.5 (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**.

6.1.2 Propositional logic

Definition 6.1.6 (Proposition). A statement that is either *true* or *false* (not both).

Definition 6.1.7 (Paradox). A statement that cannot (consistently) be assigned a truth value.

Definition 6.1.8 (Contradiction). A statement that is always *false*.

Definition 6.1.9 (Tautology). A statement that is always *true*.

Notation 6.1.10 (Truth values). The truth values *true* and *false* are denoted by $\top$ and $\perp$ respectively.

Definition 6.1.11 (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 not (**negation**): $\neg P$.

This last symbol is in fact an abbreviation for the implication $P \to \perp$.

The basic inference rule is given by **modus ponens**:

$$\text{If } P \text{ and } P \to Q\text{, then } Q. \tag{6.1}$$

The general deductive system for propositional logic is obtained by combining this rule with the following axioms:

1. If P, then $Q \to P$.
2. If $P \to Q \to R$, then $P \to Q$ implies $P \to R$.
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 $P \to Q$, then $R \to Q$ implies $P \vee R \to Q$.
8. If $\perp$, then P. This principle is often called **ex falso quodlibet**.

Remark 6.1.12 (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 2.6.34. ?? EXPLAIN THIS ??

6.1.3 Predicate logic

?? COMPLETE ??

6.2 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 2.1 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 *types*. A specific function, proof or element can then be seen as an *inhabitant* of a given type.

Definition 6.2.1 (Type judgement). A judgement 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 6.2.2 (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 (in general this depends on previously defined types).
2. **Introduction rule**: This rule gives a **constructor** of the new type (in general this depends on a **context**, i.e. a collection of existing terms).
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.

As in [35], 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 introduced. First, there is the **judgemental equality** or **definition 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 $f(5) = 25$ can be proven, but it is not true by definition (it would depend on the definition of the natural numbers). This sort of equality will be denoted by an ordinary equals sign $=$.

6.3 Basic constructions

6.3.1 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 6.3.1 (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**[1]: $\lambda x.\Phi(a) :\equiv \Phi(a)$, i.e. function application is equivalent to the subsitution 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 dinstinct 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:

Definition 6.3.2 (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

$$\prod_{a:A} B(a) : \mathcal{U}.$$

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 6.3.3 (Scope). The Π-symbol scopes over all expressions to the right of the symbol, unless delimited (similar to λ-calculus), e.g.

$$\prod_{a:A} B(a) \to C(a) \equiv \prod_{a:A} \Big(B(a) \to C(a)\Big).$$

Example 6.3.4 (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 types themselves) 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**

$$\mathrm{id} : \prod_{A:\mathcal{U}} A \to A \tag{6.2}$$

defined by

$$\mathrm{id} :\equiv \lambda(A : \mathcal{U}).\lambda(a : A).a. \tag{6.3}$$

[1] In λ-calculus this is often called β-reduction. (See the next section.)

6.3.2 λ-calculus

?? COMPLETE (e.g. Curry-Howard or even Curry-Howard-Lambek, typed vs. untyped calculus, ...)??

6.3.3 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.

Definition 6.3.5 (Equality type[2]). The type corresponding to a propositional equality is defined by the following rules:

- **Formation rule**: Given terms $a, b : A$, one can form the equlity 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
$$\mathrm{refl}_a : a = a. \tag{6.4}$$
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 [35].

  Given a type family
$$C : \prod_{a,b:A} a = b \to \mathcal{U}$$
  and a term
$$I : \prod_{a:A} C(a, a, \mathrm{refl}_a),$$
  there exists a function
$$f : \prod_{a,b:A} \prod_{p:a=b} C(a, b, p) \tag{6.5}$$
  such that
$$f(a, a, \mathrm{refl}_a) :\equiv I(a) \tag{6.6}$$
  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 6.3.6 (Mere proposition). A type $A : \mathcal{U}$ for which the type
$$\mathrm{isProp}(A) :\equiv \prod_{a,b:A} a = b \tag{6.7}$$
is inhabited.

[2]Sometimes called an **identity type**.

6.3.4 Products

As in classic set theory a basic notion is that of products. This construction is ubiquitous throughout all corners 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 6.3.7 (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 the **recursor** is defined:

$$\mathrm{rec}_{A\times B} : \prod_{C:\mathcal{U}} (A \to B \to C) \to A \times B \to C \tag{6.8}$$

  with the constraint

$$\mathrm{rec}_{A\times B}(C, f, (a, b)) :\equiv f(a)(b). \tag{6.9}$$

Example 6.3.8 (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

$$\pi_1(a, b) :\equiv a \qquad \text{and} \qquad \pi_2(a, b) :\equiv b \tag{6.10}$$

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 6.3.9 (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:

$$\mathrm{rec}_{\mathbf{1}} : \prod_{C:\mathcal{U}} C \to \mathbf{1} \to C. \tag{6.11}$$

  On the constructor $* : \mathbf{1}$ it is required to act trivially:

$$\mathrm{rec}_{\mathbf{1}}(C, c_0, *) :\equiv c_0. \tag{6.12}$$

Definition 6.3.10 (Dependent functions). One can easily generalize the above recursion functions to **induction** functions, 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 $\mathrm{rec}_{A\times B}$. 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):

$$\mathrm{ind}_{A\times B} : \prod_{C:A\times B\to\mathcal{U}} \left(\prod_{a:A}\prod_{b:B} C(a,b) \to \prod_{x:A\times B} C(x)\right), \tag{6.13}$$

$$\mathrm{ind}_{\mathbf{1}} : \prod_{C:\mathbf{1}\to\mathcal{U}} C(*) \to \prod_{x:\mathbf{1}} C(x). \tag{6.14}$$

Property 6.3.11 (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 such 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 6.3.12 (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

$$\sum_{a:A} B(a) : \mathcal{U}.$$

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 6.3.13. Dependent pair types are often called **Σ-types** (due to the notation).

Remark 6.3.14 (Scope). Like the Π-symbol, the Σ-symbol scopes over the rest of the expression unless delimited.

Definition 6.3.15 (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 recursion and induction functions are defined as follows:

$$\mathrm{rec}_{A+B} : \prod_{C:\mathcal{U}} (A \to C) \to (B \to C) \to A + B \to C, \tag{6.15}$$

$$\mathrm{ind}_{A+B} : \prod_{C:A+B\to\mathcal{U}} \left(\prod_{a:A} C(\iota_1(a))\right) \to \left(\prod_{b:B} C(\iota_2(b))\right) \to \prod_{x:A+B} C(x). \tag{6.16}$$

- **Computation rules**: The recursion function acts on the constructors as follows (the induction function virtually has the same action):

$$\mathrm{rec}_{A+B}(C, f_1, f_2, \iota_1(a)) :\equiv f_1(a), \tag{6.17}$$

$$\mathrm{rec}_{A+B}(C, f_1, f_2, \iota_2(b)) :\equiv f_2(b). \tag{6.18}$$

Definition 6.3.16 (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 $\mathbf{0}$:

$$\mathrm{rec}_{\mathbf{0}} : \prod_{C:\mathcal{U}} \mathbf{0} \to C \tag{6.19}$$

$$\mathrm{rec}_{\mathbf{0}} : \prod_{C:\mathbf{0}\to\mathcal{U}} \prod_{x:\mathbf{0}} C(x). \tag{6.20}$$

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 6.3.17 (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:

$$\mathrm{rec}_{\mathbf{2}} : \prod_{C:\mathcal{U}} C \to C \to \mathbf{2} \to C \tag{6.21}$$

with

$$\mathrm{rec}_{\mathbf{2}}(C, c_0, c_1, 0) :\equiv c_0, \tag{6.22}$$

$$\mathrm{rec}_{\mathbf{2}}(C, c_0, c_1, 1) :\equiv c_1. \tag{6.23}$$

Using this type one can prove that $A + B$ is judgementally equal to $\sum_{x:\mathbf{2}} \mathrm{rec}_{\mathbf{2}}(\mathcal{U}, A, B, x)$. 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 $\mathbf{2}$.

6.3.5 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.

6.4 Homotopy type theory

6.4.1 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 Sections 8.1 and 4.10. 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.[3]

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 6.4.1 (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 6.4.2 (Loop space). The loop space $\Omega(A, a)$ of a pointed type (A, a) is the pointed type $(a =_A a, \mathrm{refl}_a)$.

Now, the important aspect of HoTT is that the ∞-groupoid structure of a type can be derived solely from the (path) induction principle of the equality types. Some examples are given:

Property 6.4.3 (Inversion). For every type $A : \mathcal{U}$ and terms $a, b : A$, there exists a function

$$p \mapsto p^{-1} : (a = b) \to (b = a) \tag{6.24}$$

such that $\mathrm{refl}_a^{-1} :\equiv \mathrm{refl}_a$ for all $a : A$.

[3]This characterization is strongly related to the homotopy hypothesis (or theorem when using the right model for ∞-categories).

Property 6.4.4 (Concatenation). For every type $A:\mathcal{U}$ and terms $a,b,c,d:A$, there exists a function

$$p \mapsto q \mapsto p \cdot q : (a=b) \to (b=c) \to (c=d) \tag{6.25}$$

such that $\mathrm{refl}_a \cdot \mathrm{refl}_a :\equiv \mathrm{refl}_a$ 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 6.4.5. The above operations satisfy the group relations (up to higher equalities):

- $p \cdot \mathrm{refl}_b = p$ and $\mathrm{refl}_a \cdot p = p$ for all $p:a=b$.
- $p \cdot p^{-1} = \mathrm{refl}_a$ and $p^{-1} \cdot p = \mathrm{refl}_b$ 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 $p:a=b, q:b=c, r:c=d$.

6.4.2 Transport

The relation with homotopy theory and category theory becomes even stronger when looking at function types:

Property 6.4.6. Given a function $f:A\to B$, there exists an **application function**

$$\mathrm{ap}_f : (a =_A b) \to (f(a) =_B f(b)) \tag{6.26}$$

such that $\mathrm{ap}_f(\mathrm{refl}_a) :\equiv \mathrm{refl}_{f(a)}$ 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 6.4.7. 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 6.4.8 (Transport). Given a type family $P:A\to\mathcal{U}$ and an equality $p:a=_A b$, there exists a **transport function**

$$p_* : P(a) \to P(b) \tag{6.27}$$

such that $(\mathrm{refl}_a)_* :\equiv \mathrm{id}(a)$ 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 8.1.60. For every type family $P:A\to\mathcal{U}$, term $\alpha:P(a)$ and equality $p:a=b$, there exists a **lift**

$$\mathrm{lift}(p,\alpha) : (a,\alpha) = (b, p_*(\alpha)) \tag{6.28}$$

such that

$$\pi_1(\mathrm{lift}(p,u)) = p. \tag{6.29}$$

The equality $\mathrm{lift}(p,u)$ 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 $\mathrm{lift}(p,\cdot)$ 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 6.4.9. Given a dependent function $f : \prod_{a:A} P(a)$, there exists a function

$$\mathrm{apd}_f : \prod_{p:a=b} p_*(f(a)) =_{P(b)} f(b). \tag{6.30}$$

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 6.4.10. Consider two types $A, B : \mathcal{U}$ and a function $f : A \rightarrow B$. For every equality $p : a =_A b$ and term $\alpha : P(b)$, there exists an equality $\widetilde{p} : p_*(\alpha) =_{P(b)} \alpha$. Using this equality one can relate the application functions as follows:

$$\mathrm{apd}_f(p) = \widetilde{p}(f(a)) \cdot \mathrm{ap}_f(p). \tag{6.31}$$

6.4.3 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 6.4.11 (Homotopy). Consider two sections $f, g : \prod_{a:A} P(a)$. A homotopy between f and g is a term of the type

$$f \sim g :\equiv \prod_{a:A} f(a) = g(a). \tag{6.32}$$

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 6.4.12. 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

$$H(a) \cdot g(p) = f(p) \cdot H(b). \tag{6.33}$$

Using the notion of homotopy one can introduce a first kind of "equivalence":

Definition 6.4.13 (Quasi-inverse). Given a function $f : A \rightarrow B$, a quasi-inverse of f is a triple (g, α, β), where $g : B \rightarrow A$ and

$$\alpha : f \circ g \sim \mathrm{id}_B \qquad\qquad \beta : g \circ f \sim \mathrm{id}_B. \tag{6.34}$$

From a homotopy theoretical perspective one would call the pair (f, g) a homotopy equivalence. The corresponding type is given by

$$\mathrm{qInv}(f) :\equiv \sum_{g:B\rightarrow A} (f \circ g \sim \mathrm{id}_B) \times (g \circ f \sim \mathrm{id}_A). \tag{6.35}$$

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 $\mathrm{qInv}(f) \to \mathrm{isEquiv}(f)$.
2. For every function $f : A \to B$, there also exists a function $\mathrm{isEquiv}(f) \to \mathrm{qInv}(f)$.
3. For every two terms $eq_1, eq_2 : \mathrm{isEquiv}(f)$, 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.

6.4.4 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

$$(x =_{A\times B} y) \simeq (\pi_1(x) =_A \pi_1(y)) \times (\pi_2(a) =_B \pi_2(y)). \tag{6.36}$$

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[4] for the defining rules of the equality type in the case of product types:

Remark 6.4.14. The function $(\pi_1(a) = \pi_1(b)) \times (\pi_2(a) = \pi_2(b)) \to (a = b)$ 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 6.4.15. 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 $(P \times Q)(x) :\equiv P(\pi_1(x)) \times Q(\pi_2(x))$ the following equality is inhabited:

$$p_*(\alpha) = \big(p_*(\pi_1(\alpha)), p_*(\pi_2(\alpha))\big). \tag{6.37}$$

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

[4]Note that this is not a judgementally equal alternative. It is merely a convenient interpretation.

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 6.1 (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 $f(\mathrm{refl}_x)$.

Axiom 6.2 (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 ??

6.5 Modal type theory

?? COMPLETE ??

6.6 Computability theory

6.6.1 Functions

Definition 6.6.1 (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 $\mathrm{dom}(f) = S$.

Definition 6.6.2 (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 $\mathrm{dom}(f_n) = S_n$ for all $n \in \mathbb{N}$.

?? COMPLETE ??

Part II

Topology

Chapter 7

General Topology

7.1 Topological spaces

Definition 7.1.1 (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**. The **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 7.1.2 (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 7.1.3 (Pointed topological space). Let $x_0 \in X$ be any element of a topological space. The triple (X, τ, x_0) is called a pointed topological space with base point x_0.

Example 7.1.4 (Relative topology[1]). Any subset Y of a topological space (X, τ_X) can be turned into a topological space by equipping it with the following topology:

$$\tau_{\text{rel}} := \{U_i \cap Y \mid U_i \in \tau_X\}. \tag{7.1}$$

Example 7.1.5 (Discrete topology). The topology in which every subset is open (and thus also closed).

Example 7.1.6 (Indiscrete topology). The topology in which only the empty set and the space itself are open.

Definition 7.1.7 (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 7.1.8 (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 7.1.9 (Boundary). The boundary ∂Y of a subset Y of a topological space X is defined as $\overline{Y} \backslash Y^\circ$.

[1] Sometimes called the **subspace topology**.

Definition 7.1.10 (Borel set). Let $\mathcal{B}$ be the σ-algebra 2.5.2 generated by all open subsets of a topological space. The elements $B \in \mathcal{B}$ are called Borel sets.

Property 7.1.11 (Real line). For $\mathbb{R}$, the open, closed and half-open (both types) intervals all generate the same σ-algebra and, accordingly, the same Borel sets.

Definition 7.1.12 (Topological group). A group equipped with a topology such that both the multiplication and inversion morphisms are continuous.

Definition 7.1.13 (G_δ-set). A countable intersection of open sets.

Definition 7.1.14 (F_σ-set). A countable union of closed sets.

7.1.1 Neighbourhoods

Definition 7.1.15 (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 7.1.16 (Limit point). Let Y be a subset of X. A point $x \in X$ is called a limit point of Y if every neighbourhood of x contains at least one point of Y different from x.

By relaxing the last part of this definition, a slightly different notion is obtained:

Definition 7.1.17 (Adherent point). Let Y be a subset of X. A point $x \in X$ is called an adherent point of Y if every neighbourhood of x contains at least one point of S. A point x is an adherent point of Y if and only if it is an element of the closure $\overline{Y}$.

Definition 7.1.18 (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 7.1.19 (Basis). A collection $\mathcal{B} \subseteq \tau$ of open subsets of a topological space (X, τ) is a basis for (X, τ) if every $U \in \tau$ can be written as

$$U = \bigcup_{V \in \mathcal{F}} V, \tag{7.2}$$

where $\mathcal{F} \subseteq \mathcal{B}$.

Definition 7.1.20 (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 7.1.21 (First-countable space). A topological space (X, τ) for which for every point $x \in X$ there exists a countable local basis.

Property 7.1.22 (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 7.1.23 (Second-countable space). A topological space (X, τ) for which there exists a countable (global) basis.

Property 7.1.24 (Closure). Let X be a topological space. The closure of a subset $Y \subseteq X$ is given by

$$\overline{Y} = \{x \in X \mid \exists \text{ a net } (x_\alpha)_{\alpha \in I} \text{ in } X : x_\alpha \longrightarrow x\}. \tag{7.3}$$

This implies that the topology on X is completely determined by the convergence of nets 2.6.13.

[2] Sometimes called a **cluster point**.

Definition 7.1.25 (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 (7.3), 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 7.1.26. Every first-countable space is Fréchet-Urysohn and, therefore, only convergent sequences have to be considerd in these spaces.

Definition 7.1.27 (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

$$f(u) = g(u) \qquad\qquad \forall u \in N,$$

this property defines an equivalence relation denoted by $f \sim_x g$ and the equivalence classes are called germs.

7.1.2 Separation axioms

Definition 7.1.28 (Irreducible). A topological space is said to be irreducible if it is not the union of two proper closed subsets or, equivalently, if the intersection of two nonempty open subsets is again nonempty.

Definition 7.1.29 (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. The points are said to be **topologically distinguishable**. T_0-spaces are also said to carry a **Kolmogorov topology**.

Definition 7.1.30 (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 $x \notin N'$. The points are said to be **separated**. T_1-spaces are also said to carry a **Fréchet topology** (not to be confused with Fréchet spaces from functional analysis).

Definition 7.1.31 (Hausdorff space). A topological space X is a Hausdorff space or T_2-space if it satisfies the following condition:

$$\forall x, y \in X : \exists \text{ neighbourhoods } N \ni x, N' \ni y : N \cap N' = \emptyset. \tag{7.4}$$

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 7.1.32 (Closed points). Every singleton and, by extension, every finite subset is closed in a Hausdorff space.

Definition 7.1.33 (Urysohn space). A topological space is an Urysohn space or $T_{2^1/_2}$-space if every two distinct points are separated by closed neighbourhoods.

Definition 7.1.34 (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 7.1.35 (T_3-space). A space that is both regular and T_0.

Definition 7.1.36 (Normal space). A topological space such that every two closed subsets have disjoint neighbourhoods.

Definition 7.1.37 (T_4-space). A space that is both normal and T_1.

Property 7.1.38 (Nesting of axioms). A space satisfying the separation axiom T_k also satisfies all separation axioms $T_{i\leq k}$.

7.1.3 Convergence

Definition 7.1.39 (Convergence). A sequence $(x_n)_{n\in\mathbb{N}}$ in X is said to converge to a point $x \in X$ if

$$\forall \text{ neighbourhoods } U \text{ of } x(\exists N \in \mathbb{N}_0(\forall n > N : x_n \in U)). \tag{7.5}$$

The "limit" of a convergent sequence does not have to be unique:

Property 7.1.40 (Uniqueness). The limit of a converging sequence in a Hausdorff space is unique.

Property 7.1.41 (Subsequences). Every subsequence of a converging sequence converges to the same point.

7.2 Morphisms

7.2.1 Continuity

Definition 7.2.1 (Continuity). A function between topological spaces is said to be continuous if the inverse image of every open set is also open. The set of all continuous functions between two topological spaces X, Y is often denoted by $C(X, Y)$.

Definition 7.2.2 (Initial topology). Consider a collection of functions $\{f_i : X \to Y_i\}_{i\in I}$ between topological spaces. The initial topology on X with respect to this family is the coarsest topology on X for which all maps f_i are continuous.

Definition 7.2.3 (Final topology). Consider a collection of functions $\{f_i : Y_i \to X\}_{i\in I}$ between topological spaces. The final topology on X with respect to this family is the finest topology on X for which all maps f_i are continuous.

Property 7.2.4 (Continuity). Consider a function $f : X \to Y$ of topologial 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 $a \in X$.

Corollary 7.2.5. If the space Y in the previous theorem is Hausdorff, the limit $f(a)$ does not need to be known since it is unique by Property 7.1.40 above.

Remark 7.2.6. If the space X is not first-countable, one has to consider the convergence of nets 2.6.13.

Theorem 7.2.7 (Urysohn's lemma). *A topological space X is normal 7.1.36 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*

$$f(a) = 0 \qquad \textit{and} \qquad f(b) = 1. \tag{7.6}$$

The following, seemingly unrelated, theorem is actually equivalent to Urysohn's lemma:

Theorem 7.2.8 (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 7.2.9 (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 $x \in U$.

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 $x_0 \in X$.

7.2.2 Homeomorphisms

Definition 7.2.10 (Homeomorphism). A function f such that both f and f^{-1} are continuous and bijective.

Definition 7.2.11 (Embedding). A continuous function that is a homeomorphism onto its image.

Definition 7.2.12 (Local homeomorphism). A continuous function $f : X \to Y$ is a local homeomorphism if for every point $x \in X$ there exists an open neighbourhood U such that $f(U)$ is open and such that $f|_U$ is an embedding. Local homeomorphisms are also called **étale morphisms**.

Definition 7.2.13 (Covering space). Consider two topological spaces X, C 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 $\bigsqcup_i C_i$ of open sets in C where every set C_i is mapped homeomorphically onto U. The neighbourhoods U are sometimes said to be **evenly covered**.

Notation 7.2.14. 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 7.2.15 (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 7.2.16 (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 7.2.17 (Étalé space). Let X be a topological space. A topological space Y is called an étalé space over X if there exists a continuous surjection $\pi : Y \to X$ such that π is a local homeomorphism. The preimage $\pi^{-1}(x)$ of a point $x \in X$ is called the **stalk** of π over x.

Example 7.2.18. Every covering space is an étalé space.

Definition 7.2.19 (Pseudogroup ♣). Let X 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 open $U \subseteq X$.
2. If $\phi \in \mathcal{G}$, then $\phi^{-1} \in \mathcal{G}$.
3. If $V \subset U$ is open, 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 $\psi \circ \phi|_{\phi^{-1}(V\cap U')} \in \mathcal{G}$.

7.3 Associated constructions

Construction 7.3.1 (Product topology). First, consider the case with only a finite number of spaces $\{X_i\}_{i\in I}$. 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:

$$\mathcal{B} := \left\{ \prod_{i\in I} U_i \,\middle|\, U_i \in \tau_i \right\}. \tag{7.7}$$

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 7.3.2 (Disjoint union). Let $\{X_i\}_{i\in I}$ be a family of topological spaces and consider the disjoint union

$$X := \bigsqcup_{i\in I} X_i \tag{7.8}$$

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:

$$\tau_X := \big\{ U \subseteq X \,\big|\, \forall i \in I : \phi_i^{-1}(U) \text{ is open in } X_i \big\}. \tag{7.9}$$

Construction 7.3.3 (Quotient space). Consider a topological space X 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:

$$\tau_q := \big\{ U \subseteq X/Y \,\big|\, \pi^{-1}(U) \text{ is open in } X \big\}. \tag{7.10}$$

Remark 7.3.4 (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:

$$X/\emptyset = X \sqcup *. \tag{7.11}$$

Construction 7.3.5 (Wedge sum). Consider two pointed spaces $(X, x_0), (Y, y_0)$. 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 7.3.6 (Smash product). Consider two pointed topological spaces $(X, x_0), (Y, y_0)$. The smash product $X \wedge Y$ is defined as the quotient

$$X \wedge Y := (X \times Y)/(X \vee Y), \tag{7.12}$$

where $X \vee Y$ sits inside the product as the union of $X \times \{y_0\}$ and $\{x_0\} \times Y$.

Construction 7.3.7 (Suspension). Let X be a topological space. The suspension of X is defined as the following quotient space:

$$SX := (X \times [0,1])/\big\{(x,0) \sim (y,0) \text{ and } (x,1) \sim (y,1) \,\big|\, x, y \in X\big\}. \tag{7.13}$$

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:

$$\Sigma X := SX/(x_0 \times [0,1]). \tag{7.14}$$

An equivalent definition of the reduced suspension can be given in terms of the smash product:

$$\Sigma X = X \wedge S^1. \tag{7.15}$$

Example 7.3.8 (Spheres). Up to homeomorphisms the spheres are related by (reduced) suspensions:

$$SS^n \cong S^{n+1} \cong \Sigma S^n. \tag{7.16}$$

If one identifies the empty set with the (-1)-sphere, this relation can be continued to the case $n = -1$.

Construction 7.3.9 (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:

$$X \cup_f Y := (X \sqcup Y)/\{A \sim f(A)\}. \tag{7.17}$$

In categorical terms it is the pushout 4.4.59 in **Top** of the inclusion $\iota : A \hookrightarrow X$ along $f : A \to Y$.

Construction 7.3.10 (Join). Let $\{A_i\}_{i\leq n}$ be a finite collection of topological spaces. The join, denoted by $A = A_1 \circ \cdots \circ A_n$, is defined as follows. Every point of A is defined by the following data:

1. an element of the standard n-simplex 8.2.1, i.e. an n-tuple of nonnegative 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 $a_i \in A_i$.

This point in A is denoted by $t_1a_1 \oplus \cdots \oplus t_na_n$.

In the case of two spaces there exists a more intuitive construction. Let A, B be two topological spaces. The join $A \circ B$ is equal to the quotient space $(A \times B \times [0,1])/\sim$, where the relation $\sim$ is defined as follows:

- For all $a \in A$ and $b, b' \in B$: $(a, b, 0) \sim (a, b', 0)$.
- For all $a, a' \in A$ and $b \in B$: $(a, b, 1) \sim (a', b, 1)$.

This can be interpreted as collapsing one end of the cylinder $(A \times B) \times [0, 1]$ to A and the other end to B.

Property 7.3.11 (Monoidal structure ♣). The join induces a monoidal structure on the category **Top** where the tensor unit is given by the empty space $\emptyset$.

7.4 Connected spaces

Definition 7.4.1 (Connected space). A topological space that cannot be written as the disjoint union of two non-empty open sets. Equivalently, a space is connected if the only clopen sets are the empty set and the space itself.

Property 7.4.2 (Locally constant implies 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 7.4.3 (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 for every $c \in]a, b[: c \in f(X)$.*

Definition 7.4.4 (Path-connected space[3]). 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 $\varphi(1) = y$, then the space is said to be path-connected.

Property 7.4.5 (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.

Remark 7.4.6 (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 $\pi_0(X)$.

7.5 Compact spaces

7.5.1 Compactness

Definition 7.5.1 (Sequentially compact space). A topological space in which every sequence has a convergent subsequence (the sequence itself does not have to be convergent).

Definition 7.5.2 (Finite intersection property). A collection $\mathcal{F} \subseteq 2^X$ of subsets has the finite intersection property (FIP) if

$$\bigcap_{V \in \mathcal{F}'} V \neq \emptyset \tag{7.18}$$

for all finite $\mathcal{F}' \subset \mathcal{F}$.

Definition 7.5.3 (Locally finite cover). An open cover of a topological space X is said to be locally finite if every $x \in X$ has a neighbourhood that intersects only finitely many sets in the given cover.

Property 7.5.4 (First-countable spaces). A first-countable space is sequentially compact if and only if every countable open cover has a finite subcover.

Definition 7.5.5 (Lindelöf space). A space for which every open cover has a countable subcover.

Property 7.5.6. Every second-countable space is a Lindelöf space.

Definition 7.5.7 (Compact space). A topological space for which every open cover of has a finite subcover.

Theorem 7.5.8 (Heine-Borel[4]). *If a topological space X is sequentially compact and second-countable, every open cover has a finite subcover and, therefore, X is compact.*

Corollary 7.5.9 (Real numbers). A subset of $\mathbb{R}^n$ is compact if and only if it is closed and bounded.

[3] A similar notion is that of **arcwise-connectedness** where the function φ is required to be a homeomorphism.
[4] Also called the **Borel-Lebesgue** theorem.

Theorem 7.5.10 (Tychonoff's theorem). *Any product of compact topological spaces is compact under the (Tychonoff) product topology 7.3.1.*

Definition 7.5.11 (Relatively compact space). A topological space for which its closure is compact.

Definition 7.5.12 (Locally compact space). A topological space in which every point has a compact neighbourhood.

Theorem 7.5.13 (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 $f_n \longrightarrow f$ pointwise to a continuous function f, the convergence is uniform.*

Definition 7.5.14 (ω-bounded space). A topological space in which the closure of every countable subset is compact.

Definition 7.5.15 (Paracompact space). A topological space for which every open cover has a locally finite open refinement.

Property 7.5.16. Every paracompact Hausdorff space is normal.

Definition 7.5.17 (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 $\{f_i \mid \text{supp} f_i \cap U \neq \emptyset\}$ is finite.
2. **Normalization**: $\sum_i f_i = 1$.

Consider an open cover $\{V_i\}_{i\in I}$ of X. If there exists a partition of unity, also indexed by I, such that $\text{supp}(\varphi_i) \subseteq U_i$, then this partition of unity is said to be **subordinate** to the given cover.

Property 7.5.18 (Hausdorff spaces). A paracompact space is Hausdorff if and only if it admits a partition of unity subordinate to any open cover.

Definition 7.5.19 (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 7.5.20 (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

$$U^K := \{f : X \to Y \mid K \text{ compact}, U \text{ open and } f(K) \subseteq U\}. \tag{7.19}$$

Property 7.5.21 (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:

$$C(Z \times X, Y) \cong C(Z, C(X, Y)), \tag{7.20}$$

i.e. the mapping space $C(X, Y)$ is an internal hom 4.6.19 in the category **Top** and, because the topological product is the product in **Top**, $C(X, Y)$ is even an exponential object 4.6.22. For this reason the mapping spaces $C(X, Y)$ are also sometimes denoted by Y^X.

7.5.2 Compactifications

Definition 7.5.22 (Dense). A subset $V \subseteq X$ is said to be dense in a topological space X if $\overline{V} = X$.

Definition 7.5.23 (Separable space). A topological space that contains a countable, dense subset.

Property 7.5.24. Every second-countable space is separable.

Definition 7.5.25 (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 7.5.26. 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 7.5.27 (One-point compactification). Let X be a Hausdorff space. A one-point compactification or **Alexandrov compactification** is a compactification $\widehat{X}$ such that $\widehat{X} \setminus X$ is a singleton.

Example 7.5.28 (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*.

7.6 Uniform spaces

Definition 7.6.1 (Uniform structure). Consider a set X. A uniform structure on X consists of a collection $\mathfrak{U}$ of subsets $U \subseteq X \times X$ that satisfy the following properties:

1. If $U \in \mathfrak{U}$ and $U \subset V$, then $V \in \mathfrak{U}$.
2. If $U, V \in \mathfrak{U}$, then $U \cap V \in \mathfrak{U}$.
3. If $U \in \mathfrak{U}$, then $\Delta_X \subset U$.
4. If $U \in \mathfrak{U}$, there exists $V \in \mathfrak{U}$ such that $V \circ V = U$.
5. If $U \in \mathfrak{U}$, then $U^t \in \mathfrak{U}$.

The "transpose" U^t denotes the converse 2.2.5 of U and the composition $\circ$ is the relational composition 2.2.6 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 7.6.2. The first three conditions imply that a uniform structure is in particular a filter.

?? COMPLETE (Bourbaki) ??

7.7 Bornological spaces

Definition 7.7.1 (Bornology). Let X be a set. A bornology on 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 ??

7.8 Locales ♣

Property 7.8.1 (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:

$$V \cap \left(\bigcup_{i \in I} U_i \right) = \bigcup_{i \in I} (V \cap U_i). \tag{7.21}$$

This implies that the poset **Open**(X) is a frame 2.6.33.

Definition 7.8.2 (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 the opposite category:

$$\mathbf{Loc} := \mathbf{Frame}^{op}.$$

Construction 7.8.3 (From locale to topological space). There exists an adjunction

$$\mathbf{Loc} \underset{\text{Point}}{\overset{\iota}{\underset{\longrightarrow}{\overset{\longleftarrow}{\perp}}}} \mathbf{Top},$$

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 $\mathbf{Open}(X) \to 1 \equiv \Omega_{\text{Frame}} = \{0, 1\}$. Generalizing this to locales, one defines the set of points of L as the Ω_{Loc}-elements:
>
> $$\text{Point}(L) := \mathbf{Loc}(1, L).$$
>
> This set can be given a topology by declaring for every $U \in L$ the set $\{p \in \text{Point}(L) \mid p^{-1}(U) = 1\}$ to be open.

Definition 7.8.4 (Sober space). A topological space X such that the map $X \to \text{Point}(X)$ 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. Important examples are Hausdorff spaces.

?? COMPLETE ??

Chapter 8

Algebraic Topology

References for this chapter are [48, 125].

8.1 Homotopy theory

8.1.1 Homotopy

Definition 8.1.1 (Retraction). Let X be a topological space and let $A \subset X$ be a subspace. A continuous function $f : X \to A$ is called a retraction (and A is called a **retract** of X) if it satisfies $f(a) = a$ for all $a \in A$.[1]

Definition 8.1.2 (Homotopy). Let $f, g \in C(X, Y)$ where X, Y are topological spaces. If there exists a continuous function $H : X \times [0,1] \to Y$ such that $f(x) = H(x, 0)$ and $g(x) = H(x, 1)$, then f and g are said to be homotopic. 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 8.1.3 (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 8.1.4 (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 8.1.5 (Homeomorphisms). Every homeomorphism is a homotopy equivalence.

Definition 8.1.6 (Null-homotopic). A continuous function is said to be null-homotopic if it is homotopic to a constant function.

Definition 8.1.7 (Contractible space). A topological space X is said to be contractible if the identity map $\mathbb{1}_X$ is null-homotopic or, equivalently, if the space is homotopy-equivalent to a point.

Definition 8.1.8 (Good cover). Let X be a topological space with an open cover $\mathcal{U} = \{U_i\}_{i \in I}$. 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.

[1] It is a retraction of the inclusion map $A \hookrightarrow X$ in the sense of Definition 4.4.1.

[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.

Property 8.1.9 (Path space). Consider a topological space X. By Property 7.5.21 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 8.1.10 (Mapping class group). Consider a topological space X with some structure. The mapping class group of X is defined as the quotient

$$\mathrm{MCG}(X) := \mathrm{Aut}(X)/\,\mathrm{Aut}_0(X), \tag{8.1}$$

i.e. as the set of path-components of its automorphism group (path are in this case given by isotopies).

8.1.2 Homotopy groups

In this subsection it will always be assumed that the spaces are pointed 7.1.3. The generic base point will be denoted by $*$.

Definition 8.1.11 (Loop space). The set of all **loops** in a pointed topological space $(X, *)$, i.e. all continuous functions $\delta : (S^1, t_0) \to (X, *)$ for which $\delta(t_0) = *$. This space is denoted by ΩX. It can be turned into a topological space by equipping it with the *compact-open topology*.

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 8.1.12 (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 even becomes a topological group.

Remark 8.1.13 (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 4.12.7), i.e. a group up to homotopy.

Definition 8.1.14 (Fundamental group). The fundamental group $\pi_1(X, x_0)$ is defined as the loop space of (X, x_0) modulo homotopy, i.e. $\pi_1(X) := \pi_0(\Omega X)$ where π_0 denotes the set of path components 7.4.6. 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 8.1.15. 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 8.1.16 (Groups). The fundamental group of a topological group is Abelian. This follows from an Eckmann-Hilton argument 3.1.4.

Definition 8.1.17 (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
- **Morphisms**: The endpoint-preserving homotopy classes of continuous functions $f : S^1 \to X$.

[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 $S^1 \cong [0,1]/_{0\sim 1}$.

The fundamental group $\pi_1(X, x)$ can be recovered as the automorphism group of $x \in \mathrm{ob}(\mathbf{\Pi}_1(X))$.

Definition 8.1.18 (Simply-connected space). A topological space is said to be simply-connected if it is path-connected and if the fundamental group is trivial.

Definition 8.1.19 (Semilocally simply connected). A topological space is said to be semilocally simply connected if every point admits a neighbourhood U such that every loop in U can be contracted to a point.

Definition 8.1.20 (Universal covering space). A covering space 7.2.13 is said to be universal if it is simply-connected.

Universal Property 8.1.21. Let X be a topological space and let $\widetilde{X}$ be its the universal covering space. Every other covering space C of X is also covered by $\widetilde{X}$.

Property 8.1.22 (Automorphisms). Consider a topological space X and let $\widetilde{X}$ be its universal covering space. The group of deck transformations $\mathrm{Aut}(\widetilde{X})$ from Definition 7.2.16 is isomorphic to the fundamental group $\pi_1(X)$. Hence, one obtains

$$X \cong \widetilde{X}/\pi_1(X). \tag{8.2}$$

Property 8.1.23 (Existence). If the space is connected, locally path-connected and semilocally simply connected, the 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):

$$\widetilde{X} = \{[\gamma] \mid \gamma : [0, 1] \to X, \gamma(0) = x_0\}, \tag{8.3}$$

where x_0 is the base point of x_0 and $[\gamma]$ denotes the homotopy class of γ.

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 8.1.24 (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 7.4.6). This explains why the notation $\pi_0(X)$ was introduced before.

Property 8.1.25. 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 3.1.4.

Remark 8.1.26 (Relative homotopy groups). As for (co)homology 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 8.1.27 (Path-connectedness). If X is path-connected, the homotopy groups for different basepoints are isomorphic.

Property 8.1.28 (Homeomorphisms). Homeomorphic spaces have isomorphic homotopy groups.

Formula 8.1.29 (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

$$\pi_n(X \times Y, (x_0, y_0)) = \pi_n(X, x_0) \times \pi_n(Y, y_0). \tag{8.4}$$

Property 8.1.30 (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 8.1.31 (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 8.1.32 (n-connected space). A topological space is said to be n-connected if it is path-connected and if its first 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.

Definition 8.1.33 (Homotopy n-type). A topological space for which the homotopy groups π_i vanish for $i > n$.

Definition 8.1.34 (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:

$$\pi_i\Big(K(G,n)\Big) = \begin{cases} G & i = n \\ 0 & i \neq n. \end{cases} \tag{8.5}$$

It follows from Property 8.1.25 above that for $n > 1$ the group G has to be Abelian.

Property 8.1.35 (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 8.1.36 (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 8.1.37 (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 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 (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

$$\cdots \longrightarrow X_2 \longrightarrow X_1 \longrightarrow X \tag{8.6}$$

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 the space X_n is n-connected.

[4]Often called the **Moore-Postnikov tower** or **Postnikov system** (especially in category theory).

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 8.1.38 (Homotopy hypothesis ♣). **Top** *and* ∞**Grpd** *are equivalent as* $(\infty, 1)$-*categories. In particular this means that* n*-groupoids are equivalent to homotopy* n*-types.*

Remark 8.1.39. The statement of the above 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 6.

Property 8.1.40 (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 : **Top** $\to$ **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 12.2.5 of **Top** with respect to the collection of weak homotopy equivalences. (See Section 12.3 for more information.)

Recall the reduced suspension functor Σ from Definition 7.3.7. The functors Σ and Ω are related in the following way:

Property 8.1.41 (Eckmann-Hilton duality). The reduced suspension functor Σ and the loop space functor Ω form an adjunction in the category of pointed topological spaces:

$$\mathrm{Map}_*(\Sigma X, Y) \cong \mathrm{Map}_*(X, \Omega Y). \tag{8.7}$$

This also passes down to an equivalence in the associated homotopy category:

$$[\Sigma X, Y] \cong [X, \Omega Y]. \tag{8.8}$$

Corollary 8.1.42. By choosing $X = S^n$ and using the result from Example 7.3.8, one obtains the following important result for homotopy groups:

$$\pi_{n+1}(Y) \cong \pi_n(\Omega Y). \tag{8.9}$$

Theorem 8.1.43 (Freudenthal suspension theorem). *The suspension morphism in homotopy*

$$\pi_{n+k}(S^n) \to \pi_{n+k+1}(S^{n+1}) \tag{8.10}$$

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 8.1.44 (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 (Chapter 23). There it is called the **Schauder fixed point theorem**.

Theorem 8.1.45 (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 8.1.46 (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 8.1.47. The theorem is even stronger in that one can replace the volume by any other measure (See Chapter 16).

The Borsuk-Ulam theorem is equivalent to the following purely point-set theoretic theorem:

Theorem 8.1.48 (Lusternik-Schnirelmann). *For every closed cover $\{V_i\}_{1\leq i\leq n+1}$ of S^n there exists an index i such that V_i contains a pair of antipodal points, i.e. $V_i \cap (-V_i) \neq \emptyset$.*

8.1.3 CW complexes

Definition 8.1.49 (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 7.3.9.

Definition 8.1.50 (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 8.1.51 (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 8.1.52 (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 8.1.53. 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 8.1.54. Infinite-dimensional CW complexes can be obtained by taking the direct limit 3.7.2 of the sequence above.

Theorem 8.1.55 (Whitehead). *A continuous function between CW complexes is a homotopy equivalence if and only if it is a weak homotopy equivalence.*

Theorem 8.1.56 (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 8.1.57 (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 8.1.43 can be generalized to CW complexes:

Theorem 8.1.58 (Freudenthal suspension theorem). *If X is n-connected, the suspension morphism*

$$\pi_k(X) \to \pi_{k+1}(\Sigma X) \tag{8.11}$$

is an isomorphism for all $k \leq 2n$.

8.1.4 Fibrations

Definition 8.1.59 (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 $\widetilde{f}_0 : X \to E$ of $f_0 := f|_{X\times\{0\}}$ there exists a homotopy $\widetilde{f} : X \times [0,1]$ lifting f such that Diagram 8.1 commutes.

$$\begin{tikzcd}
X \arrow[r, "\widetilde{f}_0"] \arrow[d, hook, "X\times\{0\}"'] & E \arrow[d, "\pi"] \\
X \times [0,1] \arrow[r, "f"'] \arrow[ur, dashed, "\widetilde{f}"] & B
\end{tikzcd}$$

Figure 8.1: Homotopy lifting property.

Definition 8.1.60 (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 8.1.61 (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 8.1.62 (Hopf fibration). The Hopf fibration is given by

$$S^1 \hookrightarrow S^3 \to S^2. \tag{8.12}$$

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 20.7.5)

Example 8.1.63. For all $n \in \mathbb{N}$ the following sequence forms a fibration:

$$\mathrm{SO}(n) \hookrightarrow \mathrm{SO}(n+1) \to S^n. \tag{8.13}$$

$$\begin{tikzcd}
Y & X \arrow[l, "\widetilde{f}_0"'] \arrow[dl, dashed, "\widetilde{f}"'] \\
Y^{[0,1]} \arrow[u, two heads, "\pi_0"] & A \arrow[l, "f"] \arrow[u, "\iota"']
\end{tikzcd}$$

Figure 8.2: Homotopy extension property.

Definition 8.1.64 (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 $\widetilde{f}_0 : X \to Y$ of $f_0 := f|_{A\times\{0\}}$ there exists a homotopy $\widetilde{f} : X \times [0,1] \to Y$ extending f such that Diagram 8.2 commutes, where Property 7.5.21 is used to represent homotopies in terms of the path space of Y.

Definition 8.1.65 (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 8.1.66 (Mapping path space). Consider a continuous function $f : X \to Y$. The mapping path space P_f is defined as the following pullback:

$$P_f := \big\{(x, p) \in X \times Y^{[0,1]} \,\big|\, f(x) = p(0)\big\}. \tag{8.14}$$

The projection

$$\pi : P_f \to Y : (x, p) \mapsto p(1) \tag{8.15}$$

is a fibration. The homotopy type of the fibres of this fibration is called the **homotopy fibre** or **mapping fibre** of f.

Construction 8.1.67 (Mapping cylinder). Let $f : X \to Y$ be a continuous function. The mapping cylinder M_f or $\mathrm{Cyl}(f)$ is defined as follows:

$$M_f := \Big(X \times [0,1] \bigsqcup Y\Big) / \sim_f, \tag{8.16}$$

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 $\mathrm{Cone}(f)$. 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 8.1.68 (Pushouts ♣). Just like the attaching space 7.3.9 could be obtained as a pushout in **Top**, one can also characterize the mapping cone as a pushout. The cone $\mathrm{Cone}(X)$ can be obtained as the pushout of the span $\{*\} \leftarrow X \overset{\iota_0}{\to} X \times [0,1]$. The mapping cone is then obtained as the pushout of the span $\mathrm{Cone}(X) \leftarrow X \times [0,1] \to M_f$. By the *pasting law* for pushouts this implies that the mapping cone of f can be obtained as the pushout

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d] \arrow[dr, phantom, "\mathrm{po}"] & Y \arrow[d] \\ \mathrm{Cone}(X) \arrow[r] & \mathrm{Cone}(f) \end{tikzcd}$$

8.1.5 Rational homotopy theory ♣

For the theory on differential graded algebras, see Chapter 27.

Definition 8.1.69 (Rational space). A simply connected topological space X for which the homotopy groups $\pi_n(X)$ are rational vector spaces.

Definition 8.1.70 (Rational homotopy equivalence). A continuous function $f : X \to Y$ for which the induced maps on rational homotopy groups

$$\pi_n(f) \otimes \mathbb{Q} : \pi_n(X) \otimes \mathbb{Q} \to \pi_n(Y) \otimes \mathbb{Q} \tag{8.17}$$

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 8.1.71 (Rational homotopy category). Consider the category **Top** of topological spaces. The rational homotopy category is obtained as the localization 12.2.5 of **Top** with respect to the collection of rational homotopy equivalences.

Definition 8.1.72 (Polynomial differential forms). Consider the standard n-simplex Δ^n (Definition 8.2.1). On this topological space one can define differential forms analogous to those from Section 32.4. Let $\{t_i\}_{0\leq i\leq n}$ be $n+1$ generators in degree 0 (the barycentric coordinates). Together with $n+1$ associated generators dt_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:

$$\sum_{i=0}^n t_i = 1 \qquad \text{and} \qquad \sum_{i=0}^n dt_i = 0. \tag{8.18}$$

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 dt_i are identified with the de Rham differentials of the barycentric coordinates.

Morphisms $f : [m] \to [n]$ in the simplex category $\mathbf{\Delta}$ (Definition 12.1.1) 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:

$$F(t_i) := \sum_{f(j)=i} t_j. \tag{8.19}$$

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 $\Omega^\bullet_{\text{poly}} : \mathbf{sSet} \to \mathbf{dgcAlg^{op}}$. Composition with the singular set functor $\mathrm{Sing} : \mathbf{Top} \to \mathbf{sSet}$ gives the **piecewise-polynomial differential forms** functor $\Omega^\bullet_{\text{pwpoly}}$.

Definition 8.1.73 (Relative Sullivan algebra). An inclusion of DGCAs of the form

$$(A,d) \hookrightarrow (A \otimes \Lambda^\bullet V, d'),$$

where A is any DGCA and 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

  $$d'e_k \in A \otimes \Lambda^\bullet V(< k), \tag{8.20}$$

  where $V(< k) := \mathrm{span}\{e_i\}_{i<k}$.

If in addition the implication

$$i < j \implies \deg(e_i) \leq \deg(e_j) \tag{8.21}$$

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 8.1.74. The minimality condition admits an (equivalent) reformulation:

$$dV \subseteq A_{\geq 1} \otimes \Lambda^\bullet V + A \otimes \Lambda^{\geq 2} V. \tag{8.22}$$

Example 8.1.75. Consider the free DGCA $\Lambda(V)$ on a graded vector space V of finite type with $V_0 = V_1 = 0$. Then $(\Lambda(V), d)$ is a Sullivan algebra.

Definition 8.1.76 (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 8.1.77. Minimal Sullivan models are unique up to isomorphism.

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 (Section 27.7.2).

Construction 8.1.78 (Sketch, Quillen). To every simply-connected topological space X one can associated 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.

8.1.6 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 8.1.79 (Fixed-point space). Consider a topological G-space X. The set of all its fixed points is denoted by X^G.

Definition 8.1.80 (Equivariant homotopy equivalence). A continuous function that is both an equivariant function and a homotopy equivalence in the ordinary sense.

Definition 8.1.81 (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 8.1.82 (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 8.1.83 (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 8.1.84 (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 **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 **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*

$$\mathbf{Ho}(G\mathbf{Top}) \cong \mathbf{Psh}(\mathbf{Orb}_G). \tag{8.23}$$

?? COMPLETE ??

8.2 Simplicial homology

8.2.1 Simplices

Definition 8.2.1 (Simplex). A k-simplex $\sigma^k \equiv [t_0, \ldots, t_k]$ is defined as the following set:

$$\sigma^k := \left\{ \sum_{i=0}^k \lambda_i t_i \,\middle|\, \sum_{i=0}^k \lambda_i = 1 \text{ and } \lambda_i \geq 0 \right\}. \tag{8.24}$$

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 $\{t_0, \ldots, t_k\}$.

Remark 8.2.2 (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 8.2.3 (Standard simplex).

$$\Delta^k := \left\{ (x_0, \ldots, x_k) \in \mathbb{R}^{k+1} \,\middle|\, \sum_i x_i = 1 \text{ and } x_i \geq 0 \right\} \tag{8.25}$$

This simplex is also called the **probability simplex** or sometimes the **unit simplex**.

Notation 8.2.4 (Face). Consider a k-simplex $[v_0, \ldots, v_k]$. The **face** opposite to the vertex v_i is the $(k-1)$-simplex $[v_0, \ldots, \widehat{v}_i, \ldots, v_k]$ obtained by removing the vertex v_i.

Definition 8.2.5 (Simplicial complex). A simplicial complex $\mathcal{K}$ is a set of simplices satisfying the following conditions:

[5]This equivariant version of Theorem 8.1.55 is due to *Bredon*.
[6]This is the restriction of the Yoneda embedding to the subcategory $\mathbf{Orb}_G$ of $G\mathbf{Top}$.

- 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.

Definition 8.2.6 (Path-connected complex). A simplicial complex in which every two vertices are connected by an edge.

Definition 8.2.7 (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 8.2.8 (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 8.2.9 (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 3.2.47 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 8.2.10. Property 8.1.28, 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 its triangulations.

8.2.2 Simplicial homology

In this section a chain complex 5.1.1 and its associated homology theory are constructed for triangulable spaces.

Definition 8.2.11 (Chain group). Let $\mathcal{K}$ be a simplicial n-complex. The k^{th} chain group $C_k(\mathcal{K})$ is defined as the free Abelian group generated by the k-simplices in $\mathcal{K}$:

$$C_k(\mathcal{K}) := \left\{ \sum_i a_i \sigma_i \,\middle|\, \sigma_i \text{ is a } k\text{-simplex in } \mathcal{K}, a_i \in \mathbb{Z} \right\}. \tag{8.26}$$

For $k > n$ and $k < 0$, $C_k(\mathcal{K})$ is defined to be $\{0\}$.

Definition 8.2.12 (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**:

$$\partial_k \left(\sum_i a_i \sigma_i \right) = \sum_i a_i \partial_k \sigma_i, \tag{8.27}$$

- **Action on generators**

$$\partial_k [v_0, \ldots, v_k] = \sum_{i=0}^{k} (-1)^i [v_0, \ldots, \hat{v}_i, \ldots, v_k], \tag{8.28}$$

$$\partial_0 [v] = 0. \tag{8.29}$$

Property 8.2.13 (Chain condition). The boundary operators satisfy the following relation:

$$\partial_k \circ \partial_{k+1} = 0. \tag{8.30}$$

This property turns the system (C_k, ∂_k) into a chain complex 5.1.1.

Definition 8.2.14 (Cycle group). The k^{th} 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 8.2.15 (Boundary group). The k^{th} boundary group $B_k(\mathcal{K})$ is defined as the set of k-chains σ_k for which there exists a $(k+1)$-chain N such that $\partial_{k+1} N = \sigma_k$. These chains are called **boundaries**.

Definition 8.2.16 (Homology group). From Property 8.2.13 it follows that $B_k(\mathcal{K}) \subset Z_k(\mathcal{K})$ is a subgroup. One can thus define the k^{th} homology group $H_k(\mathcal{K})$ as the following quotient group:

$$H_k(\mathcal{K}) := Z_k(\mathcal{K})/B_k(\mathcal{K}). \tag{8.31}$$

Theorem 3.2.55 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 8.2.17 (Betti numbers). The ranks $R_k(\mathcal{K})$ are called the Betti numbers of $\mathcal{K}$.

Definition 8.2.18 (Euler characteristic). The Euler characteristic of a triangulable space X is defined as follows:

$$\chi(X) := \sum_i (-1)^i R_i(X). \tag{8.32}$$

This formula is sometimes called the **Poincaré** or **Euler-Poincaré** formula.

Property 8.2.19 (Isomorphisms). If two topological spaces have the same homotopy type, in particular when they are homeomorphic, they have isomorphic homology groups.

Corollary 8.2.20. 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 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 8.2.21 (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 29.1.9).

Construction 8.2.22. 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:

$$c = \sum_i g_i \sigma_i^k, \tag{8.33}$$

where $g_i \in G$ for some Abelian group G. The k^{th} homology group of X with coefficients in G is denoted by $H_k(X; G)$.

Property 8.2.23 (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 8.2.24 (Künneth formula). Let X, Y be two simplicial complexes. The homology groups of the Cartesian product $X \times Y$ with coefficients in a field F is given by

$$H_k(X \times Y; F) = \bigoplus_{k=i+j} H_i(X; F) \otimes H_j(Y; F). \tag{8.34}$$

When the requirement of F 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.

8.2.3 Relative homology

In this section the homology of a simplicial complex K "modulo" a subcomplex L is considered.

Definition 8.2.25 (Relative chain group). The k-chain group of K modulo L is defined as the following quotient group:

$$C_k(K, L) := C_k(K)/C_k(L). \tag{8.35}$$

Equivalence classes will be denoted by square brackets $[c]$, where $c \in C_\bullet(K)$.

Definition 8.2.26 (Relative boundary operator). The relative boundary operator $\overline{\partial}_k$ is defined as follows:

$$\overline{\partial}_k[c_k] := [\partial_k c_k], \tag{8.36}$$

where ∂_k on the right-hand side denotes the boundary operator in ordinary homology. This is a well-defined operation since $\partial C_\bullet(L) \subseteq C_\bullet(L)$.

Definition 8.2.27 (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:

$$H_k(K, L) := \frac{\ker(\overline{\partial}_k)}{\operatorname{im}(\overline{\partial}_{k+1})}. \tag{8.37}$$

Elements $h_k \in H_k(K, L)$ can be represented as $h_k = z_k + C_k(L)$ such that $\partial_k z_k \in C_{k-1}(L)$.

Property 8.2.28 (Homotopy invariance). Consider two 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 8.2.29 (Long sequence for a pair). The relative homology groups fit in the following (long) exact sequence:

$$\cdots \longrightarrow H_k(L) \xrightarrow{i_*} H_k(K) \xrightarrow{j_*} H_k(K, L) \xrightarrow{\partial_k} H_{k-1}(L) \longrightarrow \cdots, \tag{8.38}$$

where i_* and j_* are the homology morphisms induced by the inclusions $i : L \to K$ and $j : K \to (K, L)$.

Theorem 8.2.30 (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*

$$H_k(X, V) \cong H_k(X \backslash U, V \backslash U). \tag{8.39}$$

Definition 8.2.31 (Reduced homology). Consider the chain complex $C_\bullet(X)$. Example 8.2.34 says that a point $\{*\}$ has homology $\mathbb{Z}$ concentrated in degree 0. However, it would be nice if the point $*$ had vanishing homology. To this end one can augment 5.3.2 the chain complex $C_\bullet(X)$ of any topological space by the group $\mathbb{Z}$ to obtain the reduced homology complex $\widetilde{H}_\bullet(X)$:

$$H_n(X) \cong \begin{cases} \widetilde{H}_0(X) \oplus \mathbb{Z} & n = 0 \\ \widetilde{H}_n(X) & n \neq 0. \end{cases} \tag{8.40}$$

Property 8.2.32. 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 8.2.33 (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 8.1.65.

Given an NDR pair (X, A), there exists an isomorphism

$$H_\bullet(X, A) \cong \widetilde{H}_\bullet(X/A). \tag{8.41}$$

8.2.4 Examples

Example 8.2.34. Let X be a contractible space.

$$H_k(X) = \begin{cases} \mathbb{Z} & k = 0 \\ 0 & \text{otherwise} \end{cases} \tag{8.42}$$

Example 8.2.35. Let P be a path-connected simplicial complex.

$$H_0(P) = \mathbb{Z} \tag{8.43}$$

Furthermore, every point $p \in P$ determines a generator $\langle p \rangle \in H_0(P)$.

Example 8.2.36. The homology groups of the n-sphere S^n are given by

$$H_k(S^n) = \begin{cases} \mathbb{Z} & k = 0 \text{ or } k = n \\ 0 & \text{otherwise.} \end{cases} \tag{8.44}$$

Definition 8.2.37 (Homology sphere). A n-dimensional manifold having the same homology groups as the n-sphere.

Definition 8.2.38 (Degree). The example above says that $H_n(S^n) = \mathbb{Z}$. 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 $f(x) = dx$ where $d \in \mathbb{Z}$. This coefficient is called the degree of f.

Property 8.2.39. Two maps $f : S^n \to S^n$ have the same degree if and only if they are homotopic.

Example 8.2.40. Consider the closed (or open) disks D_n.

$$H_k(D_n, \partial D_n) = \begin{cases} \mathbb{Z} & k = n \\ 0 & \text{otherwise} \end{cases} \tag{8.45}$$

This holds for any n-skeleton X^n of a CW-complex.

[7] Some people call this a **good pair**, while others define an NDR pair more restrictively.

8.3 Singular homology

Definition 8.3.1 (Singular simplex). Recall the **standard** k-simplex Δ^k from Example 8.2.3. 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 8.3.2 (Δ-complex). Let $\{\sigma_i : \Delta^n \to X\}_{i\in I}$ be a collection of singular simplices in X, where the dimension n may depend on the index $i \in I$. This collection forms a Δ-complex on X if it satisfies the following conditions:

- The restrictions $\sigma_i|_{\overset{\circ}{\Delta}{}^n}$ 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 set in X is open if and only if it all of its inverse images σ_i^{-1} 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 8.3.3 (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 8.3.4 (Face map). The face maps are morphisms $\varepsilon_i^k : \Delta^{k-1} \to \Delta^k$ that map Δ^{k-1} onto the i^{th} face of Δ^k. They are explicitly given by

$$\varepsilon_i^k(s_0, \dots, s_{k-1}) := (s_0, \dots, s_{i-1}, 0, s_i, \dots, s_{k-1}). \tag{8.46}$$

Their defining property is the following relation:

$$\varepsilon_i^k \circ \varepsilon_j^{k-1} = \varepsilon_j^k \circ \varepsilon_{i-1}^{k-1}, \tag{8.47}$$

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 a dual picture (they work in the opposite of the simplex category $\mathbf{\Delta}$).

Definition 8.3.5 (Singular boundary operator). The singular boundary operator ∂ (the same notation as for simplicial boundary operators is used for simplicity) 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

$$\partial_k\sigma^k = \sum_{i=0}^{k}(-1)^i\sigma^k \circ \varepsilon_i^k, \tag{8.48}$$

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:

$$\partial_{k-1} \circ \partial_k = 0. \tag{8.49}$$

It follows that $S_k(X; G)$ is also a chain complex.

Definition 8.3.6 (Singular homology group). The singular homology groups of a topological space with coefficients in an Abelian group G are defined as follows:

$$H_k(X;G) := \frac{\ker(\partial_k)}{\mathrm{im}(\partial_{k+1})}. \tag{8.50}$$

Property 8.3.7 (Simplicial homology). For triangulable spaces singular homology is isomorphic to simplicial homology. When X is not triangulable, this property is not valid. The singular approach to homology is strictly more general, but it is often more difficult to compute (even in the case of triangulable spaces).

Property 8.3.8 (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:

$$f_*\left(\sum_\sigma c_\sigma \sigma\right) := \sum_\sigma c_\sigma f \circ \sigma. \tag{8.51}$$

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**:

$$f_* : H_k(X) \to H_k(Y) : \langle h\rangle \mapsto \langle f_*(h)\rangle. \tag{8.52}$$

Corollary 8.3.9. 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 8.3.10 (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 $h : \pi_1(X) \to H_1(X) : [\gamma] \mapsto \langle\gamma\rangle$ defines a group morphism. Furthermore, this map induces an isomorphism $h' : \pi_1(X)/[\pi_1(X), \pi_1(X)] \to H_1(X)$.*

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 $k \le n$ this morphism is in fact an isomorphism.

Definition 8.3.11 (Singular cohomology). The singular cohomology groups of a topological space X with coefficients in an Abelian group G are defined as follows:[8]

$$H^k(X;G) := \mathrm{Hom}(H_k(X), G), \tag{8.53}$$

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:

$$f^*g(\sigma) := g(f_*\sigma), \tag{8.54}$$

where f_* is the pushforward 8.3.8 induced by f.

Property 8.3.12 (Representability). Let X be a CW complex. There exists an isomorphism

$$[X, K(G,n)] \to H^n(X;G) \tag{8.55}$$

between the homotopy classes of maps $X \to K(G,n)$ and the n^{th} singular cohomology of X with coefficients in G. (This result can be widely generalized cf. Theorem 8.5.3 further below.)

Proof. Consider the Eilenberg-Maclane space $K(G,n)$ for G (Definition 8.1.34). By the Hurewicz theorem there exists an isomorphism $H_n(K(G,n)) \cong \pi_n(K(G,n)) \cong G$. By definition of cohomology $H^n(K(G,n);G) = \mathrm{Hom}(H_n(K(G,n)), G)$ and thus $H^n(K(G,n);G) \cong \mathrm{Hom}(G,G)$. 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 $f^*\psi$.

[8] Actually this is the *Universal Coefficient Theorem* when defining cohomology through a suitable hom-complex.

8.4 Local coefficients

For some applications one needs to generalize the notion of (co)homology. For example, when considering integration theory on manifolds (Section 32.8.3), one in general needs to take into account the nonorientability.

Consider a topological space X that admits a universal cover $\widetilde{X}$. By Property 8.1.22 one then has

$$X \cong \widetilde{X}/\pi_1(X). \tag{8.56}$$

Given a $\mathbb{Z}[\pi_1(X)]$-module A, one defines the A-twisted chain (singular) group of X as

$$C(X;A) := C(\widetilde{X}) \otimes_{\mathbb{Z}[\pi_1(X)]} A. \tag{8.57}$$

$\pi_1(X)$, the automorphism group 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).

8.5 Axiomatic approach ♣

Definition 8.5.1 (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 $H_k : \mathbf{Top} \times \mathbf{Top} \to \mathbf{Ab}$ (technically one should replace $\mathbf{Top} \times \mathbf{Top}$ by the category consisting of subspace inclusions $A \hookrightarrow X$). The axioms are as follows:

1. **Homotopy invariance**: If f, g are homotopic maps, their induced homology maps are the same:
$$f \cong g \implies \forall k \in \mathbb{N} : H_k(f) = H_k(g).$$
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
$$\cdots \longrightarrow H_k(A) \xrightarrow{i_*} H_k(X) \xrightarrow{j_*} H_k(X,A) \xrightarrow{\partial_k} H_{k-1}(A) \longrightarrow \cdots, \tag{8.58}$$
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 8.5.2. 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 8.5.3 (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* **Ho**(**Top**) *come from homotopy pushouts in* **Top**.*)*

Remark 8.5.4. If one constructs the homotopy category using the CW-model structure in **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 8.5.5 (Cohomology). Every (generalized) cohomology theory is 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

$$H^n(Y) := [Y, X_n] \tag{8.59}$$

for all $n \in \mathbb{N}$.

For reduced cohomology theories there exist suspension isomorphisms

$$\widetilde{H}^n(Y) \cong \widetilde{H}^{n+1}(\Sigma Y). \tag{8.60}$$

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 8.5.6 (Spectrum[9]). A sequence of pointed topological spaces $(X_n)_{n\in\mathbb{N}}$ such that $X_n \cong \Omega X_{n+1}$. (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:

$$(LX)_n := \varinjlim_{k\in\mathbb{N}} \Omega^k X_{n+k}. \tag{8.61}$$

This gives a *fibrant replacement functor* in the *model structure* on spectra.

Property 8.5.7. Brown's representability theorem shows that there exists a bijection between isomorphism classes of (generalized) cohomology theories and homotopy-equivalence classes of spectra.

Property 8.5.8 (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 12.1.17 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 8.5.9. There exists an (∞-)equivalence between **Sp** and ∞**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

$$\Omega^\infty X := X_0 \cong \varinjlim_{k\in\mathbb{N}} \Omega^k X_k. \tag{8.62}$$

[9]Sometimes called an Ω-spectrum to distinguish it from other (possibly more general) spectra.

8.6 Equivariant cohomology

In this section topological spaces equipped with a continuous action of a topological group G will be considered. These spaces will be called topological G-spaces or just G-spaces.

Definition 8.6.1 (Equivariant cohomology). Let X be a topological G-space for which the G-action is free. The equivariant cohomology of X is defined as

$$H_G^\bullet(X) := H^\bullet(X/G), \tag{8.63}$$

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 33.2.1. 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 $X/\!\!/G$. The G-equivariant cohomology of X is then defined as the singular cohomology of the Borel construction:

$$H_G^\bullet(X) := H^\bullet(EG \times_G X). \tag{8.64}$$

?? COMPLETE (E.G. LECTURES OF TU ON YOUTUBE) ??

Chapter 9

Sheaf Theory

A reference aimed towards the study of differential geometry is [46]. For the concept of a sheaf in category theory, see Section 13.4.

9.1 Presheaves

Definition 9.1.1 (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 $\Phi^U_V:\mathcal{F}(U)\rightarrow\mathcal{F}(V)$ 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. $W\subseteq V\subseteq U\implies\Phi^U_W=\Phi^V_W\circ\Phi^U_V$.

The set $\mathcal{F}(U)$ is called the set of **sections** over U and the morphisms Φ^U_V are called the **restriction maps**.

Definition 9.1.2 (Morphism of presheaves). Let $\mathcal{F},\mathcal{F}'$ be two presheaves on a topological space X. A morphism $\mathcal{F}\rightarrow\mathcal{F}'$ is a set of morphisms $\Psi_U:\mathcal{F}(U)\rightarrow\mathcal{F}'(U)$ that commute with the restriction maps Φ^U_V.

Alternative Definition 9.1.3 (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 $\mathcal{F}:\mathbf{Open}(X)\rightarrow\mathbf{C}$. 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 9.1.4 (Constant presheaf). Let S be any set. The constant presheaf on X with target S is defined by

$$\mathcal{F}(U):=S \tag{9.1}$$

for every open set $U\subseteq X$.

9.2 Sheaves

Definition 9.2.1 (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 9.2.2 (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 13.4.10.)

Notation 9.2.3 (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 9.2.4 (Sheaf topos ♣). The category of sheaves $\mathbf{Sh}(X)$ is in fact an elementary topos, called the sheaf topos on X. See Property 13.4.26.

Property 9.2.5. 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:

$$\mathcal{F}(U) \to \prod_{i\in I} \mathcal{F}(U_i) \rightrightarrows \prod_{i,j\in I} \mathcal{F}(U_i \cap U_j). \tag{9.2}$$

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 9.2.6 (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 2.6.12 by equipping it with the (partial) order relation

$$U \geq V \implies U \subseteq V.$$

This turns the sheaf $\mathcal{F}$ on X into a directed system. The stalk at x is defined as the following direct limit 3.7.2:

$$\mathcal{F}_x := \varinjlim_{U\ni x} \mathcal{F}(U). \tag{9.3}$$

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 9.2.7. 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 9.2.8. 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 9.2.9 (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 9.2.10 (Sheaf of sections). Consider a continuous function $f : X \to Y$. This function induces a sheaf on Y in the following way:

$$\mathcal{F}(U) := \{s : U \to X \mid f \circ s = \mathbb{1}_U\}. \tag{9.4}$$

It is the sheaf that assigns to every open set the local sections of f in the sense of Definition 4.4.1.

Construction 9.2.11 (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 a presheaf

$$\overline{\mathcal{F}}(U) := \left\{ (s_x)_{x\in U} \in \prod_{x\in U} \mathcal{F}_x \,\middle|\, \forall x \in U : \exists \text{ open } V \ni x, t \in \mathcal{F}(V) : \forall v \in V : s_v = \Phi_v^V(t) \right\}. \tag{9.5}$$

Sections of this sheaf are said to be **continuous**. This statement can be made formal using the concept of an étalé space (see Construction 9.2.15 further on). The restriction maps ρ_V^U are defined as follows:

$$\rho_V^U : (s_x)_{x\in U} \mapsto (s_x)_{x\in V}. \tag{9.6}$$

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

$$\varphi(s) : U \to \prod_{x\in U} \mathcal{F}_x : x \mapsto s_x = \Phi_x^U(s), \tag{9.7}$$

where $s \in \mathcal{F}(U)$ and $x \in U$, takes image in $\overline{\mathcal{F}}(U)$.

Universal Property 9.2.12. 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 9.2.13 (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 9.2.14. 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 9.2.15 (Étalé spaces). Let $\mathcal{F}$ be a presheaf on X and consider the disjoint union

$$\mathcal{F}^* := \bigsqcup_{x\in X} \mathcal{F}_x. \tag{9.8}$$

Define for every local section $s \in \mathcal{F}(U)$ a function $\overline{s} : U \to \mathcal{F}^* : x \mapsto s_x \in \mathcal{F}_x$. The union $\mathcal{F}^*$ can be turned into an étalé space 7.2.17 over X by equipping it with the topology with basis

$$\big\{\overline{s}(U) \mid U \text{ open in } X, s \in \mathcal{F}(U)\big\}. \tag{9.9}$$

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 9.2.16 (Paracompact spaces). Let X be a paracompact space 7.5.15 and consider a sheaf F of Abelian groups on X. For every closed subset $V \subset X$ there exists an isomorphism

$$\varinjlim_{U\supset K} F(U) \cong F(K), \tag{9.10}$$

where $F(K)$ is defined as the set of sections of the restriction of the étalé 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 9.2.17 (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 $F(V) \rightarrow F(U)$ is surjective.

Definition 9.2.18 (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 9.2.19 (Abstract nonsense). Consider the equalizer (9.2). To every presheaf $\mathcal{F}$ one can assign a separated presheaf $\mathcal{F}^{\#}$ by defining $\mathcal{F}^{\#}(U)$ as the direct limit 3.7.2 of the equalizers over all open covers of U ordered by refinement. A sheaf is obtained by applying this construction a second time.

Example 9.2.20 (Constant sheaf). Consider the constant presheaf on X with target S (Example 9.1.4). 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 $f : U \rightarrow S$.

9.3 Abelian sheaf cohomology

In this section only Abelian sheaves will be considered, i.e. sheaves with values in **Ab** (unless stated otherwise).

Property 9.3.1. 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 8, these theories are in fact specific instances.

Definition 9.3.2 (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.

9.3.1 Derived cohomology

Property 9.3.3 (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 9.3.1, one can construct nontrivial (right) derived functors of $\Gamma(X, -)$:

Definition 9.3.4 (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

$$\cdots \longrightarrow \Gamma(X, I^i) \longrightarrow \Gamma(X, I^{i+1}) \longrightarrow \cdots . \tag{9.11}$$

The cohomology group $H^0(X; \mathcal{F})$ is equal to $\Gamma(X, \mathcal{F})$.

Definition 9.3.5 (Acyclic sheaf). A sheaf is said to be acyclic if its higher cohomology groups vanish (cf. Definition 5.4.6).

Example 9.3.6 (Soft sheaves). Soft sheaves 9.2.18 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 9.3.7 (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 9.3.8 (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 $\mathrm{im}(\phi)$. In a similar way one can also define cokernels or any other notion that makes sense in Abelian categories.

Definition 9.3.9 (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 $\ker(d^i_U)/\,\mathrm{im}(d^{i-1}_U)$.

9.3.2 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 9.3.10 (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:

$$C^p(\mathcal{U}; A_\bullet) := \bigoplus_{\substack{p=k-n \\ i_0 < \cdots < i_k}} A_n(U_{i_0 \ldots i_k}). \tag{9.12}$$

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:

$$\begin{aligned}(d\omega)_{i_0 \ldots i_{k+1}} &:= \left(\partial\omega + (-1)^k \sum_{i=0}^{k} (-1)^i A_\bullet(\iota_i)\omega\right)_{i_0 \ldots i_{k+1}} \\ &= \partial\omega_{i_0 \ldots i_{k+1}} + (-1)^k \sum_{j=0}^{k+1} (-1)^j\, \omega_{i_0 \ldots i_{j-1} i_{j+1} \ldots i_{k+1}}\big|_{U_{i_0 \cdots i_{k+1}}},\end{aligned} \tag{9.13}$$

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 9.3.11 (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 (9.13) 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 (9.13) 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 9.3.12. In degrees 0 and 1 one always has

$$\check{H}^{0,1}(X;\mathcal{F}) \cong H^{0,1}(X;\mathcal{F}). \tag{9.14}$$

For a paracompact Hausdorff space, the Čech cohomology and (derived) sheaf cohomology coincide in all degrees.

Property 9.3.13 (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 $\mathcal{U}$:

$$H^{\bullet\geq 1}(U_{i_1}\cap\ldots\cap U_{i_k};\mathcal{F}) = 0 \tag{9.15}$$

for all $i_1,\ldots,i_k \in I$. If $\mathcal{U}$ is acylic with respect to $\mathcal{F}$, then

$$\check{H}^\bullet(\mathcal{U};\mathcal{F}) \cong \check{H}^\bullet(X;\mathcal{F}) \cong H^\bullet(X;\mathcal{F}). \tag{9.16}$$

Example 9.3.14 (Good covers). Consider a topological space admitting a good cover 29.1.16. 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.

9.4 Non-Abelian sheaf cohomology

9.4.1 Čech cohomology

The issue with extending Čech 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 $\{\phi_i : U_i \to G\}_{i\in I}$: $(d\phi)_{ij} := \phi_j\phi_i^{-1}$.
2. For every 1-cochain $\{\psi_{ij} : U_i\cap U_j \to G\}_{i,j\in I}$: $(d\psi)_{ijk} := \psi_{ij}\psi_{ik}^{-1}\psi_{jk}$.
3. ...

But now a major issue ariseq. With this definition, the maps d_k are not group morphisms and the sets $\ker(d_k), \mathrm{im}(d_k)$ 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 $d\phi = e$ 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

$$(\phi\cdot\psi)_{ij} := \phi_j\psi_{ij}\phi_i^{-1}. \tag{9.17}$$

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. The reason why higher cohomology $\check{H}^{\geq 2}(X;G)$ can only be defined for Abelian groups, is not completely obvious. However, when reformulating cohomology in terms of mapping spaces with values in deloopings (see Chapter 13), it becomes clear why this is the case.

9.5 Ringed spaces

Definition 9.5.1 (Ringed space). A topological space equipped with a sheaf of rings.

Definition 9.5.2 (Locally ringed space). A ringed space for which the stalk at every point is a local ring 3.6.25.

Definition 9.5.3 (Quasicoherent sheaf). Let $(X, \mathcal{O}_X)$ be a ringed space. A quasicoherent sheaf $\mathcal{E}$ on X is a sheaf of $\mathcal{O}_X$-modules that is locally the cokernel of a morphism of free modules.

This means that $\mathcal{E}$ is locally presentable in the sense that it fits into an exact sequence of the form

$$\mathcal{O}_X^{k_i}|_{U_i} \longrightarrow \mathcal{O}_X^{l_i}|_{U_i} \longrightarrow \mathcal{E}|_{U_i} \longrightarrow 0 \tag{9.18}$$

for some open cover $\{U_i\}_{i\in I}$ of X.

A quasicoherent sheaf is said to be coherent if it is of **finite type**, i.e. every point has an open neighbourhood U admitting a surjective morphism

$$\mathcal{O}_X^m|_U \longrightarrow \mathcal{E}|_U, \tag{9.19}$$

and, for every open neighbourhood V and integer $n \in \mathbb{N}$, any morphism

$$\mathcal{O}_X^n|_V \longrightarrow \mathcal{E}|_V \tag{9.20}$$

has a kernel of finite type.

Chapter 10

Metric Spaces

10.1 Definition

Definition 10.1.1 (Metric). A metric (or **distance**) on a set M is a map $d : M \times M \to \mathbb{R}^+$ that satisfies the following properties:

1. **Nondegeneracy**: $d(x, y) = 0 \iff x = y$,
2. **Symmetry**: $d(x, y) = d(y, x)$, and
3. **Triangle inequality**: $\forall x, y, z \in M : d(x, z) \leq d(x, y) + d(y, z)$.

A set M equipped with a metric d is called a **metric space**.

Definition 10.1.2 (Diameter). The diameter of a subset $U \subset (M, d)$ of a metric space is defined as follows:

$$\mathrm{diam}(U) := \sup_{x,y \in U} d(x, y). \tag{10.1}$$

Definition 10.1.3 (Bounded). A subset $U \subseteq M$ is bounded if $\mathrm{diam}(U) < \infty$.

10.2 Topology

Multiple topological notions can be (re)formulated in terms of a metric. The most important ones are given below:

Definition 10.2.1 (Open ball). An open ball centered on a point $x_0 \in M$ with radius $R > 0$ is defined as the following set:

$$B(x_0, R) := \{x \in M \mid d(x, x_0) < R\}. \tag{10.2}$$

Property 10.2.2 (Metric topology). Every metric space can be turned into a topological space by taking the open balls to be a basis.

Definition 10.2.3 (Closed ball). The closed ball $\overline{B}(x_0, R)$ is defined as the closure of the open ball $B(x_0, R)$:

$$\overline{B}(x_0, R) := \{x \in M \mid d(x, x_0) \leq R\}. \tag{10.3}$$

Definition 10.2.4 (Packing). Let (M, d) be a metric space. A ε-packing, for $\varepsilon > 0$, is a subset $\mathcal{P} \subseteq M$ such that

$$\sup_{p,q \in \mathcal{P}} d(p, q) \leq \varepsilon. \tag{10.4}$$

The **packing number** $N_P(\varepsilon, M, d)$ is defined as the greatest cardinality of a ε-packing of (M, d).

Definition 10.2.5 (Metrizable space). A topological space X is metrizable if it is homeomorphic to a metric space M or, equivalently, if there exists a metric function $d : X \times X \to \mathbb{R}$ such that it induces the topology on X.

Theorem 10.2.6 (Urysohn's metrization theorem). *Every second-countable T_3-space is metrizable.*

Definition 10.2.7 (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

$$\forall\varepsilon > 0 : \exists N_0 \in \mathbb{N} : \forall n \geq N_0 : d(x_n, a) < \varepsilon. \tag{10.5}$$

Definition 10.2.8 (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 $a \in \mathrm{dom}(f)$ if

$$\forall\varepsilon > 0 : \exists\delta_\varepsilon > 0 : \forall x \in \mathrm{dom}(f) : d(a, x) < \delta_\varepsilon \implies d'\big(f(a), f(x)\big) < \varepsilon. \tag{10.6}$$

Property 10.2.9. Let (M, d) be a metric space. The distance function $d : M \times M \to \mathbb{R}$ is a continuous function.

Definition 10.2.10 (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

$$\forall\varepsilon > 0 : \exists\delta_\varepsilon : \forall x, y \in \mathrm{dom}(f) : d(x, y) < \delta_\varepsilon \implies d'\big(f(x), f(y)\big) < \varepsilon. \tag{10.7}$$

This is clearly a stronger notion than that of continuity since the number ε is equal for all points $y \in \mathrm{dom}(f)$.

Definition 10.2.11 (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

$$d'\big(f(x), f(y)\big) \leq Cd(x, y) \tag{10.8}$$

for all $x, y \in M$.

10.3 Examples

Example 10.3.1 (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

$$d(x, y) := \max_{1\leq i\leq n} d_i(x_i, y_i) \tag{10.9}$$

this product space becomes a metric space.

Property 10.3.2 (Projections determine convergence). Let M be a product metric space. Consider the projections associated with the sets M_j:

$$\mathrm{pr}_j : M \to M_j : (a_1, \ldots, a_n) \mapsto a_j. \tag{10.10}$$

A sequence in a product metric space M converges if and only if every component $(\mathrm{pr}_j(x_m))_{m\in\mathbb{N}}$ converges in (M_j, d_j).

Example 10.3.3 (Supremum distance). Let $K \subset \mathbb{R}^n$ be a compact set. The following map defines a metric on $C(K, \mathbb{C})$:

$$d_\infty(f, g) := \sup_{x \in K} |f(x) - g(x)|. \tag{10.11}$$

Example 10.3.4 (p-metric). For every $p \geq 1$ one defines the L^p-norm on $\mathbb{R}^n$ by the following formula:

$$d_p(x, y) := \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p}. \tag{10.12}$$

Example 10.3.5 (Chebyshev distance). The Chebyshev distance is defined similarly to the supremum distance:

$$d_\infty(x, y) := \max_{1 \leq i \leq n} |x_i - y_i|. \tag{10.13}$$

This metric is also sometimes called the **maximum metric** or L^∞-metric.

Remark 10.3.6. 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 $\{1, \ldots, n\} \to \mathbb{R}$ equipped with the supremum distance. Another justification is the following relation:

$$d_\infty(x, y) = \lim_{p \to \infty} d_p(x, y), \tag{10.14}$$

which is also the origin of the name L^∞-metric.

10.4 Completeness

Definition 10.4.1 (Cauchy sequence). A sequence $(x_n)_{n \in \mathbb{N}}$ in a metric space (M, d) is Cauchy (or has the Cauchy property) if

$$\forall \varepsilon > 0 : \exists N \in \mathbb{N} : \forall m, n \geq N : d(x_m, x_n) < \varepsilon. \tag{10.15}$$

A metric space (M, d) is said to satisfy the Cauchy criterion if a sequence converges to a point $x \in M$ if and only if it is Cauchy.

Definition 10.4.2 (Complete metric space). A metric space that satisfies the Cauchy criterion.

Property 10.4.3. 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.

Definition 10.4.4 (Polish space). A separable, completely metrizable space.

Property 10.4.5. All uncountable Polish spaces are Borel isomorphic (Definition 16.1.28).

Property 10.4.6. All open and closed subsets of a Polish space are Polish. So are G_δ-subsets.

10.4.1 Injective metric spaces

Definition 10.4.7 (Metric retraction). Let (M, d) be a metric space. A function $f : X \to X$ is said to be a retraction of metric spaces if:

- f is idempotent, and
- f is non-expansive, i.e. the following relation holds for all $x, y \in M$:

$$d(f(x), f(y)) \leq d(x, y). \tag{10.16}$$

The image of f is called a (metric) retract of M.

Definition 10.4.8 (Injective metric space). A metric space M is said to be injective if whenever M is isometric to a subspace Y of a metric space X, then Y is a metric retract of X.

Property 10.4.9. Every injective metric space is complete.

10.4.2 Convex metric spaces

Definition 10.4.10 (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

$$d(x, z) = d(x, y) + d(y, z). \tag{10.17}$$

Property 10.4.11 (Convex sets). A closed subset of Euclidean space is a convex metric space if and only if it is a convex set 14.8.1.

Definition 10.4.12 (Hyperconvex space). A convex space for which the set of closed balls has the Helly property 2.4.5.

Theorem 10.4.13 (Aronszajn & Panitchpakdi). *A metric space is injective if and only if it is hyperconvex.*

?? WHY IS THIS HERE ??

10.5 Compactness

See Section 7.5 for the general theory of compact spaces.

Theorem 10.5.1 (Stone). *Every metric space is paracompact.*

Definition 10.5.2 (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 $\forall F \in \mathcal{F} : \mathrm{diam}(F) \leq \varepsilon$.
- For every $\varepsilon > 0$ there exists a finite subset $E \subset M$ such that $M \subseteq \bigcup_{x \in E} B(x, \varepsilon)$.

Definition 10.5.3 (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 he covering number is sometimes called the **metric entropy**.

Property 10.5.4 (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 10.5.5. *For a metric space M the following statements are equivalent:*

- *M is compact.*
- *M is sequentially compact.*
- *M is complete and totally bounded.*

Theorem 10.5.6 (Heine-Cantor). *Let M, M' be two metric spaces with M compact. Every continuous function $f : M \to M'$ is also uniformly continuous.*

Definition 10.5.7 (Equicontinuity). Let X be a topological space and let M be a metric space. A collection $\mathcal{F}$ of maps $X \to M$ is equicontinuous in $x \in X$ if for all $\varepsilon \geq 0$ there exists a neighbourhood U of x such that

$$\forall f \in \mathcal{F}, \forall x' \in U : d(f(x), f(x')) \leq \varepsilon. \tag{10.18}$$

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

$$\forall f \in \mathcal{F} : f(U) \cap V \neq \emptyset \implies f(U) \subset W. \tag{10.19}$$

Property 10.5.8. 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 10.5.9 (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* (10.3.3).
- *$\mathcal{F}$ is equicontinuous, closed under uniform convergence and $\{f(x) \mid f \in \mathcal{F}\}$ is totally bounded for every $x \in K$.*

Definition 10.5.10 (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 which is the same as saying that it is the image of a Polish space under a continuous bijection. (In particular, every Polish space is Lusin.)

Definition 10.5.11 (Suslin space[1]). The image of a Polish space under a continuous function. (In particular, every Lusin space is Suslin.)

Theorem 10.5.12 (Lusin-Suslin). *A subset of a Polish space is Lusin if and only if it is Borel.*

[1] Sometimes written as "Souslin".

Chapter 11

Algebraic Geometry

References for this chapter are [43, 124]. For the basics on ring theory and ideals, see Section 3.6. In order to not confuse the letter k, often used for fields, with various indices and dimensions fields will be denoted by the capital letter K.

11.1 Algebraic numbers

11.1.1 Polynomials

Definition 11.1.1 (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: $R[X] \equiv R[x_1, \ldots, x_n]$.

Definition 11.1.2 (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 11.1.3 (Monic polynomial). A polynomial for which the highest degree term has coefficient 1.

Theorem 11.1.4 (Fundamental theorem of algebra). *Consider a $\mathbb{C}$-valued polynomial of degree ≥ 1. It has at least one root in $\mathbb{C}$.*

Corollary 11.1.5. 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 $a_1, \ldots, a_k \in \mathbb{C}$ and $n_1, \ldots, n_k \in \mathbb{N}$.

Definition 11.1.6 (Transcendental element). Consider a field K and a field extension L/K (Definition 3.8.1). An element $x \in L$ for which there exist no nontrivial polynomials p over K such that $p(x) = 0$, is said to be transcendental, otherwise it is said to be **algebraic**.

Definition 11.1.7 (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 only monic polynomials are considered, a slightly different notion is obtained:

Definition 11.1.8 (Integral dependence). Consider a commutative ring R and a subring S. An element $r \in R$ is said to be integrally dependent on S if it is the root of a monic polynomial in $S[x]$.

Remark 11.1.9. 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.

11.1.2 Roots

Formula 11.1.10 (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:

$$\sum_{1\leq i_1\leq\ldots\leq i_k\leq n}\prod_{j=1}^{k} r_{i_j} = (-1)^k\frac{a_{n-k}}{a_n}, \tag{11.1}$$

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

$$r_1 + r_2 + \cdots + r_n = -\frac{a_{n-1}}{a_n}, \tag{11.2}$$

$$r_1 r_2 \cdots r_n = (-1)^n\frac{a_0}{a_n}. \tag{11.3}$$

Example 11.1.11. For quadratic polynomials ax^2+bx+c one recovers the following well-known formulas:

$$r_1 + r_2 = -\frac{b}{a}, \tag{11.4}$$

$$r_1 r_2 = \frac{c}{a}. \tag{11.5}$$

11.1.3 Ideals

Theorem 11.1.12 (Weak Nullstellensatz). *Let K be an algebraically closed field K and consider the polynomial ring $R = K[x_1,\ldots,x_n]$. An ideal $I \subset R$ is maximal if and only if it is of the form*

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

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

Corollary 11.1.13. There exists a bijection between K^n and the set of maximal ideals of $K[x_1,\ldots,x_n]$.

Corollary 11.1.14. Consider a collection of polynomials $\{f_i\}_{i\in I} \subset K[x_1,\ldots,x_n]$. If these polynomials do not have a common root, the ideal they generate is the unit ideal.

11.2 Varieties

From here on K is assumed to be an algebraically closed field. For notational simplicity and to differentiate between K^n as a vector space and as a set, the notion of affine space is introduced:

Definition 11.2.1 (Affine space). $\mathbb{A}^n$ is defined as the underlying set of the vector space K^n:

$$\mathbb{A}^n := \{(a_1,\ldots,a_n) \in K^n\}. \tag{11.6}$$

Definition 11.2.2 (Algebraic set). Consider a finite set of polynomials in $K[x_1, \ldots, x_n]$. 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 $I \subset K[x_1, \ldots, x_N]$ to be

$$V(I) := \big\{(a_1, \ldots, a_n) \in \mathbb{A}^n \mid \forall f \in I : f(a_1, \ldots, a_k) = 0\big\}. \tag{11.7}$$

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 [124]) make no distinction between general algebraic sets and affine varieties.

Property 11.2.3. By *Hilbert's basis theorem* 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 11.2.4 (Hilbert's Nullstellensatz). *Let J be an ideal in $K[x_1, \ldots, x_n]$ and let $\sqrt{J}$ denote its radical. The following relation holds for all J:*

$$I(V(J)) = \sqrt{J}. \tag{11.8}$$

Similar to the case of the weak Nullstellensatz, one obtains the following result

Corollary 11.2.5. There exists a bijection between the algebraic subsets of $\mathbb{A}^n$ and the radical ideals in $K[x_1, \ldots, x_n]$. The irreducible algebraic sets correspond to the prime ideals (by the *Noetherian decomposition theorem*).

Definition 11.2.6 (Morphism of varieties). Let $V_1 \subset \mathbb{A}^{n_1}, V_2 \subset \mathbb{A}^{n_2}$ be two algebraic sets. A morphism $\varphi : V_1 \to V_2$ is a function that can be expressed in the following way:

$$\varphi(x_1, \ldots, x_{n_1}) = \big(f_1(x_1, \ldots, x_{n_1}), \ldots, f_{n_2}(x_1, \ldots, x_{n_1})\big), \tag{11.9}$$

where $f_i \in K[x_1, \ldots, x_{n_1}]$ for all $i \leq n_2$.

A closely related notion is that of rational maps:

Definition 11.2.7 (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 where $f_U : U \to Y$, 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 if there exists a rational map $g : Y \to X$ such that $f \circ g = \mathbb{1}_Y$ and $g \circ f = \mathbb{1}_X$.

Definition 11.2.8 (Coordinate ring). Consider the polynomial ring $K[x_1, \ldots, x_n]$ and let V be an algebraic set in $\mathbb{A}^n$. The coordinate ring of V is defined as the following quotient:

$$\Gamma(V) := K[x_1, \ldots, x_n]/I(V). \tag{11.10}$$

The elements of this ring are the 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 $K(V)$. This field is called the **function field** of V and the elements of $K(V)$ 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 K-morphism on the associated affine ring by precomposition. This gives rise to the following property:

Property 11.2.9 (Affine varieties and 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 11.2.10 (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:

$$\dim(X) := \sup_{n\in\mathbb{N}}(\exists\ \mathfrak{p}_0 \subsetneq \mathfrak{p}_1 \subsetneq \cdots \subsetneq \mathfrak{p}_n), \tag{11.11}$$

where all $\mathfrak{p}_i$ are prime ideals in $\Gamma(X)$. 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

$$\dim(X) = \max_{p\in X} \dim_p(X). \tag{11.12}$$

11.2.1 Topology

A topology on an affine variety can be constructed in the following way:

Definition 11.2.11 (Zariski topology). A set in $\mathbb{A}^n$ is said to be closed exactly if it is an algebraic set. A basis for this topology is given by the zero loci $B_f = \{x \in \mathbb{A}^n \mid f(x) \neq 0\}$ for $f \in K[x_1, \ldots, x_n]$. 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 in $\Gamma(V)$.

The following property shows that the Zariski topology is very different from the topologies that occur in for example analysis:

Property 11.2.12 (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. From here on the content of Chapter 9 on sheaf theory will become a prerequisite.

Definition 11.2.13 (Structure sheaf). Consider an affine variety X and denote its associated coordinate ring by $\Gamma(X)$. Now, for any point $x \in X$, one can consider the set of functions $m_x \subset R$ that vanish on x. This is a prime ideal, so one can construct the localization of R at m_x:

$$\mathcal{O}_x := R_{m_x} = \{f/g \mid f, g \in R \wedge g(x) \neq 0\}. \tag{11.13}$$

For every open subset $U \subset X$ one can then define the ring of functions on U as follows:

$$\mathcal{O}_X(U) := \bigcap_{x \in U} \mathcal{O}_x. \tag{11.14}$$

$\mathcal{O}_X$ is a sheaf with stalks given by $\mathcal{O}_x$. By Property 3.6.27 all stalks $\mathcal{O}_x$ are local rings and, hence, $(X, \mathcal{O}_X)$ is a locally ringed space 9.5.2. 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 $f, g \in R$ 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 11.2.14. 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: $\mathcal{O}_X(B_f) = R_f$, where R_f denotes the localization 3.6.8 of R at f. In particular

$$\Gamma(X, \mathcal{O}_X) = \Gamma(X), \tag{11.15}$$

where on the left-hand side Γ denotes the global sections functor 9.2.9. This property explains the notation $\Gamma(X)$ introduced before.

Remark 11.2.15. Both the rings $\mathcal{O}_X(U)$ and $\mathcal{O}_x$ are subrings of the function field $K(X)$.

Alternative Definition 11.2.16 (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 11.2.17 (Morphism). A continuous function between affine varieties $f : X \to Y$ such that precomposition by f preserves regular functions.

Property 11.2.18 (Identity theorem). If two regular maps coincide on a nonempty open subset, they are equal.

Definition 11.2.19 (Generic stalk). For the construction of the stalk of the structure sheaf over a point 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 was 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:

$$\mathcal{O}_{\tilde{x}} := \varinjlim_{U \subset X} \mathcal{O}_X(U). \tag{11.16}$$

This stalk is called the generic stalk of X and it is isomorphic to $K(X)$.

11.2.2 Varieties

Definition 11.2.20 (Prevariety). Let X be a topological space equipped with a sheaf $\mathcal{O}_X$ of K-valued functions. The space X is called a prevariety if X is connected and if there exists a finite covering $\{U_i\}_{i \in I}$ of X such that every couple $(U_i, \mathcal{O}_X|_{U_i})$ is an affine variety.

Definition 11.2.21 (Morphism). Consider two prevarieties $(X, \mathcal{O}_X)$ and $(Y, \mathcal{O}_Y)$. A morphism between them is a continuous function $f : X \to Y$ such that

$$g \in \Gamma(V, \mathcal{O}_Y) \implies gf \in \Gamma(f^{-1}V, \mathcal{O}_X) \tag{11.17}$$

for all open sets $V \subset Y$, i.e. a morphism of prevarieties is just a morphism of ringed spaces.

Remark 11.2.22. 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 called the **function field** $K(X)$. It coincides with the function field of every open affine subset of X.

Construction 11.2.23 (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 7.3.9 $U \cup_f Y$ 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 11.2.24 (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 11.2.25. The motivation for this definition is Property 7.1.31. 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 11.2.23 above.

The following two closure properties are very important:

Property 11.2.26. Consider a prevariety morphism $f : X \to Y$ where Y is a variety. The graph of f is closed in $X \times Y$.

Property 11.2.27. 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.

11.2.3 Projective varieties

Definition 11.2.28 (Projective space). Consider the vector space K^n. The projective space $\mathbb{P}_{n-1}(K)$ or $K\mathbb{P}^{n-1}$ is defined as the quotient of K^n under the following equivalence relation:

$$(x_1, \dots, x_n) \sim (y_1, \dots, y_n) \iff \exists \lambda \in K^\times : \forall i \leq n : x_i = \lambda y_i. \tag{11.18}$$

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

$$K_{\text{hom}}[x_0, \dots, x_n] \subset K[x_0, \dots, x_n]$$

consisting of all homogeneous polynomials. This definition implies that

$$\forall f \in K_{\text{hom}}[x_0, \dots, x_n] : f(\lambda x_0, \dots, \lambda x_n) = 0 \iff f(x_0, \dots, x_n) = 0$$

and, hence, that the zero loci of homogeneous polynomials are well-defined subsets of the projective space $\mathbb{P}_n(K)$.

[1]Sometimes also called a **separated prevariety**.

Definition 11.2.29 (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 $K[x_0, \ldots, x_n]$ that is generated by homogeneous polynomials. The projective algebraic set $V_P(I)$ is defined as the zero locus of I:

$$V_P(I) := \{x \in \mathbb{P}_n(K) \mid \forall f \in I : f(x) = 0\}. \tag{11.19}$$

Given a projective algebraic set $V \in \mathbb{P}_n(K)$, one can define the ideal $I_p(V)$ as follows:

$$I_P(V) := (f \in K_{\text{hom}}[x_0, \ldots, x_n] \mid \forall x \in V : f(x) = 0), \tag{11.20}$$

i.e. the ideal $I_P(V)$ is generated by all homogeneous polynomials that vanish on V. The **Zariski topology** on $\mathbb{P}_n(K)$ is defined such that the closed sets are exactly the projective algebraic sets.

Theorem 11.2.30 (Projective Nullstellensatz). *For all homogeneous ideals I, except $I_0 = (x_1, \ldots, x_n)$, one finds that*

$$I_P(V_P(I)) = \sqrt{I}. \tag{11.21}$$

Corollary 11.2.31. As before this implies that there exists a bijection between the projective algebraic sets in $\mathbb{P}_n(K)$ and the homogeneous radical ideals (except for I_0) in $K[x_0, \ldots, x_n]$.

Definition 11.2.32 (Coordinate ring). As for affine algebraic sets, the coordinate ring of a projective algebraic set V is defined as the following quotient:

$$\Gamma(V) := K[x_0, \ldots, x_n]/I_P(V). \tag{11.22}$$

The construction of regular functions on affine varieties 11.2.13 cannot be extended to projective spaces in a straightforward manner. Consider for example a polynomial $f \in K[x_0, \ldots, x_n]$. This polynomial does not form a well-defined function on a projective algebraic set $V_P(I) \subset \mathbb{P}_n(K)$ even if f is homogeneous, since changing the homogeneous coordinates on $V_P(I)$ 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 $V_P(I)$.

Since the ideal I is homogeneous, the quotient $R = K[x_0, \ldots, x_n]/I$ is a graded algebra. Denote by $K(X)$ the zeroth order part of the localization of R by the homogeneous elements:

$$K(X) := \{f/g \mid \exists n \in \mathbb{N} : f, g \in R_n\}. \tag{11.23}$$

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:

$$\mathcal{O}_x := \{f/g \in K(X) \mid g(x) \neq 0\}. \tag{11.24}$$

This ring has a maximal ideal $I_x = \{f/g \in K(X) \mid f(x) = 0, g(x) \neq 0\}$ such that all elements in $\mathcal{O}_x$ are invertible and, by Property 3.6.26; $\mathcal{O}_x$ 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:

$$\mathcal{O}_X(U) = \bigcap_{x \in U} \mathcal{O}_x. \tag{11.25}$$

Property 11.2.33 (Variety). For every projective variety $X \subset \mathbb{P}_n(K)$ 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 11.2.24.

Property 11.2.34 ($\mathbb{A}^n$ in $\mathbb{P}_n(K)$). Consider the affine variety $\mathbb{A}^n$. This set admits a bijective mapping onto an open subset of $\mathbb{P}_n(K)$ as follows:

$$\varphi : \mathbb{A}^n \to U_0 : (x_1, \dots, x_,n) \mapsto [1 : x_1 : \cdots : x_n]. \tag{11.26}$$

It can be shown that this map is a homeomorphism if both spaces are equipped with the Zariski topology.

Property 11.2.35 (Schubert decomposition). The projective space $\mathbb{P}_n(K)$ admits a decomposition of the form

$$\mathbb{P}_n(K) = \bigcup_{i=0}^{n} K^i, \tag{11.27}$$

where the union is set-theoretic. However, one can refine this to a statement in topology. The projective space $\mathbb{P}_n(K)$ admits the structure of a CW complex with one k-cell in every dimension (namely $\mathbb{A}^k$). These cells are called **Bruhat cells** or **Schubert cells**. (The precise distinction is of no relevance here.)

Example 11.2.36 (Finite fields). Consider a finite field $\mathbb{F}_q$. Using the above decomposition one can easily compute the cardinality of $\mathbb{P}_n(\mathbb{F}_q)$:

$$|\mathbb{P}_n(\mathbb{F}_q)| = \sum_{i=0}^{n} |\mathbb{F}_q^i| = \sum_{i=0}^{n} q^i \equiv [n+1]_q, \tag{11.28}$$

For example, the **Fano plane** $\mathbb{F}_2(\mathbb{P}^2)$ has cardinality 7.

Construction 11.2.37 (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 as $X \backslash V(f_1, \dots, f_k)$. By definition these functions do not all vanish simultaneously on Y and, hence, there exists a well-defined map

$$f : Y \to \mathbb{P}_{k-1}(K) : x \mapsto \Big[f_1(x) : \cdots : f_k(x)\Big].$$

The graph of this morphism is closed in $Y \times \mathbb{P}_{k-1}(K)$ by Property 11.2.26, but not in $X \times \mathbb{P}_{k-1}(K)$. 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).

If X is irreducible, there exists a birational morphism $X \to \widetilde{X}$.

Property 11.2.38 (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:

$$\widetilde{X} \subseteq \{(x, y) \in X \times \mathbb{P}_{n-1}(K) \mid \forall i, j \leq n : y_i f_j(x) = y_j f_i(x)\}. \tag{11.29}$$

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 $\mathbb{P}_{n-1}(K)$.

11.3 Schemes ♣

11.3.1 Spectrum of a ring

Definition 11.3.1 (Spectrum). Let R be a commutative ring. The spectrum $\mathrm{Spec}(R)$ 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 $V_I := \{P \subseteq I \mid I \text{ is an ideal}, P \text{ is a prime ideal}\}$. A basis for this topology is given by the sets $D_f := \{P \not\ni f \mid f \in R, P \text{ is a prime ideal}\}$.

Property 11.3.2. $\mathrm{Spec}(R)$ is a compact T_0-space.

Definition 11.3.3 (Structure sheaf). Given a spectrum $X = \mathrm{Spec}(R)$, 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 11.3.4. The spectrum $\mathrm{Spec}(R)$ together with its structure sheaf forms a ringed space.

Definition 11.3.5 (Affine scheme). A locally ringed space that is isomorphic to the spectrum of a commutative ring.

Property 11.3.6. There exists an equivalence of categories $\mathbf{AffSch} \cong \mathbf{CRing}^{op}$.

11.3.2 Zariski tangent space

Definition 11.3.7 (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:

$$C_0X := V(\{f^{\text{in}} \mid f \in I\}), \tag{11.30}$$

where f^{in} denotes the **initial part** of f, i.e. the sum of the smallest degree monomials in f.

Definition 11.3.8 (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:

$$T_xX := V(\{f^{[1]} \mid f \in I(X)\}), \tag{11.31}$$

where $f^{[1]}$ denotes the linear part of the polynomial f.

Property 11.3.9. For $x = 0$ one obtains that $I(0) \equiv (x_1, \ldots, x_n)/I(X)$. Moreover, there exists a natural isomorphism

$$I(0)/I(0)^2 \cong \mathrm{Hom}_K(T_0X, K). \tag{11.32}$$

The tangent space at 0 can accordingly also be ontained 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 11.3.10 (Zariski tangent space). Consider an affine variety 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 $x \in X$.

Definition 11.3.11 (Smooth variety). An affine variety is said to be smooth 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.

[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.

11.4 Algebraic groups

Definition 11.4.1 (Linear algebraic group). A subgroup of $\mathrm{GL}(n, F)$ defined by a (finite) set of polynomials in the matrix coefficients.

Property 11.4.2. From the definition it is immediately clear that intersections of algebraic groups are again algebraic.

?? COMPLETE ??

Chapter 12

Model theory ♣

General references for this chapter are [28, 73]. For more on monoidal model categories see [106]. A good reference for the section on simplicial spaces and, in particular, the theory of Segal spaces is [87]. For more on Reedy model structures see [72]. A gentle introduction to the theory of homotopy (co)limits can be found in [121, 123].

12.1 Simplicial sets

Definition 12.1.1 (Simplex category). The simplex category $\mathbf{\Delta}$ has as objects the posets of the form $[n] := \{0 < \ldots < n\}$ and as morphisms the order-preserving maps.

Definition 12.1.2 (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 12.1.3 (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 **C** is a **C**-valued presheaf on $\mathbf{\Delta}$.

Remark 12.1.4. 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 12.1.5 (Standard simplex). For every n, the standard simplicial n-simplex $\Delta[n]$ is defined as the Yoneda embedding $\mathbf{\Delta}(-, [n])$. One can also define a functor $\Delta_{\text{top}} : \mathbf{\Delta} \rightarrow \mathbf{Top}$ that maps $[n]$ to the standard topological n-simplex Δ^n (Definition 8.2.3).

Property 12.1.6. 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 $\Delta[n] \rightarrow X$.

Property 12.1.7 (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 $\delta_{n,i} : [n-1] \rightarrow [n]$ that misses the i^{th} element.
- For every n and $i \leq n$, the unique map $\sigma_{n,i} : [n+1] \rightarrow [n]$ that duplicates the i^{th} 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 some fundamental relations that are called the **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 $i < j$,
- $d_i \circ s_j = \mathrm{id}$ 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 12.1.8 (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

$$X_1 \underset{d_1}{\overset{d_0}{\rightrightarrows}} X_0 \times X_0.$$

This defines a fucntor $\pi_0 : \mathbf{sSet} \to \mathbf{Set}$. By base change 4.7.3, this also induces a functor on simplicial categories.

Construction 12.1.9 (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 $\boldsymbol{\Delta}$ or the cube category ⧈). Every such functor induces an adjunction

$$\begin{tikzcd} \mathbf{C} \arrow[r, phantom, "\perp"] \arrow[r, bend right, "N"'] & \mathbf{Psh}(\mathbf{S}). \arrow[l, bend right, "|-|"'] \end{tikzcd} \tag{12.1}$$

The **realization functor** $|-|$ is defined as the left Kan extension $\mathrm{Lan}_{\mathcal{Y}}F$. 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, i.e. for $\mathbf{Psh}(\mathbf{S}) \equiv [\mathbf{S}^{op}, \mathcal{V}]$. Furthermore, if one assumes that $\mathbf{C}$ is copowered over $\mathcal{V}$, the realization functor can be expressed as a coend:

$$|X| = \int^{s \in \mathbf{S}} Xs \cdot Fs. \tag{12.2}$$

Example 12.1.10 (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 $\boldsymbol{\Delta} \hookrightarrow \mathbf{Cat}$ induces the nerve functor

$$N : \mathbf{Cat} \to \mathbf{sSet} : \mathbf{C} \mapsto \mathbf{Cat}(-, \mathbf{C}). \tag{12.3}$$

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 faithfull.

Example 12.1.11 (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:

$$|\cdot| := \mathrm{Lan}_{\mathcal{Y}}\Delta_{\mathrm{top}}. \tag{12.4}$$

An application of the Yoneda lemma shows that the geometric realization can be expressed as a functor tensor product 4.7.8:

$$|X| = X \otimes_{\mathbf{\Delta}} \Delta_{\mathrm{top}} = \int^{n\in\mathbf{\Delta}} X_n \cdot \Delta^n. \tag{12.5}$$

This formula can easily be generalized to the category of simplical 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

$$|X| := \bigsqcup_{n\in\mathbb{N}} X_n \times \Delta^n / \sim, \tag{12.6}$$

where the equivalence relation identifies the points (x, f_*y) and (f^*x, y) for all morphisms $f \in \mathrm{hom}(\mathbf{\Delta})$. The morphisms f^*, f_* are the ones induced by X and $\Delta_{\mathrm{top}} : \mathbf{\Delta} \hookrightarrow \mathbf{Top}$. As an immediate example one obtains

$$|\Delta[n]| = \Delta^n, \tag{12.7}$$

which shows that $\Delta[n]$ really deserves to be called the standard n-simplex.

Example 12.1.12 (Singular set). Given a topological space X one can define a simplicial set $\mathrm{Sing}(X)$. Its components are defined as the set of morphisms from the standard (topological) n-simplex to X:

$$\mathrm{Sing}(X)_n := \mathbf{Top}(\Delta^n, X). \tag{12.8}$$

This is the object of relevance in the definition of singular (co)homology as given in Section 8.3.

Property 12.1.13 (Classifying space). For a (discrete) group G one can construct two important objects: the delooping $\mathbf{B}G$ and the classifying space BG (Definitions 4.10.2 and 33.2.1). As their notations imply there exists a relation between these space. By first taking the nerve of $\mathbf{B}G$ and then passing to its geometric realization, one obtains BG. In fact this method can be applied to any monoid A to obtain the so-called (two-sided) *bar construction*.

12.1.1 Homological algebra

In this section simplicial sets are related to homological algebra (Chapter 5). A basic introduction is [81].

Construction 12.1.14 (Alternating face map complex). From a simplicial Abelian group A, one can construct a connective chain complex as follows. For every $n \in \mathbb{N}$:

$$(CA)_n := A_n. \tag{12.9}$$

The boundary maps δ_n are defined as the alternating sum of the face maps:

$$\delta_n := \sum_{i=1}^{n}(-1)^i d_i. \tag{12.10}$$

Every group A_{n+1} contains a subgroup $D(A_n)$ generated by the degeneracy maps:

$$D(A_n) := \left\langle \bigcup_{i=1}^{n} s_i(A_n) \right\rangle. \tag{12.11}$$

If these degenerate simplices are quotiented out, the **normalized complex** is obtained.

This construction can be generalized to any simplicial group:

Construction 12.1.15 (Moore complex). Let G be a simplicial group. For every $n \in \mathbb{N}$:

$$(NG)_n := \bigcap_{i=1}^{n} \ker(d_i^n). \tag{12.12}$$

The differential ∂_n is given by the zeroth face map d_0^n.

Property 12.1.16 (Equivalences). 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 12.1.17 (Dold-Kan correspondence). *The functor that maps simplicial Abelian groups to normalized chain complexes gives an equivalence of categories* $\mathbf{sAb} \to \mathbf{Ch}^+(\mathbf{Ab})$.

12.2 Localization

Definition 12.2.1 (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. Any two composable morphisms $f, g \in \hom(\mathbf{W})$ satisfy the "2-out-of-3 property": If any two of $\{f, g, f \circ g\}$ are in $\hom(\mathbf{W})$, so is the third.

Definition 12.2.2 (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 3 properties:

1. Every morphism in $\mathbf{C}$ factorizes as a composition $g \circ f$ where $f \in L$ and $g \in R$.
2. L consists of exactly those morphisms in $\mathbf{C}$ that have the left lifting property 4.4.23 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.[1] This was later fixed by introducing the 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 including closure as above.

[1]Model categories defined using the "strong" notion of weak factorization system were then called **closed model categories**.

Definition 12.2.3 (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. Any three composable morphisms $f, g, h \in \hom(\mathbf{W})$ satisfy the "2-out-of-6 property": If $f \circ g$ and $g \circ f$ are in $\hom(\mathbf{W})$, then so are f, g, h and $f \circ g \circ h$.

It is not hard to see that every homotopical category is a category with weak equivalences.

Definition 12.2.4 (Homotopical functor). Consider two homotopical categories $\mathbf{C}, \mathbf{D}$. A functor $F : \mathbf{C} \to \mathbf{D}$ is said to be homotopical if it preserves weak equivalences.

Definition 12.2.5 (Gabriel-Zisman localization). Consider a category $\mathbf{C}$ with a collection of morphisms $M \subset \hom(\mathbf{C})$. The localization of $\mathbf{C}$ with respect to M is constructed by adding for each morphism $f \in M$ a formal inverse to $\hom(\mathbf{C})$.

More specifically, the localization consists of a category $\mathbf{C}[M^{-1}]$ and a functor $F_M : \mathbf{C} \to \mathbf{C}[M^{-1}]$ that inverts M, i.e. it maps all morphisms in M to isomorphisms, with the property that F_M is universal with respect to inverting M.

Universality here means that for every other category $\mathbf{D}$ and functor $F : \mathbf{C} \to \mathbf{D}$ that inverts all morphisms in M the following conditions are satisfied:

- There exists a functor $Z_F : \mathbf{C}[M^{-1}] \to \mathbf{D}$ such that $Z_F \circ F_M$ is naturally isomorphic to F.
- The precomposition functor $F_M^* : [\mathbf{C}[M^{-1}], \mathbf{D}] \to [\mathbf{C}, \mathbf{D}]$ is fully faithfull.

Definition 12.2.6 (Homotopy category II). 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 functor $\mathbf{C} \to \mathbf{Ho}(\mathbf{C})$ is sometimes denoted by $\gamma_{\mathbf{C}}$.

Remark 12.2.7 (Size issues). When $\mathbf{C}$ is small, so is its localization. However, even in the case were $\mathbf{C}$ is locally small, its localization might be large.

Definition 12.2.8 (Reflective localization). Consider an adjunction $F \dashv G : \mathbf{C} \to \mathbf{D}$. G is fully faithful, i.e. defines a reflective subcategory 4.2.27, 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 12.2.9 (Derived functor). Consider a homotopical functor $F : \mathbf{A} \to \mathbf{B}$ and let γ be the composition $\gamma = \gamma_{\mathbf{B}} \circ F$ with the localization functor $\gamma_{\mathbf{B}} : \mathbf{B} \to \mathbf{Ho}(\mathbf{B})$. The derived functor $\mathbf{Ho}(F) : \mathbf{Ho}(\mathbf{A}) \to \mathbf{Ho}(\mathbf{B})$ is the functor obtained by the universal property of $\mathbf{Ho}(\mathbf{A})$ applied to γ.

This definition can be rephrased in terms of Kan extensions. Consider a homotopical functor $F : \mathbf{A} \to \mathbf{B}$. The left and right derived functors are defined as the following Kan extensions:

$$LF := \mathrm{Ran}_{\gamma_{\mathbf{A}}}(\gamma_{\mathbf{B}} \circ F) \tag{12.13}$$

$$RF := \mathrm{Lan}_{\gamma_{\mathbf{A}}}(\gamma_{\mathbf{B}} \circ F). \tag{12.14}$$

In fact, one can drop the assumption that $\mathbf{B}$ has weak equivalences (here F should map weak equivalences to isomorphisms). In this case one simply has to replace $\gamma_{\mathbf{B}} \circ F$ by F in the above formulas.

Example 12.2.10 (Derived category). Consider an Abelian category $\mathbf{A}$ together with its category of chain complexes $\mathbf{Ch}(\mathbf{A})$. The derived category $\mathcal{D}(\mathbf{A})$ is defined as the localization of $\mathbf{Ch}(\mathbf{A})$ at the collection of quasi-isomorphisms.

Remark 12.2.11. In this case it can be shown that one can first restrict to the naive homotopy category $\mathbf{K}(\mathbf{A})$, consisting of chain complexes and chain maps up to chain-homotopy, and then localize at the collection of quasi-isomorphisms.

12.3 Model categories

Definition 12.3.1 (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 12.2.1.
2. $(\mathrm{Cof}, \mathrm{Fib} \cap W)$ and $(\mathrm{Cof} \cap W, \mathrm{Fib})$ are weak factorization systems 12.2.2.

The morphisms in $\mathrm{Fib} \cap W$ and $\mathrm{Cof} \cap W$ are said to be **acyclic** or **trivial**.

Remark. That W contains all isomorphisms in fact follows from the property that any class of morphisms satisfying a lifting property contains all isomorphisms.

Definition 12.3.2 (Model category). A complete and cocomplete category equipped with a model structure.[2]

Definition 12.3.3 (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 12.3.4 (Fibrant object). An object in a model category is said to be fibrant if the morphism to the terminal object is a fibration. Dually, an object in a model category is said to be cofibrant if the morphism to the initial object is a cofibration.

Property 12.3.5 (Model structure on functor categories). Consider a (small) category $\mathbf{A}$ and a model category $\mathbf{B}$. In certain cases the functor category $[\mathbf{A}, \mathbf{B}]$ 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.

Property 12.3.6 (Resolution). In a model category the (co)completeness 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^{c} by suitably factorizing the morphisms to the initial and terminal objects. These replacements are also sometimes called **resolutions** or **approximations**.

If the weak factorization system is in fact functorial 4.3.5, (co)fibrant replacement defines an endofunctor that is weakly equivalent to the identity functor.

Definition 12.3.7 (Quillen adjunction). Let $\mathbf{A}, \mathbf{B}$ be two model categories. An adjunction

$$\begin{tikzcd} \mathbf{B} \arrow[r, shift right, "G"'] \arrow[r, phantom, "\perp"] & \mathbf{A} \arrow[l, shift right, "F"'] \end{tikzcd}$$

[2] *Quillen*'s original definition only required the existence of finite limits and finite colimits.

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, 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, G) is called a **Quillen equivalence**.

Quillen equivalences can also be characterized on the level of homotopy categories:

Property 12.3.8. Let $F \dashv G$ be a Quillen adjunction. This pair is a Quillen equivalence if and only if the left (resp. right) derived functor LF (resp. RG) is an equivalence.

Property 12.3.9 (Derived adjunction). If $(F \dashv G)$ is a Quillen adjunction, the derived functors (LF, RG) also form an adjunction.

Property 12.3.10 (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 12.3.11 (Topological spaces). A first example of model structures is the category **Top** of topological spaces (Chapters 7 and 8). This category can be endowed with a model structure by taking the weak equivalences to be the weak homotopy equivalences 8.1.31 and by taking the fibrations to be the Serre fibrations 8.1.60.

Example 12.3.12 (Simplicial sets). As a second example consider the category **sSet** of simplicial sets 12.1.2. 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 realizations and by taking the fibrations to be Kan fibrations. With this structure the fibrant objects are the Kan complexes and the cofibrations are the levelwise injections, i.e. the cofibrations are the monomorphisms.

Notation 12.3.13 (Quillen's model structure). The model structure defined in the above example is generally called Quillen's model structure on simplicial sets and it is denoted by $\mathbf{sSet}_{Quillen}$.

Property 12.3.14 (Quillen). The adjoint pair of geometric realization and singular set functors (12.1.11 and 12.1.12) 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.

12.3.1 Monoidal structures

Definition 12.3.15 (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 $a \otimes y \sqcup_{a\otimes x} b \otimes x$ 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

$$\begin{tikzcd} a \otimes x \arrow[r] \arrow[d] & b \otimes x \arrow[d] \\ a \otimes y \arrow[r] & b \otimes y. \end{tikzcd}$$

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**.

Definition 12.3.16 (Quillen bifunctor). Any 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 4.7.14. Consider such a triple of bifunctors $(\otimes, \mathrm{hom}_L, \mathrm{hom}_R)$.

- The bifunctor $-\otimes- : \mathbf{C} \times \mathbf{D} \to \mathbf{E}$ is said to be **left Quillen** if its pushout-product of cofibrations is again a cofibration and if the acyclicity of any of the domain morphisms implies the acyclicity of the result.
- The bifunctors $\mathrm{hom}_L : \mathbf{C}^{op} \times \mathbf{E} \to \mathbf{D}$ and $\mathrm{hom}_R : \mathbf{D}^{op} \times \mathbf{E} \to \mathbf{C}$ 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 domain morphisms implies the acyclicity of the result.

It can be shown that if one of these bifunctors is Quillen, the other two are also Quillen.

Remark 12.3.17. The fact that in the two-variable adjunction approach one does not mention preservation of (co)limits follows from the property that left (resp. right) adjoints preserve colimits (resp. limits).

Definition 12.3.18 (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 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 12.3.19 (Enriched model category). Let $\mathcal{V}$ be a monoidal model category. A category $\mathbf{M}$ is called a $\mathcal{V}$-enriched model category if is satisfies the following conditions:

1. $\mathbf{M}$ is a $\mathcal{V}$-enriched category that is both powered and copowered over $\mathcal{V}$.
2. The underlying category $\mathbf{M}_0$ is a model category.
3. The copower is a left Quillen bifunctor or, equivalently, the power is a right Quillen bifunctor.

Example 12.3.20 (Simplicial model category). A model category enriched over the standard model category of simplicial sets $\mathbf{sSet}_{Quillen}$. It can be shown that the full subcategory of a simplicial model category on fibrant-cofibrant objects is enriched over Kan-complexes.

12.3.2 Homotopy category

Definition 12.3.21 (Homotopy). Before being able to construct homotopies in general model categories, thereby generalizing the ideas from Section 8.1, one has to define the counterpart of the unit interval in a model category $\mathbf{C}$. To this end consider an object $x \in \mathrm{ob}(\mathbf{C})$. By

$$\begin{tikzcd}
x \arrow[r] \arrow[dr, "f"'] & \mathrm{Cyl}(x) \arrow[d, "\eta"] & x \arrow[l] \arrow[dl, "g"] \\
& y &
\end{tikzcd}$$

(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 12.1: Homotopies in a model category.

taking the product and coproduct of two copies of x, one can construct the unique diagonal and codiagonal morphisms $\Delta : x \to x \times x$ and $\nabla : x \sqcup x \to x$. By factorizing these morphisms in $\mathbf{C}$, two weak equivalences are obtained:

$$x \to \mathrm{Path}(x), \tag{12.15}$$

$$\mathrm{Cyl}(x) \to x. \tag{12.16}$$

These objects are called the **path object** and **cylinder object** respectively. 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 $\eta : \mathrm{Cyl}(x) \to y$ such that Diagram 12.1a commutes. Analogously, a right homotopy between f and g is a morphism $\lambda : x \to \mathrm{Path}(y)$ such that Diagram 12.1b commutes. The existence of homotopies induces an equivalence relation on morphisms.

Example 12.3.22 (Topological spaces). Consider the category **Top**. A cylinder object $\mathrm{Cyl}(X)$ 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 8.1.2.

A path object $\mathrm{Path}(X)$ 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 8.1.9.

Property 12.3.23. If x is cofibrant, every left homotopy induces a right homotopy. Dually, if y is fibrant, every right homotopy induces a left homotopy.

Corollary 12.3.24. Whenever x is cofibrant and y is fibrant, the relations of being left homotopic (or equivalently right homotopic) coincide on $\mathbf{C}(x, y)$ and, in particular, define equivalence relations. The equivalence classes are denoted by $[x, y]$.

Property 12.3.25 (Stability under composition). Homotopies are preserved under both precomposition and postcomposition by arbitrary morphisms.

Property 12.3.26 (Weak equivalences). A morphism is a weak equivalence if and only if it is a homotopy inverse.

Definition 12.3.27 (Homotopy equivalence). Two objects 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 12.2.5 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 12.3.28 (Homotopy category I). Let $\mathbf{C}$ be a model category. The homotopy category $\mathbf{Ho}(\mathbf{C})$ is the category defined by the following data:

- **Objects**: $\mathrm{ob}(\mathbf{C})$
- **Morphisms**: $[x^{\mathrm{cf}}, y^{\mathrm{cf}}]$

In fact it is easier to restrict to the subcategory $\mathbf{C_{cf}}$ of $\mathbf{Ho}(\mathbf{C})$ 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 12.3.29. The homotopy category of a model category is equivalent to those of the full subcategories on (co)fibrant objects:

$$\mathbf{Ho}(\mathbf{C}) \cong \mathbf{Ho}(\mathbf{C_f}) \cong \mathbf{Ho}(\mathbf{C_c}) \cong \mathbf{Ho}(\mathbf{C_{cf}}). \tag{12.17}$$

Theorem 12.3.30 (Whitehead). *A weak equivalence between objects that are both fibrant and cofibrant is a homotopy equivalence.*

Property 12.3.31. A Quillen equivalence between model categories induces an equivalence of homotopy categories.

Property 12.3.32 (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 the underlying category of a category 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. To formalize this notion, the following definition is introduced:

Definition 12.3.33 (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. $\mathbf{C}$ admits finite products.
2. Fibrations and acyclic fibrations are preserved under arbitrary pullbacks.
3. Every object admits a **good path object**, i.e. a factorization of the product map $x \to x \times x$ as the composition of a weak equivalence and a fibration.
4. All objects are **fibrant**, i.e. the terminal map $x \to 1$ is a fibration for all objects $x \in \mathrm{ob}(\mathbf{C})$.

Theorem 12.3.34 (Factorization lemma). *Let $\mathbf{C}$ be a category of fibrant objects. Any morphism $f : x \to y$ 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 12.3.35 (Ken Brown's lemma). *Let* $\mathbf{A}$ *be a category of fibrant objects and let* $\mathbf{B}$ *be a category with weak equivalences. If a functor* $F : \mathbf{A} \to \mathbf{B}$ *maps acyclic fibrations to weak equivalences, it preserves all weak equivalences.*

Remark 12.3.36. 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 12.2.9):

Alternative Definition 12.3.37 (Derived functor). Let $F : \mathbf{A} \to \mathbf{B}$ 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 12.3.38 (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.

12.3.3 Reedy model structure

Consider a complete and cocomplete category $\mathbf{M}$ (in the remainder of this section this will often be $\mathbf{sSet}$). For any full subcategory inclusion $\mathbf{B} \hookrightarrow \mathbf{A}$ one obtains an induced **truncation** (or **restriction**) functor $\mathrm{tr} : \mathbf{M}^{\mathbf{A}} \to \mathbf{M}^{\mathbf{B}}$. The left and right adjoints of this functor are respectively called the **skeleton** and **coskeleton** functors.

Formula 12.3.39. The adjoint functors are defined by Kan extensions and hence we can express them in terms of (co)ends and weighted (co)limits:

$$\mathrm{sk}(X)x := \int^{y\in\mathbf{B}} \mathbf{A}(y,x) \cdot Xy = \mathrm{colim}^{\mathbf{A}(-,x)} \tag{12.18}$$

$$\mathrm{cosk}(X)x := \int_{y\in\mathbf{B}} [\mathbf{A}(x,y), Xy] = \lim{}^{\mathbf{A}(x,-)} X. \tag{12.19}$$

Definition 12.3.40 (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 12.3.41 (Simplicial nerve). The nerve functor $N : \mathbf{Cat} \to \mathbf{sSet}$ from Sefinition 12.1.10 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 $\mathbf{Set}$ (under the isomorphism $C_2 \cong C_1 \times_{C_0} C_1$ which will be called the first *Segal condition* in Definition 12.4.19).

Definition 12.3.42 (Reedy category). A category $\mathbf{C}$ equipped with a **degree** function $\mathrm{ob}(\mathbf{C}) \to \alpha$, where α is an ordinal 2.6.19, and two wide subcategories $\mathbf{C}^{\pm}$ that satisfy the following conditions:

1. Nontrivial morphisms in $\mathbf{C}^{+}$ (strictly) increase the degree.
2. Nontrivial morphisms in $\mathbf{C}^{-}$ (strictly) decrease the degree.

3. All morphisms admit a unique factorization as a morphism in $\mathbf{C}^-$ followed by a morphism in $\mathbf{C}^+$. This factorization is sometimes called the **(canonical) Reedy factorization**.

Property 12.3.43 (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 12.3.44 (Isomorphisms are trivial). A morphism in a Reedy category is an isomorphism if and only if it is trivial.

Example 12.3.45. 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 12.3.46 (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 comonads (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 are defined as the restrictions of $\mathbf{sk}_{n-1}$ and $\mathbf{cosk}_{n-1}$ to the degree n subcategory of $\mathbf{R}$. The counit of the skeleton adjunction and the unit of the coskeleton adjunction 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_rX can be shown to be isomorphic to the colimit of X over $\mathbf{R}^+(r)$.

Example 12.3.47 (Simplicial objects). The above property allows to give a nice interpretation to latching objects in the case of $\mathbf{R} = \mathbf{\Delta}^{op}$. Using the *Eilenberg-Zilber lemma* it can be shown that the n^{th} latching object of a simplicial object is given by its collection of degenerate n-simplices.

Definition 12.3.48 (Boundary). The boundary of a representable presheaf $\mathbf{R}(-, r)$ is defined as the latching object of the Yoneda embedding $\mathcal{Y} : \mathbf{R} \rightarrow \mathbf{Psh}(\mathbf{R})$ at r. It is denoted by $\partial\mathbf{R}(-, r)$. It can be shown that $\partial\mathbf{R}(-, r)$ consists of exactly those morphisms that are not in $\mathbf{R}^-$ or, equivalently, as $\mathbf{sk}_{n-1}\mathbf{R}(-, r)$. The latching map coincides with the canonical inclusion $\partial\mathbf{R}(-, r) \hookrightarrow \mathbf{R}(-, r)$ if r is of degree n.

Formula 12.3.49. One can show that the latching and matching objects can be obtained through (co)limits weighted by boundaries:

$$M_rX \cong \lim^{\partial\mathbf{R}(r,-)} X \tag{12.20}$$

$$L_rX \cong \mathrm{colim}^{\partial\mathbf{R}(-,r)} X. \tag{12.21}$$

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 12.3.50 (Relative matching and latching objects). Consider the (weighted) colimit bifunctor colim : $\mathbf{Psh}(\mathbf{R}) \times [\mathbf{R}, \mathbf{M}] \rightarrow \mathbf{M}$. The Leibniz construction 12.3.15 allows to define the relative latching object of $f : X \rightarrow Y$ at $r \in \mathrm{ob}(\mathbf{R})$ as the Leibniz product of the boundary inclusion $\partial\mathbf{R}(-, r) \hookrightarrow \mathbf{R}(-, r)$ and f.

By Equations (4.52) and (12.21), the relative latching map is of the form $Xr \sqcup_{L_rX} L_rY \rightarrow Yr$ and the relative matching map is of the form $Xr \rightarrow Yr \times_{M_rY} M_rX$.

Property 12.3.51 (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 $r \in \mathrm{ob}(\mathbf{R})$
- **Cofibrations**: morphisms for which the relative latching map is a cofibration (in $\mathbf{M}$) for all $r \in \mathrm{ob}(\mathbf{R})$

Property 12.3.52 (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 way too abstract and difficult at this point) ??

12.4 Simplicial spaces

12.4.1 Kan complexes

Definition 12.4.1 (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 k^{th} 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 12.4.2 (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 12.4.3 (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 12.4.4 (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. 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 (due to *Boardman* and *Vogt*):

Definition 12.4.5 (Quasicategory[3]). A simplicial set for which all inner horns have (not necessarily unique) fillers. This condition is sometimes called the **Boardman condition**.

At this point the first instance of a ‘‘homotopy category’’ can be defined:

Definition 12.4.6 (Homotopy category II). 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 edges f, g and h.

Property 12.4.7 (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 12.4.4:

Theorem 12.4.8 (Joyal). *A quasicategory is a Kan complex if and only if its homotopy category is a groupoid.*

[3]Some authors such as *Joyal* call these **logoi** (singular: **logos**).

12.4.2 Simplicial localization

Definition 12.4.9 (Homotopy category III). Consider a simplicially enriched category $\mathbf{C}$. Its homotopy category $\pi_0(\mathbf{C})$ is defined as follows:

- **Objects**: $\mathrm{ob}(\mathbf{C})$
- **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 12.4.10 (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. [4]
- 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 12.4.11 (Hammock localization). Let $(\mathbf{C}, W)$ be a category with weak equivalences. Its hammock localization (or **simplicial localization**) $L^H\mathbf{C}$ is the simplicially enriched category constructed as follows:

- **Objects**: $\mathrm{ob}(\mathbf{C})$
- **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 composable morphisms.

Property 12.4.12. Let $(\mathbf{C}, W)$ be a category with weak equivalences. It can be shown that

$$\mathrm{Ho}(L^H\mathbf{C}) \cong \mathbf{C}[W^{-1}]. \tag{12.22}$$

This construction gives a more explicit description of the homotopy category. Furthermore, if $\mathbf{M}$ is a simplicial model category, the categories $\mathbf{M_{cf}}$ and $L^H\mathbf{M}$ are Dwyer-Kan equivalent.

Property 12.4.13. Quillen equivalent model categories have Dwyer-Kan equivalent simplicial localizations.

Property 12.4.14. Up to Dwyer-Kan equivalence, every simplicially enriched category can be obtained as the simplicial localization of a category with weak equivalences.

Definition 12.4.15 (Simplicial resolution). Consider an object x 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.

[4] This is a generalization of the ordinary definition for categories. (Simplicial categories will give models for ∞-categories.)

Definition 12.4.16 (Bergner model structure). The catgory of simplicially enriched categories admits the following model structure:

- **Weak equivalences**: Dwyer-Kan equivalences
- **Cofibrant objects**: *simplicial computads*
- **Fibrant objects**: **Kan**-enriched categories

Construction 12.4.17 (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**: $\mathrm{ob}(\mathbf{C})$
- **Morphisms**: $\hom(\mathfrak{C}\mathbf{C})_n := \hom(F^{n+1}\mathbf{C})$, 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.

12.4.3 Segal spaces

Example 12.4.18 (Bisimplicial sets). One can also look at simplicial objects in **sSet** (these are in particular **sSet**-enriched). Such objects are often called bisimplicial sets or **simplicial spaces**[5]. Since **sSet** is a model category, Property 12.3.5 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 12.4.19 (Segal space). Consider a fibrant object X in the injective (or Reedy) 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 $n \geq 1$:[6]

$$X_n \simeq X_1 \times_{X_0} \cdots \times_{X_0} X_1 \quad (n \text{ factors}). \tag{12.23}$$

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 12.4.20 (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 12.4.21 (Mapping space). Consider a Segal space X. For every two points $x, y \in X_0$, the mapping space $\mathrm{Map}(x, y)$ 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 element for $x \in X_0$ is given by $s_0(x)$.

The following two definitions should be compared to definitions 12.4.9 and 12.4.10:

Definition 12.4.22 (Homotopy category IV). Consider a Segal category X. Its homotopy category $\mathbf{Ho}(X)$ is defined as follows:

- **Objects**: X_0
- **Morphisms**: $\pi_0\mathrm{Map}(x, y)$

[5]The latter name follows from the fact that topological spaces and simplicial sets are (Quillen-)equivalent.

[6]If the Reedy condition is omitted, the limit on the right-hand side has to be replaced by a *homotopy limit*.

Two points $f, g \in \mathrm{Map}(x, y)$ are said to be **homotopic** if they are identified in $\mathbf{Ho}(X)$. A point $f \in \mathrm{Map}(x, y)$ 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 $X_{\mathrm{hoequiv}} \subset X_1$ consists of the components that contain homotopy equivalences.[7]

Definition 12.4.23 (Dwyer-Kan equivalence II). A map F of Segal spaces such that:

1. the induced map on mapping spaces is a weak equivalence.
2. the induced map between homotopy categories is an equivalence (of categories).

Definition 12.4.24 (Complete Segal space). A Segal space X for which the map $s_0 : X_0 \to X_{\mathrm{hoequiv}}$ is a weak equivalence.

Property 12.4.25 (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.

12.4.4 Coherent nerve

The nerve and realization functors 12.1.9 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**: $[n] \equiv \{0, 1, \dots, n\}$
- **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**[8]) 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 12.4.26. The cosimplicial object $C : \mathbf{\Delta} \to \mathbf{sSetCat}$ is in fact just the free resolution functor $\mathfrak{C}$ in the Bergner model structure 12.4.17 restricted to the subcategory $\mathbf{\Delta}$.

Property 12.4.27. The homotopy coherent nerve functor N_Δ is uniquely determined by the following relation:

$$\mathbf{sSet}(\Delta[n], N_\Delta \mathbf{C}) = \mathbf{sSetCat}(C[n], \mathbf{C}). \tag{12.24}$$

Property 12.4.28 (Simplicial realization). The left adjoint of the homotopy coherent nerve functor satisfies the following relation for all finite ordinals $[n]$:

$$|\Delta[n]| \cong C[n]. \tag{12.25}$$

Property 12.4.29 (Quasicategories). If $\mathbf{C}$ is **Kan**-enriched, its coherent nerve is a quasicategory. By Example 12.3.20 this allows to associate a quasicategory to any simplicial model category by passing to the coherent nerve of its fibrant-cofibrant subcategory.

[7]It should be noted that if any point in a component is a homotopy equivalence, all points in that component are homotopy equivalences.

[8]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.

At this point it is interesting to reconsider the simplicial nerve construction 12.1.10. Although the functor $N : \mathbf{Cat} \to \mathbf{sSet}$ is fully faithfull, there is a problem 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 12.3.12 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 groupoid, does one recover the converse statement.

A first solution is to use a different model structure with less weak equivalences:

Definition 12.4.30 (Joyal model structure). The category **sSet** of simplicial sets 12.1.2 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 12.4.5

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 are 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 12.4.31 (Classifying diagram). Consider a (small) category $\mathbf{C}$ together with the functor category $\mathbf{C}^{[n]}$ where $[n]$ is the totally ordered set on $n+1$ elements. The classifying diagram of $\mathbf{C}$ is the bisimplicial set $\widetilde{N}\mathbf{C}$ defined levelwise as follows:

$$\widetilde{N}\mathbf{C}_k := N(\mathrm{Core}(\mathbf{C})), \tag{12.26}$$

where N is the ordinary nerve functor and Core denotes the core functor 4.10.4.

The reason for why this construction is better for distinguishing categories and groupoids comes 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 12.4.32. If $\mathbf{C}$ is small, then $\widetilde{N}\mathbf{C}$ is a complete Segal space.

?? COMPLETE ??

12.5 Cofibrant generation

12.5.1 Transfinite constructions

Before proceeding, some notions from ordinary category need to be specialized to the context of regular cardinals 2.6.27. In this section the symbol κ will always denote such a regular cardinal.

Definition 12.5.1 (κ-filtered category). A category in which every diagram with less than κ arrows admits a cocone.

Definition 12.5.2 (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 12.5.3 (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 4.4.45 also holds in the κ-context, i.e. a category has all κ-directed colimits if and only if it has all κ-filtered colimits (and analogously for limits).

Definition 12.5.4 (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**.

Definition 12.5.5 (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 $\mathbf{C}$ contains a small subcategory $\mathbf{D}$ of κ-small objects that generates all objects by κ-filtered colimits, i.e. $\mathbf{C} \cong \mathrm{Ind}_\kappa(\mathbf{D})$.

Definition 12.5.6 (Locally presentable category). A cocomplete, accessible category. It can be shown that such categories are also complete and, hence, bicomplete.

Property 12.5.7. Every locally presentable category can be obtained as a full reflective subcategory of a presheaf category under an **accessible embedding**, i.e. an embedding that preserves filtered colimits.

Property 12.5.8 (Gabriel-Ulmer duality). There exists an equivalence between locally finitely presentable categories and categories of left exact copresheaves from small complete categories. This gives a 2-equivalence $\mathbf{Lex}^{op} \cong \mathbf{LFP}$.

Theorem 12.5.9 (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 12.5.10 (Transfinite composition). Consider a category $\mathbf{C}$ with a collection of morphisms $I \subseteq \hom(\mathbf{C})$ and let α be an infinite ordinal 2.6.19. A (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 $D_0 \to D\alpha \equiv \operatorname{colim} D$.

Definition 12.5.11 (Cell complex). Consider a cocomplete category $\mathbf{C}$ with a collection of morphisms $I \subseteq \hom(\mathbf{C})$. A **relative** I-cell complex is a transfinite composition of pushouts (of coproducts[9]) 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 12.5.12 (Relative cell complexes). The set of all relative I-cell complexes is often denoted by $\mathrm{cell}(I)$.

[9]It can be shown that closure under coproducts follows automatically.

12.5.2 Cofibrant generation

The following is a famous result by *Quillen*:

Theorem 12.5.13 (Small object argument). *Let* $\mathbf{C}$ *be a locally presentable category with a collection of morphisms* $I \subseteq \mathrm{hom}(\mathbf{C})$*. Every morphism in* $\mathbf{C}$ *can be factorized as the composition of a morphism in* $\mathrm{rlp}(I)$ *followed by a morphism in* $\mathrm{cell}(I)$*.*

Remark 12.5.14. This theorem can be generalized to cocomplete categories where the morphisms in I are small relative[10] to $\mathrm{cell}(I)$. Sets of morphisms with this property are said to **admit a small object argument**.

Definition 12.5.15 (Cofibrantly generated model category). Consider a model category $\mathbf{C}$. This category is said to be cofibrantly generated by two sets of morphisms $I, J \subseteq \mathrm{hom}(\mathbf{C})$ if it satisfies the following conditions:

1. I and J both admit the small object argument.
2. The fibrations are given by $\mathrm{rlp}(J)$.
3. The acyclic fibrations are given by $\mathrm{rlp}(I)$.

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 $\mathrm{cell}(I)$.

3^*. The acyclic fibrations are the retracts (in the arrow category) of $\mathrm{cell}(J)$.

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 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 $\mathrm{Id} : \mathbf{C}_0 \leftrightarrows \mathbf{C} : \mathrm{Id}$ 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 $S \subset \mathrm{hom}(\mathbf{C})$ of cofibrations with cofibrant domain. First, the notion of "S-local objects" is introduced:

Definition 12.5.16 (Local object). A fibrant object x 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 12.5.17. Every weak equivalence is an S-local weak equivalence: $W \subset W_S$.

Construction 12.5.18 (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 as the same category but with the following model structure:

[10] *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.

- **Cofibrations**: $\mathrm{cof}(\mathbf{C})$
- **Acyclic cofibrations**: cofibrations that are S-local weak equivalences

If it exists, the Bousfield localization $L_S\mathbf{M}$ of $\mathbf{M}$ at S is the universal left Quillen functor $\gamma : \mathbf{M} \to L_S\mathbf{M}$ such that its left derived functor sends the image of S to isomorphisms in $\mathbf{Ho}(\mathbf{M})$. The identity functor gives a Quillen adjunction between $\mathbf{M}$ and $L_S\mathbf{M}$ and the induced adjunction on homotopy categories defines a reflective localization.

Remark 12.5.19. The notions of local object/morphisms can be slightly generalized (they give equivalent localizations). A **S-local morphism** $f : a \to b$ is a morphism such that for all objects $x \in \mathrm{ob}(\mathbf{M})$ for which the map

$$\mathbf{M}(g^{\mathrm{c}}, x^{\mathrm{f}}) : \mathbf{M}(d^{\mathrm{c}}, x^{\mathrm{f}}) \to \mathbf{M}(c^{\mathrm{c}}, x^{\mathrm{f}}) \tag{12.27}$$

is a weak equivalence for all $g \in S$ (x is said to be **S-local**), the map

$$\mathbf{M}(f^{\mathrm{c}}, x^{\mathrm{f}}) : \mathbf{M}(b^{\mathrm{c}}, x^{\mathrm{f}}) \to \mathbf{M}(a^{\mathrm{c}}, x^{\mathrm{f}}) \tag{12.28}$$

is also a weak equivalence. This is the homotopical version of Definition 4.2.19, after cofibrant (resp. fibrant) replacement of the domain (resp. codomain), the Yoneda embedding of every S-local object is weak equivalence, where S-local objects are those objects whose Yoneda embedding maps morphisms in S to weak equivalences.

Definition 12.5.20 (Combinatorial model category). A locally presentable, cofibrantly generated model category.

Theorem 12.5.21 (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).*

?? FINISH ??

12.6 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 12.6.1 (Homotopy (co)products). In ordinary categories the universal property of a product (and dually for coproducts) characterizes it as an object with an isomorphism

$$\mathbf{C}(-, c) \cong \prod_{i\in I} \mathbf{C}(-, c_i) \tag{12.29}$$

such that every I-indexed collection of morphisms $f_i : x \to c_i$ 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 the hom-space $\mathbf{M}(x, c_i)$. This leads to the definition of a homotopy product as an object c together with a natural weak (homotopy) equivalence

$$\psi_x : \mathbf{C}(x, c) \simeq \prod_{i\in I} \mathbf{C}(x, c_i). \tag{12.30}$$

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 12.6.2 (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 $(\mathrm{co})\lim : [\mathbf{B}, \mathbf{A}] \to \mathbf{A}$.

Remark 12.6.3. The reason for why (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 12.6.4 (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 12.3.37.

Example 12.6.5 (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.[11] If the Reedy structure is such that the constant functor preserves cofibrations, then this functor is left Quillen and Ken Brown's lemma 12.3.35 implies that its right Quillen adjoint, the limit functor, preserves weak equivalences. In this case one can define the **homotopy limit** $\mathrm{holim}\, D$ as the functor $\lim(D \circ Q_{\mathrm{f}})$, where Q_{f} is the fibrant replacement-functor (which in this case acts pointwise). A dual construction gives rise to **homotopy colimits**.

12.6.1 Simplicially enriched diagrams

In the setting where diagrams are enriched over **sSet**, one can define homotopy (co)limits in a more sophisticated way.

Definition 12.6.6 (Homotopy colimit). Consider a diagram $D : \mathbf{I} \to \mathbf{C}$ with $\mathbf{C}$ copowered over **sSet**. The homotopy colimit of D is defined as the following tensor product 4.7.19:

$$\mathrm{hocolim}\, D := N(-/\mathbf{I}) \otimes_{\mathbf{I}} D = \int^{i\in\mathbf{I}} N(i/\mathbf{I}) \cdot Di, \tag{12.31}$$

where N is the nerve functor 12.1.10.

A similar definition for homotopy limits can be used when the category is powered over **sSet**:

Definition 12.6.7 (Homotopy limit). Consider a diagram $D : \mathbf{I} \to \mathbf{C}$ with $\mathbf{C}$ powered over **sSet**. The homotopy limit of D is defined as the following hom-like object:

$$\mathrm{holim}\, D := \int_{i\in\mathbf{I}} [N(\mathbf{I}/i), Di]. \tag{12.32}$$

This is exactly the characterization of the homotopy limit as the **sSet**-natural transformations between $N(\mathbf{I}/-)$ and D.

[11] This is also true when $\mathbf{R}$ is not Reedy.

Remark 12.6.8 (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 12.3.28 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 12.6.9 (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 $D : \mathfrak{C}\mathbf{I} \rightarrow \mathbf{C}$. 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 $\mathfrak{C}(\mathbf{I} \times \Delta[1]) \rightarrow \mathbf{C}$ 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:

$$\mathbf{CohDgrm}(\mathbf{I}, \mathbf{C}) \cong (N\mathbf{C})^{\mathbf{I}}, \tag{12.33}$$

where N is the simplicial nerve functor.

?? COMPLETE ??

12.7 ∞-categories

12.7.1 Simplicial approach

The first approach to ∞-category theory is the simplicial one. The motivation is Property 12.4.4, which relates the categorical structure to the existence of certain horn fillers. The generalization is then given by the notion of quasicategories 12.4.5.

Theorem 12.7.1 (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 12.5.21, can be presented by simplicial presheaves.*

Chapter 13

Topos theory ♣

The main reference for this chapter is [6, 118]. For an introduction to stacks and descent theory, see [71].

13.1 Elementary topoi

Definition 13.1.1 (Subobject classifier). Consider a finitely complete category (in fact, the existence of a terminal object suffices). A subobject classifier is a mono[1] $\mathtt{true} : 1 \hookrightarrow \Omega$ 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:

$$\begin{tikzcd}
x \arrow[r, "\exists!"] \arrow[d, hook, "\phi"'] \arrow[dr, phantom, "\text{pb}"] & 1 \arrow[d, hook, "\mathtt{true}"] \\
y \arrow[r, "\exists!\chi"'] & \Omega
\end{tikzcd}$$

Figure 13.1: Subobject classifier.

Alternative Definition 13.1.2. Consider a well-powered category $\mathbf{C}$. The assignment of subobjects $\mathrm{Sub}(x)$ to an object $x \in \mathrm{ob}(\mathbf{C})$ defines a contravariant functor $\mathrm{Sub} : \mathbf{C} \to \mathbf{Set}$. A subobject classifier Ω is a representation of this functor, i.e. the following isomorphism is natural in x:

$$\mathrm{Sub}(x) \cong \mathbf{C}(x, \Omega). \tag{13.1}$$

Example 13.1.3 (Indicator function). The category **Set** has a subobject classifier, the 2-element set $\{\mathtt{true}, \mathtt{false}\}$. The morphism $\chi : S \to \Omega$ is the indicator function

$$\chi_S(x) = \begin{cases} \mathtt{true} & x \in S \\ \mathtt{false} & x \notin S. \end{cases} \tag{13.2}$$

Definition 13.1.4 (Elementary topos). An elementary topos is a finitely complete Cartesian closed category admitting a subobject classifier. Equivalently, one can define an elementary topos as a finitely complete category in which all power objects exist.

[1]The symbol for this morphism will become clear in Section 13.2.

The power object Px of $x \in \mathrm{ob}(\mathcal{E})$ is related to the subobject classifier Ω by the following relation:

$$Px = \Omega^x. \tag{13.3}$$

Remark 13.1.5 (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 13.1.6 (Fundamental theorem of topos theory). *Let $\mathcal{E}$ be a topos. For every object $x \in \mathrm{ob}(\mathcal{E})$ the slice category $\mathcal{E}_{/x}$ is also a topos. The subobject classifier is given by $\pi_2 : \Omega \times x \to x$.*

Property 13.1.7 (Balanced). All monos in a topos are regular. Hence, every mono arises as an equalizer and every epic equalizer is necessarily an isomorphism. It follows that every topos is balanced 4.4.5.

Property 13.1.8 (Epi-mono factorization). Every morphism $f : x \to y$ in a topos factorizes uniquely as an epi followed by a mono:

$$x \overset{e}{\twoheadrightarrow} z \overset{m}{\hookrightarrow} y. \tag{13.4}$$

The mono is called the **image** of f.

13.2 Internal logic

In this subsection general finitely complete categories that admit a subobject classifier are considered (they do not have to be elementary topoi).

Definition 13.2.1 (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 13.2.2 (Internal Heyting algebra). For all objects x in an elementary topos, the poset of subobjects $\mathrm{Sub}(x)$ has the structure of a Heyting algebra 2.6.34. Hence, every topos canonically gives an external Heyting algebra, namely $\mathrm{Sub}(1)$. Furthermore, every power object is an internal Heyting algebra. This in particular includes the subobject classifier $\Omega = P1$.

?? COMPLETE ??

13.3 Geometric morphisms

Definition 13.3.1 (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 13.3.2 (Logical morphism). Let $\mathcal{E}, \mathcal{F}$ be (elementary) topoi. A morphism $f : \mathcal{E} \to \mathcal{F}$ is called a logical morphism if it preserves finite limits, exponential objects and subobject classifiers.

Property 13.3.3. If a logical morphism has a left adjoint then it also has a right adjoint.

Definition 13.3.4 (Geometric morphism). Let $\mathcal{E}, \mathcal{F}$ be (elementary) topoi. A geometric morphism $f : \mathcal{E} \to \mathcal{F}$ consists of an adjunction

$$\begin{tikzcd} \mathcal{E} \arrow[r, shift right, "f_*"'] \arrow[r, phantom, "\perp"] & \mathcal{F} \arrow[l, shift right, "f^*"'] \end{tikzcd}$$

where the left adjoint is left exact. The right adjoint f_* is called the **direct image** part of f and the left adjoint is called the **inverse image** part. If f^* itself has a left adjoint, then f is said to be **essential**.

Definition 13.3.5 (Geometric embedding). A geometric morphism for which the direct image part is fully faithful.

Property 13.3.6 (Characterization of geometric embeddings). Let $f : \mathcal{E} \to \mathcal{F}$ be a geometric embedding and let $W \subset \hom(\mathcal{F})$ 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 and the localization $\mathcal{F}[W^{-1}]$ at W (Definition 12.2.5).

Property 13.3.7 (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 13.3.8 (Topological spaces). Every continuous function $f : X \to Y$ induces a geometric morphism

$$\mathbf{Sh}(X) \underset{f_*}{\overset{f^*}{\underset{\longrightarrow}{\overset{\longleftarrow}{\perp}}}} \mathbf{Sh}(Y), \tag{13.5}$$

where the direct image functor f_* is defined as

$$f_*F(U) := F(f^{-1}U) \tag{13.6}$$

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.

By the previous example 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 13.3.9 (Point). A point of a topos $\mathcal{E}$ is a geometric morphism $\mathbf{Set} \to \mathcal{E}$.

Notation 13.3.10 (Category of topoi). The category of elementary topoi and geometric morphisms is a 2-category. It is denoted by **Topos**.

In fact, 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 [6] and the "inverse image convention" is used, i.e. a 2-morphism $f \Rightarrow g$ consists of natural transformations $f^* \Rightarrow g^*$ and $g_* \Rightarrow f_*$.

13.4 Grothendieck topos

Definition 13.4.1 (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 $\mathbf{C}_{/x}$. This means that S is a subset of $\mathrm{ob}(\mathbf{C}_{/x})$ that is closed under precomposition, i.e. if $y \to x \in S$ and $z \to y \in \mathrm{hom}(\mathbf{C})$, $z \to y \to x \in S$.

All of this can be summarized by saying that a sieve on an object $x \in \mathrm{ob}(\mathbf{C})$ is a subfunctor of the hom-functor $\mathbf{C}(-, x)$.

Example 13.4.2 (Maximal sieve). Let $\mathbf{C}$ be a category. The maximal sieve on $x \in \mathrm{ob}(\mathbf{C})$ is the collection of all morphisms $\{f \in \mathrm{hom}(\mathbf{C}) \mid \mathrm{cod}(f) = x\}$ or, equivalently, all of $\mathrm{ob}(\mathbf{C}_{/x})$.

Example 13.4.3 (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:

$$f^*S(x) = \big\{g \,\big|\, f \circ g \in S(y)\big\}. \tag{13.7}$$

Property 13.4.4 (Presheaf topos). Consider the presheaf category $\mathbf{Psh}(\mathbf{C})$ for an arbitrary (small) category $\mathbf{C}$. This category is an elementary topos where the subobject classifier is defined on each object in the following way:

$$\underline{\Omega}(x) := \{S \mid S \text{ is a sieve on } x\}. \tag{13.8}$$

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 $\texttt{true} : \underline{1} \hookrightarrow \underline{\Omega}$ is defined as the natural transformation assigning to every object its maximal sieve. For every subobject $\underline{K} \hookrightarrow \underline{X}$ the characteristic morphism χ_K is defined as follows. Consider an object $c \in \mathrm{ob}(\mathbf{C})$ and element $x \in \underline{X}(c)$. The component $\chi_K|_c$ is then given by

$$\chi_K|_c(x) := \{f \in \mathbf{C}(d, c) \mid \underline{X}(f)(x) \in \underline{K}(d)\}. \tag{13.9}$$

The following definition is due to *Giraud* (for the original definition using the notion of a *cover*, see the end of this section):

Definition 13.4.5 (Grothendieck topology). A Grothendieck topology on a category is a map J assigning to every object a collection of sieves satisfying the following conditions:

1. **Identity**[2]: For every object x 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. 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 $x \in \mathrm{ob}(\mathbf{C})$ is called a **cover**[3] of x if the sieve generated by these morphisms is a covering sieve on x.

Example 13.4.6 (Topological spaces). These conditions have the following interpretation in the case of topological spaces:

- The collection of all open subsets covers a space U.

[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.

- 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 13.4.7 (Site). A (small) category equipped with a Grothendieck topology J.

Definition 13.4.8 (Matching family). Consider a presheaf $F \in \mathbf{Psh}(\mathbf{C})$ together with a sieve S on $x \in \mathrm{ob}(\mathbf{C})$. A matching family for S with respect to F is a natural transformation $\alpha : S \Rightarrow F$ between S, regarded as a subfunctor of $\mathbf{C}(-, x)$, and F.

More explicitly, it is an assignment of an element $x_f \in Fd$ to every morphism $(f : y \to x) \in S$ such that

$$F(g)(x_f) = x_{f\circ g} \tag{13.10}$$

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:

$$F(f')(x_f) = F(g')(x_g). \tag{13.11}$$

Given such a matching family, one calls an element $a \in Fx$ an **amalgamation** if it satisfies

$$F(f)(a) = x_f \tag{13.12}$$

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 13.4.9. 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 13.4.10 (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 4.2.19:

Alternative Definition 13.4.11 (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 $\eta \in \mathcal{S}$:

$$Fx \cong \mathbf{Psh}(\mathcal{Y}x, F) \xrightarrow{\mathbf{Psh}(\eta, F)} \mathbf{Psh}(S, F). \tag{13.13}$$

This is also called the **descent condition** for sheaves. In this context the collection of matching families $\mathrm{Match}(S, F) := \mathbf{Psh}(S, F)$ for a sieve S with respect to a presheaf F is often called the **descent object** of S with respect to F.

[4] If there exists at most one amalgamation, the presheaf is said to be **separated**.

Example 13.4.12 (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 $(\mathbf{Open}(X), J_{\mathbf{Open}(X)})$. 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 13.4.8 above, gives the second part of Definition 9.2.1. 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.

Cover, Matching family, Unique amalgamation

$$\begin{tikzcd}[column sep=small]
 & U_i \cap U_j \arrow[dl, hook] \arrow[dr, hook] & & & & x_i|_{U_i\cap U_j} = x_j|_{U_i\cap U_j} \arrow[dl] \arrow[dr] & & & & x_i|_{U_i\cap U_j} = x_j|_{U_i\cap U_j} \arrow[dl] \arrow[dr] & \\
U_i \arrow[dr, hook] & & U_j \arrow[dl, hook] \arrow[r, "(1)"] & {} & x_i & & x_j \arrow[r, "(2)"] & {} & x_i = x|_{U_i} \arrow[dr] & & x_j = x|_{U_j} \arrow[dl] \\
 & U & & & & & & & & x &
\end{tikzcd}$$

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.

Example 13.4.13 (Canonical topology). The canonical topology on a category is the largest 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 13.4.14 (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 13.4.15 (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 13.4.16. Every Grothendieck topos is an elementary topos.

Property 13.4.17. For every Grothendieck topos there exists a site of definition for which the Grothendieck topology is (sub)canonical.

Construction 13.4.18 (Sheafification). Given a presheaf $\mathcal{F}$, one can construct a sheaf $\overline{\mathcal{F}}$ along the same lines of Construction 9.2.19.

Definition 13.4.19 (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:

$$\Gamma : \mathcal{E} \to \mathbf{Set} : x \mapsto \mathcal{E}(1, x). \tag{13.14}$$

When $\mathcal{E}$ is the sheaf topos over a topological space, this is exactly the global sections functor 9.2.9. 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.

A different approach for defining sheaf topoi is through an embedding of sheaves into presheaves.

Definition 13.4.20 (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 12.2.1 closed under pullbacks along morphisms out of representable presheaves.

Property 13.4.21 (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 $\mathrm{im}(f) \to Y$ 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 $X \in \mathrm{ob}(\mathbf{C})$ if $F \hookrightarrow \mathcal{Y}X$ is a local epi.

Alternative Definition 13.4.22 (Sheaf topos). A category $\mathbf{Sh}(\mathbf{C})$ equipped with a geometric embedding into $\mathbf{Psh}(\mathbf{C})$.

Proof of equivalence. By Property 13.3.6 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 13.4.23 (Descent condition). This is essentially a restatement of the descent condition 13.4.11. Covering sieves, regarded as subfunctors, are in particular local isomorphisms. Stability of sieves under pullback together with the co-Yoneda lemma 4.4.68, which says that every presheaf is a colimit of representables, generates the full collection of local isomorphisms.

As a last point the weaker notion of coverages is introduced:

Definition 13.4.24 (Coverage). Let $\mathbf{C}$ be a category. A coverage on $\mathbf{C}$ is a map that assigns to every object $x \in \mathrm{ob}(\mathbf{C})$ a collection of families $\{f : y \to x\} \subset \hom(\mathbf{C})$ 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.

It should be clear that every coverage generates a sieve (the smallest sieve containing the coverage). Furthermore, although coverages are weaker and easier to handle, they are in fact equivalent for the purpose of sheaf theory:

Property 13.4.25. 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.

13.4.1 Topological sheaves

See Chapter 9 for the application of sheaves to topology.

Property 13.4.26 (Presheaf topos). Consider the presheaf category

$$\mathbf{Psh}(X) := \mathbf{Psh}(\mathbf{Open}(X))$$

over a topological space (X, τ). Unpacking Property 13.4.4 shows that this category is an elementary topos where the subobject classifier Ω is defined as follows:

$$\Omega(U) := \{V \in \tau \mid V \subseteq U\}. \tag{13.15}$$

Construction 13.4.27 (Sheaves and étalé bundles). Let X be a topological space. The functor

$$I : \mathbf{Open}(X) \to \mathbf{Top}_{/X} : U \mapsto (U \hookrightarrow X)$$

induces the following adjunction:

$$\mathbf{Top}_{/X} \underset{\Gamma}{\overset{E}{\underset{\longrightarrow}{\overset{\longleftarrow}{\perp}}}} \mathbf{Psh}(X). \tag{13.16}$$

The slice category on the right-hand side is the category of (topological) bundles (Chapter 31) 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 obtain the **étalé space** and **sheaf categories** respectively:

$$\mathbf{Et}(X) \cong \mathbf{Sh}(X). \tag{13.17}$$

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 étalé spaces 7.2.17.

Property 13.4.28 (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 just Construction 9.2.15.

The fact that the counit of the adjunction 13.4.27 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 13.4.29 (Petit and gros topoi). 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 $\mathbf{Sh}(X)$. It represents X as some kind of generalized space. (By Construction 13.4.27 the objects in a small topos are the étalé spaces over a given base space.) However, one can also build a topos whose objects are generalized spaces. To this end, choose a site S of "probes" and call the sheaf topos $\mathbf{Sh}(S)$ a gros topos. See Section 42.2 for more information.

Property 13.4.30 (Localic reflection). Mapping a topological space to its sheaf of continuous sections defines a functor $\mathbf{Sh} : \mathbf{Top} \to \mathbf{Topos}$ by Example 13.3.8. When restricted to the full subcategory of sober spaces 7.8.4, this functor becomes fully faithful. Generalizing to sober locales even gives a reflective inclusion 4.2.27.

This property states that no information is lost when regarding (sober) topological spaces as sheaf topoi. This also explains the name "petit topos".

13.4.2 Lawvere-Tierney topology

Definition 13.4.31 (Lawvere-Tierney topology). As noted in Section 13.2 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 meet-semilattice, where the meet is given by the pullback of morphisms. This internal poset, viewed as an internal category, admits the construction of a closure operator $j : \Omega \to \Omega$ (Definition 4.3.25) satisfying the following condition:

$$j \circ \wedge = \wedge \circ (j \times j). \tag{13.18}$$

This condition states (in a nontrivial way) that j is (internally) order-preserving.

Remark 13.4.32. The condition satisfied by the unit morphism in the definition of a closure operator can also be reformulated in this context as follows:

$$j \circ \texttt{true} = \texttt{true}. \tag{13.19}$$

The Lawvere-Tierney operator also induces a "closure operator" on all posets $\mathrm{Sub}(x)$ in the topos. Given an object x and a subobject $u \in \mathrm{Sub}(x)$, one defines the closure $j_*(u) \in \mathrm{Sub}(x)$ as the subobject classified by the characteristic morphism $j \circ \chi_u : x \to \Omega$.

Definition 13.4.33 (Dense object). Given a Lawvere-Tierney topology $j : \Omega \to \Omega$, a subobject $u \in \mathrm{Sub}(x)$ is said to be dense (in x) if it satisfies $j_*(u) = x$.

Alternative Definition 13.4.34 (Sheaf). Given a Lawvere-Tierney topology $j : \Omega \to \Omega$ on a topos $\mathcal{E}$, one calls an object $s \in \mathrm{ob}(\mathcal{E})$ a j-sheaf if for all dense morphisms $u \hookrightarrow x$ the induced map

$$\mathcal{E}(x, s) \to \mathcal{E}(u, s)$$

is a bijection.

Property 13.4.35. For the presheaf topos on a small category $\mathbf{C}$, the Grothendieck topologies on $\mathbf{C}$ and Lawvere-Tierney topologies on $\mathbf{Psh}(\mathbf{C})$ are equivalent.

Sketch of proof. Since a Grothendieck topology assigns to every object a collection of sieves, Property 13.4.4 implies that $J(x) \subseteq \Omega_{\mathbf{Psh}}(x)$ for all $x \in \mathrm{ob}(\mathbf{C})$. 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 $x \in \mathrm{ob}(\mathbf{C})$. From the conditions of Lawvere-Tierney topologies one can then prove that this collection satisfies the conditions of a Grothendieck topology.

Remark 13.4.36. It follows that Lawvere-Tierney topologies generalize Grothendieck topologies from presheaf topoi to arbitrary elementary topoi.

13.5 Stacks

13.5.1 2-sheafs

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 13.5.1 (2-presheaf). Consider a 2-category $\mathbf{C}$. A 2-presheaf on $\mathbf{C}$ is a pseudofunctor $F : \mathbf{C}^{op} \to \mathbf{Cat}$. 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 13.5.2 (2-coverage). Virtually the same as an ordinary coverage 13.4.24, 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 13.4.11), one can define 2-sheaves through a descent condition:

Definition 13.5.3 (2-sheaf). A 2-presheaf $F : \mathbf{C}^{op} \to \mathbf{Cat}$ 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:

$$Fc \cong \mathbf{Psh}_2(\mathcal{Y}c, F) \to \mathbf{Psh}_2(S, F), \tag{13.20}$$

where the fist equivalence is just the 2-Yoneda lemma.

Remark 13.5.4. 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 $\mathbf{Set} \hookrightarrow \mathbf{Cat}$ 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 13.5.5 (Prestack of groupoids). Consider a category $\mathbf{C}$. A prestack of groupoids on $\mathbf{C}$ is a **Grpd**-valued prestack on $\mathbf{C}$.

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**: The natural transformations $\alpha : F \Rightarrow G$ (note that the components are themselves functors).
- **Morphisms**: The "strict modifications" in the sense that they map objects in $\mathbf{C}$ to natural transformations satisfying the whiskering condition (see also Definition 4.9.12)

$$\mathbb{1}_{Ff} \cdot \mathfrak{m}_b = \mathfrak{m}_a \cdot \mathbb{1}_{Gf}. \tag{13.21}$$

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 13.5.6 (Čech groupoid). Consider a site $(\mathbf{C}, J)$. To every covering family $\mathcal{U} \equiv \{f_i : x_i \to x\}_{i \in I}$ one can assign an internal groupoid in presheaves $C(\mathcal{U})$ consisting of the following data:

- **Objects**: $\bigsqcup_i \mathcal{Y}x_i$
- **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 $y \in \mathrm{ob}(\mathbf{C})$ the groupoid consisting of the following data:

- **Objects**: The pairs $(i, g_i : y \to x_i)$ where $x_i \in \mathcal{U}$.
- **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 13.4.8, shows that one could presume that the Čech groupoid is related to the matching families. This intuition is indeed correct:

Property 13.5.7 (Matching families). Any ordinary presheaf F can be considered to be **Grpd**-valued by postcomposing with the embedding $\mathbf{Set} \hookrightarrow \mathbf{Grpd}$. For any covering family $\mathcal{U}$, there exists an isomorphism

$$[\mathbf{C}^{op}, \mathbf{Grpd}](C(\mathcal{U}), F) \cong \mathbf{Psh}_2(\mathcal{U}, F). \tag{13.22}$$

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 42 to define (higher) smooth groupoids.

People 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 13.5.8 (Čech nerve). Consider a morphism $f : y \to x$ in a category $\mathbf{C}$. The Čech nerve $C_\bullet(f)$ is the simplicial object 12.1.3 that is defined as the $(k+1)$-fold pullback of f along itself in degree k. For a covering family $\mathcal{U} \equiv \{f_i : x_i \to x\}$, the Čech nerve is defined as $C_\bullet(\mathcal{U}) := C_\bullet(\bigsqcup_i x_i \to x)$.

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* (31.1) known for example from the study of fibre bundles.

13.5.2 Stacks on a 1-site

For the definition of stacks, one needs the notions of fibred categories or, equivalently, pseudofunctors as defined in Section 4.3.1. 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 in $(f : c \to d)\mathbf{B}$ with $d \in \mathrm{im}(\Pi)$ and each lift $y \in \mathbf{A}_d$ there exists at least one inverse image in $(\widetilde{f} : x \to y) \in \mathbf{A}$ of f. By the Grothendieck construction every fibred category gives rise to a pseudofunctor $F : \mathbf{B}^{op} \to \mathbf{Cat}$ by sending objects to their fibres under Π and sending morphisms f to their pullback functors f^*.

Definition 13.5.9 (Descent datum). Consider a category $\mathbf{C}$ with a covering family $\mathcal{U} \equiv \{f_i : x_i \to x\}$ and a pseudofunctor $F : \mathbf{C}^{op} \to \mathbf{Cat}$. 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 (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**:

$$\pi_{ik}^* f_{ik} = \pi_{ij}^* f_{ij} \circ \pi_{jk}^* f_{jk}. \tag{13.23}$$

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

$$\pi_i^* \phi_i \circ f_{ij} = f'_{ij} \circ \pi_j^* \phi_j. \tag{13.24}$$

The category of descent data for $\mathcal{U}$ with respect to F will be denoted by $\mathrm{Descent}(\mathcal{U}, F)$.

Construction 13.5.10. Consider an object ξ in Fx. From this object one can construct a descent datum as follows. The objects 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 $Fx \to \mathrm{Descent}(\mathcal{U}, F)$. It can be shown that this construction is independent of a choice of cleavage up to equivalence.

Definition 13.5.11 (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 $x \in \mathrm{ob}(\mathbf{C})$, the functor $Fx \to \mathrm{Descent}(\mathcal{U}, F)$ is fully faithful.
- F is called a **stack** if for each covering family $\mathcal{U}$ on $x \in \mathrm{ob}(\mathbf{C})$ the functor $Fx \to \mathrm{Descent}(\mathcal{U}, F)$ is an equivalence.

This is a generalization of the descent condition 13.4.11. This can be seen by observing that $\mathrm{Descent}(\mathcal{U}, F) \cong \mathbf{Psh}_2(S(\mathcal{U}), F)$, where $S(\mathcal{U})$ is the sieve generated by $\mathcal{U}$ regarded as a fibred category.

A more conceptual (although completely equivalent) generalization from (1-)sheaves to 2-sheaves can be obtained by starting from Property 13.5.7. There it was shown that matching families for (1-)presheaves can be obtained as natural transformations from the Čech groupoid.

Property 13.5.12 (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 $(c, \{c_i\}, \{c_{ij}\}, \{c_{ijk}\})$, where $c_i \in Fx_i$, 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:

$$\mathrm{Descent}(\mathcal{U}, F) \cong [\mathbf{C}^{op}, \mathbf{Cat}](C(\mathcal{U}), F). \tag{13.25}$$

?? FINISH THIS ??

Remark 13.5.13 (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 $\mathbf{C}^{op} \to \mathbf{Set}$, 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 morphism 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 (13.11).

13.6 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 12 will be used.

Definition 13.6.1 (∞-groupoid). Objects of the full simplicial subcategory of $\mathbf{sSet}_{Quillen}$ on Kan complexes. From Property 12.4.4, 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 13.6.2 ($(\infty, 1)$-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 12.1.10:

$$[\mathbf{C}, \mathbf{D}] := |\mathbf{sSet}(N\mathbf{C}, N\mathbf{D})|. \tag{13.26}$$

Property 13.6.3 (Čech model structure). For any small category $\mathbf{C}$, the ∞-category of $\infty\mathbf{Grpd}$-valued ∞-sheaves can be represented by the category $[\mathbf{C}^{op}, \mathbf{sSet}]$ of simplicial presheaves on $\mathbf{C}$ by a theorem of *Lurie* 12.7.1, 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 $[\mathbf{C}^{op}, \mathbf{sSet}]$ 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

$$\mathrm{Hom}(M, X) \to \mathrm{Hom}(\mathcal{C}(\mathcal{U}), X) \tag{13.27}$$

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 13.4.11:

Definition 13.6.4 (∞-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:

$$Fx \simeq \mathbf{Psh}_\infty(\mathcal{C}(\mathcal{U}), F) \tag{13.28}$$

for all covers $\mathcal{U}$ of x, where $\simeq$ denotes a weak equivalence.

Definition 13.6.5 (∞-stack). An $(\infty, 1)$-sheaf taking values in $\infty\mathbf{Grpd}$.

Property 13.4.19 can be generalized as follows:

Property 13.6.6. 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 13.6.7 (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 $\mathbf{B}G$. The global sections functor maps this stack to the delooping groupoid BG.

Definition 13.6.8 (Mapping stack). Consider two ∞-stacks $X, Y \in \mathbf{Sh}_{(\infty,1)}(\mathbf{C})$. The mapping stack is defined as follows:

$$[X, Y](U) := \mathbf{Sh}_{(\infty,1)}(\mathbf{C})(X \times U, Y), \tag{13.29}$$

where on the right-hand side, U denotes the representable ∞-stack.

?? FINISH (PERHAPS MOVE infinity-CATEGORY STUFF TO CHAPTER ON MODEL THEORY) ??

13.7 Cohomology

In this section, cohomology will be generalized to the ∞-categorical setting.

First, take a topological space X and an ∞-groupoid G. Geometric realization 12.1.11 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:

$$H(X;G) := \pi_0\mathbf{Top}(X, |G|). \tag{13.30}$$

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

$$\overline{H}(X;G) := \pi_0\mathbf{Sh}_{(\infty,1)}(X)(X, \mathrm{LConst}\, G) \cong \pi_0 \circ \Gamma \circ \mathrm{LConst}(G). \tag{13.31}$$

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 13.7.1 (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

  $$H(X;A) := \pi_0\mathbf{H}(X,A) = \mathrm{Hom}_{\mathbf{Ho(H)}}(X,A), \tag{13.32}$$

  where $\mathbf{Ho}(\mathbf{H})$ is the homotopy category 12.3.28 of $\mathbf{H}$, is called the intrinsic cohomology of X with coefficients in A.

If the object A admits an n-delooping $\mathbf{B}^nA$, the n^{th} cohomology group of X is defined as

$$H^n(X;A) := H(X;\mathbf{B}^nA). \tag{13.33}$$

Example 13.7.2 (Singular cohomology). Consider a topological space X. For every group G one can define the first delooping 4.10.2, 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 13.7.3 (Group cohomology). Consider a (discrete) group G together with its delooping groupoid $\mathbf{B}G$. The group cohomology 3.4.1 of a group with coefficients in an Abelian group A is given by the intrinsic cohomology of $\infty\mathbf{Grpd}$ of the delooping groupoids:

$$H(G;A) \cong \pi_0\infty\mathbf{Grpd}(\mathbf{B}G, \mathbf{B}A). \tag{13.34}$$

By replacing the topos $\mathbf{H}$ by a slice topos $\mathbf{H}_{/X}$ one obtains twisted cohomology:

Definition 13.7.4 (Twisted cohomology). Consider a $(\infty,1)$-topos $\mathbf{H}$ with some object $X \in \mathrm{ob}(\mathbf{H})$. 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 (hte connected components of) the following mapping space:

$$\mathbf{H}_{/\mathbf{BAut}(A)}(\chi, \rho_A). \tag{13.35}$$

13.8 Cohesion

In this section the terminology "(Grothendieck) topos **over** a topos $\mathcal{S}$" will mean a topos equipped with a geometric morphism to $\mathcal{S}$.

Definition 13.8.1 (Local topos). Consider a topos $\mathcal{E}$ over a base topos $\mathcal{S}$. $\mathcal{E}$ is said to be ($\mathcal{S}$-)local if the geometric morphism $(f^* \dashv f_*) : \mathcal{E} \leftrightarrows \mathcal{S}$ 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 4.9.16.
- $f^!$ is Cartesian closed 4.6.23.

If one takes $\mathcal{S} = \mathbf{Set}$, the conditions are automatically satisfied since all functors are **Set**-indexed.

The right adjoint is sometimes called the **codiscrete object functor** coDisc (in fact, this terminology is applied more generally when $\mathcal{E}$ is just any category). If this functor exists, $\mathcal{E}$ is said to have **codiscrete objects**.

Property 13.8.2. A topos is local if and only if 1 is tiny 4.4.40.

Definition 13.8.3 (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

$$\Pi_0 : \mathcal{E} \to \mathbf{Set} : \bigsqcup_{i \in I} X_i \mapsto I \tag{13.36}$$

left adjoint to the discrete object functor (which is itself left adjoint to the global section functor). This functor is suitably called the **connected components functor**.

A topos is locally connected if and only if its global section geometric morphism is essential. More generally, a topos over some base topos $\mathcal{S}$ 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 13.8.4 (Connected topos). A topos over a base 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 the requirement that X is 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 13.8.5. Every local topos is connected.

Definition 13.8.6 (Cohesive topos). A local, strongly connected topos. This implies the existence of an adjoint quadruple $(\Pi_0, \mathrm{Disc}, \Gamma, \mathrm{coDisc})$ where both Disc and coDisc are fully faithful.

Property 13.8.7 (Cohesive modalities). The adjoint quadruple on a cohesive topos induces an adjoint triple of modalities 4.3.25, i.e. idempotent (co)monads (see Section 6.5 for a formal introduction in the context of type theory):

$$(\textstyle\int \dashv \flat \dashv \sharp) := (\mathrm{Disc} \circ \Pi_0 \dashv \mathrm{Disc} \circ \Gamma \dashv \mathrm{coDisc} \circ \Gamma). \tag{13.37}$$

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.

?? COMPLETE (e.g. work by Schreiber) ??

Part III

Calculus

Chapter 14

Calculus

14.1 General definitions

Definition 14.1.1 (Domain). A connected, open subset of $\mathbb{R}^n$.

Definition 14.1.2 (Factorial).

$$n! := n(n-1)\cdots 1 \tag{14.1}$$

Definition 14.1.3 (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

$$\forall f \in \mathcal{F}, x \in X : |f(x)| \leq F(x). \tag{14.2}$$

14.2 Sequences

Definition 14.2.1 (Limit superior). Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of real numbers. The limit superior is defined as follows:

$$\limsup_{n\to\infty} x_n := \inf_{n\geq 1} \sup_{k\geq n} x_k. \tag{14.3}$$

Definition 14.2.2 (Limit inferior). Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of real numbers. The limit superior is defined as follows:

$$\liminf_{n\to\infty} x_n := \sup_{n\geq 1} \inf_{k\geq n} x_k. \tag{14.4}$$

Property 14.2.3. A sequence $(x_n)_{n\in\mathbb{N}}$ converges pointwise if and only if

$$\limsup_{n\to\infty} x_n = \liminf_{n\to\infty} x_n. \tag{14.5}$$

14.3 Continuity

Definition 14.3.1 (Darboux function). A function that has the intermediate value property.

Theorem 14.3.2 (Darboux). *Every differentiable function defined on a closed interval has the intermediate value property 7.4.3.*

Corollary 14.3.3 (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 14.3.4 (Weierstrass's extreme value theorem). *Let $I = [a, b] \subset \mathbb{R}$ be a closed interval and let f be a continuous function defined on I. Then f attains a minimum and maximum at least once on I.*

Definition 14.3.5 (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

$$\sum_i (y_i - x_i) < \delta_\varepsilon, \tag{14.6}$$

the function f satisfies

$$\sum_i |f(y_i) - f(x_i)| < \varepsilon. \tag{14.7}$$

Property 14.3.6. The different types of continuity form the following hierarchy:

Lipschitz-continuous $\subset$ absolutely continuous $\subset$ uniformly continuous $\subset$ continuous.

Definition 14.3.7 (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:

$$V_{a,b}(f) := \sup_{P \in \mathcal{P}} \sum_{i=0}^{|P|-1} |f(x_{i+1}) - f(x_i)|, \tag{14.8}$$

where the supremum is taken over all partitions of $[a, b]$.

Property 14.3.8. Every function of bounded variation can be decomposed as the difference of two monotonically increasing functions.

Example 14.3.9. Every absolutely continuous function is of bounded variation.

14.4 Convergence

Definition 14.4.1 (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

$$\forall x \in \mathrm{dom}(f_n) : \lim_{n\to\infty} f_n(x) = f(x). \tag{14.9}$$

Definition 14.4.2 (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

$$\lim_{n\to\infty} \sup_{x\in\mathrm{dom}(f_n)} |f_n(x) - f(x)| = 0. \tag{14.10}$$

14.5 Series

14.5.1 Convergence tests

Property 14.5.1. A necessary condition for the convergence of a series $\sum_{i=n}^{\infty} a_i$ is that $\lim_{n\to\infty} a_n = 0$.

Property 14.5.2 (Absolute/conditional convergence). If $S' = \sum_{i=1}^{\infty} |a_i|$ converges, so does the series $S = \sum_{i=1}^{\infty} a_i$. Moreover, S is said to be absolutely convergent. If S converges but S' does not, S is said to be conditionally convergent.

Definition 14.5.3 (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 $S_{a,k} \geq S_{b,k}$.

Method 14.5.4 (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 14.5.5 (Maclaurin-Cauchy integral test). Let f be a nonnegative, continuous and monotonically decreasing function defined on the interval $[n, \infty[$. 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 14.5.6. 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 14.5.7. If the integral in the previous theorem converges, the series is bounded in the following way:

$$\int_n^{\infty} f(x)dx \leq \sum_{i=n}^{\infty} a_i \leq f(n) + \int_n^{\infty} f(x)dx. \tag{14.11}$$

Method 14.5.8 (d'Alembert's ratio test). Define the quantity

$$R := \lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right|. \tag{14.12}$$

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 14.5.9 (Cauchy's root test). Define the quantity

$$R := \limsup_{n\to\infty} \sqrt[n]{|a_n|}. \tag{14.13}$$

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 14.5.10 (Radius of convergences). The number $1/R$ is called the radius of convergence.

Remark 14.5.11. 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 14.5.12 (Gauss's test). If $a_n > 0$ for all n, one can write the ratio of successive terms as follows:

$$\left|\frac{a_n}{a_{n+1}}\right| = 1 + \frac{h}{n} + \frac{B(n)}{n^k}, \tag{14.14}$$

where $k > 1$ and $B(n)$ is a bounded function when $n \longrightarrow \infty$. The series converges if $h > 1$ and diverges otherwise.

Definition 14.5.13 (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

$$f(x) - \sum_{i=0}^{n} a_i x^i = O(x^{n+1}) \tag{14.15}$$

for all $x \in \mathbb{R}$.

14.6 Differentiation

Formula 14.6.1 (Derivative). Consider a function $f : \mathbb{R} \to \mathbb{R}$. The following limit is called the derivative of f at x (if it exists):

$$f'(x) := \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}. \tag{14.16}$$

If the derivative exists at every point of some interval I, then f is said to be differentiable on I.

Theorem 14.6.2 (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*

$$f'(c) = \frac{f(b) - f(a)}{b - a}. \tag{14.17}$$

Definition 14.6.3 (Differentiablity class). Let I be a set and let f be a function defined on I. If f is n times continuously differentiable on I (i.e. $f^{(i)}$ exists and is continuous for $i = 1, \dots, n$), then f is said to be of class $C^n(I)$.

Definition 14.6.4 (Smooth function). A function f is said to be smooth if it is of class C^∞.

Theorem 14.6.5 (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 14.6.6 (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:

$$f(y) = f(x) + f'(x)(y - x) + \frac{f''(x)}{2}(y - x)^2 + \cdots = \sum_{n=0}^{\infty} \frac{f^{(n)}(x)}{n!}(y - x)^n. \tag{14.18}$$

For the special case $x = 0$ the name **Maclaurin series** is sometimes used.

Definition 14.6.7 (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 14.6.8 (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:*

$$f(x) = f(x_0) + \sum_{i=1}^{n} (x^i - x_0^i) g_i(x_0), \tag{14.19}$$

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 14.6.9. This lemma gives a finite order approximation of the Taylor expansion of f.

Theorem 14.6.10 (Schwarz[1]). *Consider a twice differentibale function $f \in C^2(\mathbb{R}^n, \mathbb{R})$. The mixed partial derivatives of f coincide for all indices $i, j \leq n$:*

$$\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right) = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right). \tag{14.20}$$

Formula 14.6.11 (Derivative of $f(x)^{g(x)}$). Consider a function of the form

$$u(x) = f(x)^{g(x)}.$$

After taking the logarithm and applying the standard rules of differentiation, one can obtain the following expression:

$$\frac{d}{dx}\left[f(x)^{g(x)}\right] = f(x)^{g(x)}\left(\frac{dg}{dx}(x)\ln[f(x)] + \frac{g(x)}{f(x)}\frac{df}{dx}(x)\right). \tag{14.21}$$

Definition 14.6.12 (Euler operator). On the space $C^{r>1}(\mathbb{R}^n, \mathbb{R})$, the Euler operator $\mathbb{E}$ is defined as follows:

$$\mathbb{E} := \sum_{i=1}^{n} x_i \frac{\partial}{\partial x^i}. \tag{14.22}$$

Theorem 14.6.13 (Euler). *Let f be a homogeneous function, i.e.*

$$f(\lambda x_1, \ldots, \lambda x_n) = \lambda^n f(x_1, \ldots, x_n). \tag{14.23}$$

This function satisfies the following equality:

$$\mathbb{E}(f) = n f(x_1, \ldots, x_n). \tag{14.24}$$

?? COMPLETE (add multidimensional extensions) ??

14.7 Integration theory

Definition 14.7.1 (Improper Riemann integral).

$$\int_{-\infty}^{\infty} f(x)dx = \lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_a^b f(x)dx \tag{14.25}$$

One-sided improper integrals are defined in a similar fashion.

[1] Also called **Clairaut's theorem**.

Theorem 14.7.2 (First fundamental theorem of calculus). *Let f be a continuous function defined on an open interval I and consider an element $c \in I$. The following theorem establishes the relation between integration and differentiation:*

$$\exists F \in C^1(I) : \forall x \in I : F'(x) = f(x). \tag{14.26}$$

Furthermore, the function F is uniformly continuous on I and is given by the following integral:

$$F(x) = \int_c^x f(x')dx'. \tag{14.27}$$

Remark 14.7.3. 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.

Theorem 14.7.4 (Second fundamental theorem of calculus). *Let f be a C^1-function defined on the interval $[a, b]$.*

$$\int_a^b f'(x)dx = f(b) - f(a) \tag{14.28}$$

Formula 14.7.5 (Differentiation under the integral sign[2]).

$$\frac{d}{dx}\int_{a(x)}^{b(x)} f(x,y)dy = f(x,b(x)) \cdot b'(x) - f(x,a(x)) \cdot a'(x) + \int_{a(x)}^{b(x)} \frac{\partial f(x,y)}{\partial x}dy \tag{14.29}$$

Definition 14.7.6 (Borel transform). Consider the following function:

$$F(x) := \sum_{n=0}^{\infty} \frac{a_n}{n!}x^n. \tag{14.30}$$

If

$$\int_0^\infty e^{-t}F(xt)dt < \infty \tag{14.31}$$

for all $x \in \mathbb{R}$, then F is called the Borel transform of

$$f(x) = \sum_{n=0}^{\infty} a_n x^n. \tag{14.32}$$

Furthermore, the integral gives a convergent expression for f.

Proof. The function F is defined as follows:

$$F(x) := \sum_{n=0}^{\infty} \frac{a_n}{n!}x^n. \tag{14.33}$$

[2] This is a more general version of the *Leibniz integral rule*.

The Borel transform gives:

$$\begin{aligned}\int_0^\infty F(xt)e^{-t}dt &= \sum_{n=0}^N \int_0^\infty \frac{a_n}{n!}x^n t^n e^{-t}dt \\ &= \sum_{n=0}^N \frac{a_n}{n!}x^n \int_0^\infty t^n e^{-t}dt \\ &= \sum_{n=0}^N \frac{a_n}{n!}x^n \Gamma(n+1) \\ &= \sum_{n=0}^N a_n x^n,\end{aligned}$$

where the definition of the Gamma function 14.7.9 was used on line 3 and the relation (14.38) between the factorial function and the Gamma function was used on line 4. □

Theorem 14.7.7 (Watson). *The Borel transform F is unique if the function f is holomorphic on the domain* $\{z \in \mathbb{C} \mid |\arg(z)| < \frac{\pi}{2} + \varepsilon\}$.

14.7.1 Euler integrals

Formula 14.7.8 (Beta function). The beta function (also known as the **Euler integral of the first kind**) is defined as follows:

$$B(x,y) := \int_0^1 t^{x-1}(1-t)^{y-1}dt. \tag{14.34}$$

Formula 14.7.9 (Gamma function). The gamma function (also known as the **Euler integral of the second kind**) is defined as follows:

$$\Gamma(x) := \int_0^\infty t^{x-1}e^{-t}dt. \tag{14.35}$$

Formula 14.7.10. The following formula relates the beta and gamma functions:

$$B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}. \tag{14.36}$$

Property 14.7.11 (Recursion). The gamma function satisfies the following recursion relation for all points x in its domain:

$$\Gamma(x+1) = z\Gamma(x). \tag{14.37}$$

Formula 14.7.12 (Factorial). For integers $n \in \mathbb{N}$ the gamma function can be expressed in terms of the factorial:

$$\Gamma(n) = (n-1)!\,. \tag{14.38}$$

Formula 14.7.13 (Stirling). This formula (originally stated for the factorial of natural numbers) gives an asymptotic expansion of the gamma function:

$$\ln\Gamma(z) \approx z\ln z - z + \frac{1}{2}\ln\left(\frac{2\pi}{z}\right). \tag{14.39}$$

14.7.2 Gaussian integrals

Formula 14.7.14 (n-dimensional Gaussian integral). An integral of the form

$$I\Big(A,\vec{\boldsymbol{b}}\Big) := \int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\vec{\boldsymbol{x}}\cdot A\vec{\boldsymbol{x}} + \vec{\boldsymbol{b}}\cdot\vec{\boldsymbol{x}}\right) d^n x, \tag{14.40}$$

where A is a real symmetric matrix. By performing the transformation $\vec{\boldsymbol{x}} \to A^{-1}\vec{\boldsymbol{b}} - \vec{\boldsymbol{x}}$ and diagonalizing A, one can obtain the following expression:

$$I\Big(A,\vec{\boldsymbol{b}}\Big) = \sqrt{\frac{(2\pi)^n}{\det(A)}} \exp\left(\frac{1}{2}\vec{\boldsymbol{b}}\cdot A^{-1}\vec{\boldsymbol{b}}\right). \tag{14.41}$$

More generally one has the following result:

$$\int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\vec{\boldsymbol{x}}\cdot A\vec{\boldsymbol{x}}\right) f(\vec{\boldsymbol{x}}) d^n x = \sqrt{\frac{(2\pi)^n}{\det(A)}} \exp\left(\frac{1}{2}\sum_{i,j=1}^{n} A^{-1}_{ij}\partial_i\partial_j\right) f(\vec{\boldsymbol{x}})\Bigg|_{\vec{\boldsymbol{x}}=0}. \tag{14.42}$$

This result is sometimes called **Wick's lemma**.

Corollary 14.7.15. A functional generalization is given by:

$$\begin{aligned} I(iA,iJ) &= \int [d\varphi] \exp\left(-i\int_{\mathbb{R}^n\times\mathbb{R}^n} \varphi(x)A(x,y)\varphi(y)\, d^n x d^n y + i\int_{\mathbb{R}^n} \varphi(x)J(x)\, d^n x\right) \\ &= C\det(A)^{-1/2} \exp\left(\frac{i}{2}\int_{\mathbb{R}^n\times\mathbb{R}^n} J(x)A^{-1}(x,y)J(y)\, d^n x d^n y\right), \end{aligned} \tag{14.43}$$

where the analytic continuation $I(iA,iJ)$ of Equation (14.41) was used. One should pay attention to the normalization factor C which is infinite in general.

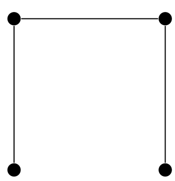
Method 14.7.16 (Feynman diagrams). The expansion of the exponential in the general expression for Gaussian integrals admits a diagrammatic expression. Let $f(\vec{\boldsymbol{x}})$ be a polynomial function of the coordinates.

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:

$$\int_{\mathbb{R}^n} \exp\left(-\frac{1}{2}\vec{\boldsymbol{x}}\cdot A\vec{\boldsymbol{x}}\right) x^{i_1}\cdots x^{i_k} d^n x = \sqrt{\frac{(2\pi)^n}{\det(A)}} \sum_{\sigma\in S_k} A^{-1}_{\sigma(i_1)\sigma(i_2)}\cdots A^{-1}_{\sigma(i_{k-1})\sigma(i_k)}. \tag{14.44}$$

To every coordinate dimension one can assign a vertex in the place, i.e. the object x_i can be interpreted as a real-valued function on the set of k 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 14.7.17 (Feynman diagrams). Some simple example 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:

$A^{-1}_{11}A^{-1}_{12}A^{-1}_{22} =$ 

Remark 14.7.18 (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 even imperative since, as mentioned above, the normalization factor diverges for infinite-dimensional spaces.

14.8 Convexity

Definition 14.8.1 (Convex set). A subset of X of a vector space V (Definition 20.1.1) 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 14.8.2 (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 $\lambda \in [0,1]$:

$$f\big(\lambda x + (1-\lambda)y\big) \leq t\lambda(x) + (1-\lambda)f(y). \tag{14.45}$$

For the definition of a **concave** function the inequality has to be turned around.

Property 14.8.3 (Linear map). A function $f : X \to \mathbb{R}$ is linear if and only if it is both convex and concave.

Theorem 14.8.4 (Karamata's inequality). *Consider an interval $I \subset \mathbb{R}$ and let $f : I \to \mathbb{R}$ be a convex function. If $(x_1, \ldots, x_n)$ is a tuple that majorizes $(y_1, \ldots, y_n)$, i.e.*

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \tag{14.46}$$

and

$$x_{(1)} + \cdots + x_{(k)} \geq y_{(1)} + \cdots + y_{(k)} \tag{14.47}$$

for all $k \leq n$, where $x_{(i)}$ denotes the i^{th} largest element of $(x_1, \ldots, x_n)$, then

$$\sum_{i=1}^{n} f(x_i) \geq \sum_{i=1}^{n} f(y_i). \tag{14.48}$$

The following inequality can be derived directly from the definition of convexity by induction:

Theorem 14.8.5 (Jensen's inequality). *Let f be a convex function and consider a point $\{a_i\}_{i \leq n}$ in the probability simplex Δ^n (Definition 8.2.3).*

$$f\left(\sum_{i=1}^{n} a_i x_i\right) \leq \sum_{i=1}^{n} a_i f(x_i). \tag{14.49}$$

Definition 14.8.6 (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

$$x \longrightarrow f'(x). \tag{14.50}$$

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

$$f^*(x^*) := \sup_x \big(x^*x - f(x)\big). \tag{14.51}$$

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^*:

$$\frac{df^*}{dx^*}(x^*) = x(x^*). \tag{14.52}$$

Property 14.8.7 (Alternative characterization). In fact, up to an additive constant, the condition

$$(f^*)' = (f')^{-1} \tag{14.53}$$

uniquely determines the Legendre transformation.

Remark 14.8.8. The above definitions can easily be extended to higher dimensions ($n \geq 2$).

Chapter 15

Complex Analysis

15.1 Complex algebra

The set of complex numbers $\mathbb{C}$ forms a 2-dimensional vector space over the field of real numbers 20. At the same time, the operations of complex addition and complex multiplication also turn the complex numbers into a field.

Definition 15.1.1 (Complex conjugate). Complex conjugation

$$\overline{\cdot} : a + bi \mapsto a - bi \tag{15.1}$$

is an involution 2.3.17. It is sometimes denoted by z^* instead of $\overline{z}$, but unless this would cause confusion the former notation will be used.

Formula 15.1.2 (Real/imaginary part). A complex number z can also be written as $\mathrm{Re}(z) + i\,\mathrm{Im}(z)$, where

$$\mathrm{Re}(z) := \frac{z + \overline{z}}{2} \tag{15.2}$$

$$\mathrm{Im}(z) := \frac{z - \overline{z}}{2i}. \tag{15.3}$$

Definition 15.1.3 (Argument). Let z be a complex number expressed in *polar form*: $z = re^{i\theta}$. The number θ is called the argument of z and it is denoted by $\arg(z)$.

Definition 15.1.4 (Riemann sphere). Consider the one-point compactification $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ (Definition 7.5.27). 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:

$$\begin{aligned} z + \infty &:= \infty \\ z * \infty &:= \infty \\ \frac{z}{\infty} &:= 0. \end{aligned} \tag{15.4}$$

Since there exists no multiplicative inverse for ∞, the Riemann sphere is not a field.

15.2 Holomorphic functions

Definition 15.2.1 (Holomorphic function). A function f is said to be holomorphic on an open set $U \subseteq \mathbb{C}$ if it is complex differentiable at every point $z_0 \in U$, i.e. for every point $z_0 \in U$ the following limit exists:

$$f'(z_0) := \lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0}. \tag{15.5}$$

Definition 15.2.2 (Biholomorphic function). A holomorphic function f for which f^{-1} is also holomorphic.

Definition 15.2.3 (Entire). A function that is holomorphic at every point $z \in \mathbb{C}$.

Property 15.2.4 (Cauchy-Riemann conditions). A holomorphic function f satisfies the following conditions:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad \text{and} \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}. \tag{15.6}$$

These conditions can be combined into one equation using the so-called **Wirtinger derivative**:

$$\frac{\partial f}{\partial \overline{z}} = 0. \tag{15.7}$$

Theorem 15.2.5 (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 15.2.6 (Laplace equation). The functions u, v satisfying the CR conditions are harmonic functions, i.e. they satisfy the Laplace equation.

Property 15.2.7 (Level sets). The functions u, v satisfying the CR conditions have orthogonal level curves 2.3.9.

Property 15.2.8 (Real functions). Consider a real-valued function f defined on the complex plane. If it is holomorphic, the CR conditions imply that f is a constant.

Theorem 15.2.9 (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.*

15.3 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 15.3.1 (Contour integral). The contour integral of a function $f(z) = u(z)+iv(z)$ is defined as the following line integral:

$$\int_{z_1}^{z_2} f(z)dz = \int_{(x_1,y_1)}^{(x_2,y_2)} \big[u(x,y) + iv(x,y)\big](dx + idy). \tag{15.8}$$

Theorem 15.3.2 (Cauchy's integral theorem[2]). *Let Ω be a simply-connected subset of $\mathbb{C}$ and let f be a holomorphic function on Ω. For every closed, rectifiable (i.e. of finite length) contour C in Ω:*

$$\oint_C f(z)dz = 0. \tag{15.9}$$

Corollary 15.3.3 (Freedom of contour). The contour integral of a holomorphic function depends only on the limits of integration and not on the contour connecting them.

[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**.

Formula 15.3.4 (Cauchy's integral formula). Let Ω be a connected subset of $\mathbb{C}$ and let f be a holomorphic function on Ω. Consider a contour C in Ω. For every point z_0 inside C one can express the function f as follows:

$$f(z_0) = \frac{1}{2\pi i}\oint_C \frac{f(z)}{z - z_0}dz. \tag{15.10}$$

Corollary 15.3.5 (Analytic function). Let Ω be a connected subset of $\mathbb{C}$ and let C be a closed contour in Ω. If f is holomorphic on Ω, then f is analytic 14.6.7 on Ω and

$$f^{(n)}(z_0) = \frac{1}{2\pi i}\oint_C f(z)\frac{n!}{(z - z_0)^{n+1}}dz. \tag{15.11}$$

Furthermore, the derivatives are also holomorphic on Ω.

Theorem 15.3.6 (Morera). *If f is continuous on a connected open set Ω and $\oint_C f(z)dz = 0$ for every closed contour C in Ω, then f is holomorphic on Ω.*

Theorem 15.3.7 (Liouville). *Every bounded entire function is constant.*

Theorem 15.3.8 (Sokhotski-Plemelj). *Let f be a continuous complex-valued function defined on the real line and let $a < 0 < b$, then*

$$\lim_{\varepsilon\to 0^+}\int_a^b \frac{f(x)}{x \pm i\varepsilon}dx = \mp i\pi f(0) + \mathcal{P}\int_a^b \frac{f(x)}{x}dx, \tag{15.12}$$

where $\mathcal{P}$ denotes the Cauchy principal value.

15.4 Laurent series

Definition 15.4.1 (Laurent series). If f is a analytic function defined on an **annulus**, i.e. a ring-shaped region, then f can be expanded as the following series:

$$f(z) = \sum_{n=-\infty}^{\infty} a_n(z - z_0)^n \qquad \text{with} \qquad a_n = \frac{1}{2\pi i}\oint \frac{f(z')}{(z' - z_0)^{n+1}}dz'. \tag{15.13}$$

The subseries containing all negative degree terms is called the **principal part** of the Laurent series.

Property 15.4.2 (Convergence of Lauren series). The Laurent series of an analytic function f 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*.

Definition 15.4.3 (Analytic continuation). Consider an analytic function f 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 15.4.4 (Schwarz's reflection principle). *Let f be analytic on the upper half plane. If $z \in \mathbb{R} \implies f(z) \in \mathbb{R}$, then*

$$f(\overline{z}) = \overline{f(z)}. \tag{15.14}$$

15.5 Singularities

15.5.1 Poles

Definition 15.5.1 (Pole). A function f has a pole of order $m > 0$ at a point z_0 if its Laurent series at z_0 satisfies $\forall n < -m : a_n = 0$ and $a_{-m} \neq 0$.

Definition 15.5.2 (Meromorphic). A function f 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 15.5.3 (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 15.5.4 (Frobenius transformation). To study the behaviour of a function f at $z \to \infty$, 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 15.5.5 (Polynomials). An entire function f is polynomial if and only if it has a pole at ∞.

Theorem 15.5.6 (Casorati-Weierstrass). *Let f 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 15.5.7. If f is an entire nonpolynomial 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 15.5.8 (Picard's little theorem). *The range of a nonconstant entire function is the complex plane with at most a single exception.*

Theorem 15.5.9 (Picard's great theorem). *Let f 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.*

15.5.2 Branch cuts

Formula 15.5.10 (Roots). Let $z \in \mathbb{C}$. The n^{th} roots[4] of $z = re^{i\theta}$ are given by

$$\left\{ \sqrt[n]{r}\exp\left(i\frac{\theta + 2\pi k}{n}\right) \,\middle|\, k \in \{0, 1, \ldots, n\} \right\}. \tag{15.15}$$

Formula 15.5.11 (Complex logarithm). The natural logarithm can be continued to the complex plane (as a multi-valued function) as follows:

$$\mathrm{LN}(z) := \{\ln(r) + i(\theta + 2\pi k) \mid k \in \mathbb{Z}\}. \tag{15.16}$$

Definition 15.5.12 (Branch). The problem with the previous two formulas is that they represent multi-valued 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 $\arg(\mathrm{LN}) \in \,]\alpha, \alpha + 2\pi]$ is often denoted by LN_α or $\log_\alpha$.

[3] Polynomials are excluded due to the property above.
[4] Also see the fundamental theorem of algebra 11.1.4.

Definition 15.5.13 (Principal value). The principal value of a multi-valued complex function is defined as the value associated with a choice of branch for which $\arg(f) \in]-\pi, \pi]$.

Definition 15.5.14 (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 15.5.15 (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 15.5.16. Consider the complex function

$$f(z) = \frac{1}{\sqrt{(z-z_1)\cdots(z-z_n)}}.$$

This function has singularities at $z_1, \ldots, z_n$. If 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.

15.5.3 Residue theorem

Definition 15.5.17 (Residue). By applying Formula 15.3.1 to a polynomial function, one finds

$$\int_C (z-z_0)^n dz = 2\pi i \delta_{n,-1}, \tag{15.17}$$

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 15.5.18. The residue of a complex function f at a pole z_0 is denoted by

$$\mathrm{Res}[f(z)]_{z=z_0}.$$

Formula 15.5.19. For a pole of order m, the residue is calculated as follows:

$$\mathrm{Res}[f(z)]_{z=z_j} = \lim_{z\to z_0} \frac{1}{(m-1)!}\left(\frac{\partial}{\partial z}\right)^{m-1} (f(z)(z-z_0)). \tag{15.18}$$

For essential singularities the residue can be found by writing out the Laurent series explicitly.

Theorem 15.5.20 (Residue theorem). *If f is a meromorphic function on Ω and if C is a closed contour in Ω that contains the poles z_j of f, then*

$$\oint_C f(z)dz = 2\pi i \sum_j \mathrm{Res}[f(z)]_{z=z_j}. \tag{15.19}$$

For poles on the contour C, only half of the residue contributes to the integral.

Formula 15.5.21 (Argument principle). Let f be a meromorphic function 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:

$$\frac{1}{2\pi i}\oint_C \frac{f(z)}{f'(z)}dz = Z_f(C) - P_f(C). \tag{15.20}$$

Definition 15.5.22 (Winding number). Let f be a meromorphic function 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:

$$\mathrm{Ind}_f(a) := \frac{1}{2\pi i}\oint_C \frac{f'(z)}{f(z)-a}dz. \tag{15.21}$$

This number is always an integer.

15.6 Limit theorems

Theorem 15.6.1 (Small limit theorem). *Let f be a function that is holomorphic almost everywhere on $\mathbb{C}$ 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*

$$\int_C f(z)dz = i\alpha A$$

with

$$A = \lim_{\varepsilon\to 0} f(z).$$

Theorem 15.6.2 (Great limit theorem). *Let f be a function that is holomorphic almost everywhere on $\mathbb{C}$ 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*

$$\int_C f(z)dz = i\alpha B$$

with

$$B = \lim_{R\to\infty} f(z).$$

Theorem 15.6.3 (Jordan's lemma). *Let g 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*

$$\lim_{R\to\infty} f(z) = 0,$$

then

$$\int_C g(z)dz = 0.$$

Chapter 16

Measure Theory and Lebesgue Integration

The main references for this chapter are [8, 9].

16.1 Measure theory

16.1.1 General definitions

Definition 16.1.1 (Measure). Let X be a set and let Σ be a σ-algebra over X. 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**: $\forall i \neq j : E_i \cap E_j = \emptyset \implies \mu(\bigcup_{n=1}^{\infty} E_n) = \sum_{i=n}^{\infty} \mu(E_n)$.

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 16.1.2. 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 2.5.10.

Definition 16.1.3 (Measure space). The pair (X, Σ) is called a measurable space and the triple (X, Σ, μ) is called a measure space. The elements $E \in \Sigma$ are called **measurable sets**.

Definition 16.1.4 (Null set). A set $A \subset \mathbb{R}$ is said to be null if $\mu(A) = 0$.

Definition 16.1.5 (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:

$$\mu\big(\{x \in X \,\big|\, \neg P(x)\}\big) = 0, \tag{16.1}$$

i.e. it holds everywhere except for a null set.

Definition 16.1.6 (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 16.1.7 (Completion). Let $\mathcal{F} \subseteq \mathcal{G}$ be σ-algebras over a set X. $(X, \mathcal{G}, \overline{\mu})$ is called the completion of $(X, \mathcal{F}, \mu)$ if:

[1] In probability theory this is often called **almost surely**.

1. $\forall A \in \mathcal{F} : \overline{\mu}(A) = \mu(A)$,
2. $(X, \mathcal{G}, \overline{\mu})$ is complete, and
3. $\mathcal{G}$ is the smallest σ-algebra for which the foregoing conditions hold.

Definition 16.1.8 (Borel measure). Consider a topological space together with its Borel σ-algebra 7.1.10. Any measure defined on this measurable space is called a Borel 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 16.1.9 (Regular measure). If μ is an outer measure on a topological space X whose completion contains the Borel sets, it 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**: $\mu(B) = \inf\left\{\mu(O) \,\middle|\, O \text{ open}, O \supset B\right\}$, and
2. **Inner regularity**: $\mu(B) = \sup\left\{\mu(K) \,\middle|\, K \text{ closed and measurable}, K \subset B\right\}$.

If X is locally compact Hausdorff, the measure is said to be **regular** if it satisfies the following conditions for every measurable set B:

1. **Outer regularity**: $\mu(B) = \inf\left\{\mu(O) \,\middle|\, O \text{ open}, O \supset B\right\}$, and
2. **Inner regularity**: $\mu(B) = \sup\left\{\mu(K) \,\middle|\, K \text{ compact and measurable}, K \subset B\right\}$.

Note that for locally compact spaces, inner regularity uses compact subsets. Authors that use the second definition of regularity often call these latter measures **tight**. 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 16.1.10 (Radon measure). A Borel measure on a Hausdorff space that is outer regular, inner regular on open sets and **locally finite**, i.e. every point has a neighbourhood of finite measure. When restricted to locally-compact Hausdorff spaces, this is equivalent to requiring that every compact subset has finite measure.

Property 16.1.11. Let X be either a locally compact, separable metric space or a Polish space. Every finite Borel measure on X is tight. Slightly weaker, every finite Borel measure on a metric space is regular.

Definition 16.1.12 (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 $\forall n \in \mathbb{N} : \mu(A_n) < \infty$.

16.1.2 Lebesgue measure

Definition 16.1.13 (Lebesgue outer measure). Let $X \subseteq \mathbb{R}$ be a set. The (Lebesgue) outer measure of X is defined as follows:

$$\lambda^*(X) := \inf\left\{\sum_{n=1}^\infty l(I_n) \,\middle|\, (I_n)_{n\in\mathbb{N}} \text{ a sequence of open intervals that covers } X\right\}. \tag{16.2}$$

Property 16.1.14 (Intervals). The outer measure of an interval I equals its length: $\lambda^*(I) = l(I)$.

Property 16.1.15 (Translation). The outer measure is translation-invariant: $\lambda^*(A+t) = \lambda^*(A)$ for all $A \subset \mathbb{R}$ and $t \in \mathbb{R}$.

Property 16.1.16. The Lebesgue outer measure is an outer measure in the sense of Definition 16.1.1.

Theorem 16.1.17 (Carathéodory's criterion). *Let X be a subset of $\mathbb{R}$. If X satisfies the following equation, it is said to be* ***Lebesgue measurable****:*

$$\forall A \subseteq \mathbb{R} : \lambda^*(A) = \lambda^*(A \cap X) + \lambda^*(A \cap X^c). \tag{16.3}$$

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.*

The construction of Lebesgue measurable sets from the Lebesgue outer measure can be generalized to arbitrary sets:

Construction 16.1.18 (Carathéodory's extension theorem[2]). Every outer measure μ^* gives rise to a σ-algebra consisting of those sets that satisfy Carathéodory's criterion 16.1.17 with respect to μ^*. Furthermore, consider a **premeasure** μ_0, i.e. a σ-additive function defined on an algebra of sets 2.5.1 such that $\mu_0(\emptyset) = 0$. Definition 16.1.13 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.

Corollary 16.1.19. 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 16.1.20. Any countable set is null with respect to the Lebesgue outer measure.

Property 16.1.21 (Regularity). The Lebesgue measure is a regular Borel measure. For every $A \subset \mathbb{R}$ there exists a sequence $(O_n)_{n\in\mathbb{N}}$ of open sets such that

$$A \subset \bigcap_{n=1}^{\infty} O_n \qquad \text{and} \qquad \lambda\left(\bigcap_{n=1}^{\infty} O_n\right) = \lambda^*(A), \tag{16.4}$$

and for every $E \in \mathcal{M}$ there exists a sequence $(F_n)_{n\in\mathbb{N}}$ of closed sets such that

$$\bigcup_{n=1}^{\infty} F_n \subset E \qquad \text{and} \qquad \lambda\left(\bigcup_{n=1}^{\infty} F_n\right) = \lambda(E). \tag{16.5}$$

Property 16.1.22. 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 16.1.23. Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of sets in $\mathcal{M}$. The following two properties apply:

$$\forall i \in \mathbb{N} : A_i \subseteq A_{i+1} \implies \lambda\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \lambda(A_n) \tag{16.6}$$

$$\forall i \in \mathbb{N} : A_i \supseteq A_{i+1} \wedge \lambda(A_1) < \infty \implies \lambda\left(\bigcap_{i=n}^{\infty} A_n\right) = \lim_{n\to\infty} \lambda(A_n). \tag{16.7}$$

[2] Also called the **Hahn-Kolmogorov theorem**.

Remark 16.1.24. This property is valid for every σ-additive set function.

Construction 16.1.25 (Restriction). Let $A \in \mathcal{M}$ have nonzero measure. The restriction of the Lebesgue measure to the set B is defined as follows:

$$\mathcal{M}_A := \{A \cap B \mid B \in \mathcal{M}\} \qquad \text{and} \qquad \forall E \in \mathcal{M}_A : \lambda_A(E) := \lambda(E). \tag{16.8}$$

It can be shown that the measure space $(A, \mathcal{M}_A, \lambda_A)$ is complete.

16.1.3 Measurable functions

Definition 16.1.26 (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.

Definition 16.1.27 (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:

Example 16.1.28 (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 16.1.29 (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 16.1.30. The inclusion $\mathcal{B} \subset \mathcal{M}$ implies that every Borel-measurable function is also Lebesgue-measurable.

Property 16.1.31. The class of Borel/Lebesgue-measurable functions defined on $E \in \mathcal{M}$ forms an algebra.

Example 16.1.32. The following types of functions are Lebesgue-measurable:

- monotonic functions,
- continuous functions,
- indicator functions, and
- compositions of measurable functions.

Corollary 16.1.33. Let f, g be (Lebesgue-)measurable functions and let $F : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a continuous function. The composition $F(f(x), g(x))$ is also measurable.

Property 16.1.34. Let f be a Lebesgue-measurable function. The level set $\{x \mid f(x) = a\}$ is measurable for all $a \in \mathbb{R}$.

Property 16.1.35. Define the following functions (which are measurable if f is measurable as a result of the previous properties):

$$f^+(x) := \max(f, 0) = \begin{cases} f(x) & f(x) > 0 \\ 0 & f(x) \leq 0, \end{cases} \tag{16.9}$$

$$f^-(x) := \max(-f, 0) = \begin{cases} 0 & f(x) > 0 \\ -f(x) & f(x) \leq 0. \end{cases} \tag{16.10}$$

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 16.1.36 (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:

$$f_*\mu(A) := \mu\big(f^{-1}(A)\big). \tag{16.11}$$

Definition 16.1.37 (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

$$T_*\mu = \mu. \tag{16.12}$$

These functions form the morphisms in the category **Meas** of measure spaces.[3]

Definition 16.1.38 (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:

$$T(A) = A \implies \mu(A) = 0 \vee \mu(X \backslash A) = 0. \tag{16.13}$$

This is equivalent to stating that for every set $A \in \Sigma$ with positive measure the following condition holds:

$$\mu\left(\bigcup_{n=1}^{\infty} T^{-n}(A)\right) = 1. \tag{16.14}$$

Property 16.1.39. 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 16.1.40 (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:

$$\lim_{n\to\infty} \mu\big(T^{-n}(A) \cap B\big) = \mu(A)\mu(B) \tag{16.15}$$

Property 16.1.41. All mixing transformations are ergodic.

Property 16.1.42 (Additivity). Every measurable, additive function $f : \mathbb{R} \to \mathbb{R}$ is linear.

Corollary 16.1.43. 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 $\lambda \in \mathbb{R}$.

16.1.4 Limit operations

Property 16.1.44. 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 $\max\limits_{i\leq k}(f_i)$

[3]The notation **Meas** is also sometimes used to denote the larger category of measurable spaces and measurable functions.

- $\inf_{n\in\mathbb{N}}(f_n)$ and $\sup_{n\in\mathbb{N}}(f_n)$
- $\liminf_{n\to\infty}(f_n)$ and $\limsup_{n\to\infty}(f_n)$

Property 16.1.45. If f is a measurable function and g is a function such that $f = g$ almost everywhere, then g is measurable as well.

Corollary 16.1.46. 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 16.1.47 (Essential supremum).

$$\operatorname{ess\,sup}(f) := \inf\{z \in \mathbb{R} \mid f \leq z \text{ a.e.}\} \tag{16.16}$$

Definition 16.1.48 (Essential infimum).

$$\operatorname{ess\,inf}(f) := \sup\{z \in \mathbb{R} \mid f \geq z \text{ a.e.}\} \tag{16.17}$$

Property 16.1.49. 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.
- $\operatorname{ess\,sup}(f) \leq \sup(f)$ and $\operatorname{ess\,inf}(f) \geq \inf(f)$.

The latter pair of inequalities becomes a pair of equalities if f is continuous.

Property 16.1.50. 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 16.1.51 (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 canonical 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.

16.2 Lebesgue integral

16.2.1 Simple functions

Definition 16.2.1 (Indicator function).

$$\mathbb{1}_A(x) := \begin{cases} 1 & x \in A \\ 0 & x \notin A. \end{cases} \tag{16.18}$$

Definition 16.2.2 (Simple function). A function $f : X \to \mathbb{R}$ on a measurable space (X, Σ) that can be expressed as

$$f(x) = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}(x) \tag{16.19}$$

for some $\{a_i \geq 0\}_{i\leq n}, \{A_i\}_{i\leq n} \subset \Sigma$ and $n \in \mathbb{N}$.

Definition 16.2.3 (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 16.2.4 (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 $A \in \Sigma$ with respect to μ is given by

$$\int_A \varphi\, d\mu := \sum_{i=1}^n a_i \mu(A \cap A_i). \tag{16.20}$$

As usual, if the domain of integration is not mentioned explicitly, an integral over the whole space X is implied.

Example 16.2.5. 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:

$$\int_\mathbb{R} \mathbb{1}_\mathbb{Q}\, d\lambda = 1 \times \lambda(\mathbb{Q}) + 0 \times \lambda(\mathbb{R}\backslash\mathbb{Q}) = 0, \tag{16.21}$$

where the measure of the rational numbers is 0 because it is a countable set (Corollary 16.1.20).

16.2.2 Measurable functions

Definition 16.2.6 (Integral for nonnegative functions). The definition for simple functions can be generalized to nonnegative measurable functions f as follows:

$$\int_A f\, d\mu := \sup\left\{\int_A \varphi\, d\mu \,\middle|\, \varphi \text{ a simple function such that } \varphi \leq f\right\}. \tag{16.22}$$

This integral is always nonnegative.

Formula 16.2.7. The following equality allows to change the domain of integrals:

$$\int_A f\, d\mu = \int_X f \mathbb{1}_A\, d\mu. \tag{16.23}$$

Property 16.2.8. The Lebesgue integral over a null set is 0.

Theorem 16.2.9 (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 16.2.10 (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.

16.2.3 Integrable functions

Definition 16.2.11 (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

$$\int_A f\, d\mu := \int_A f^+\, d\mu - \int_A f^-\, d\mu. \tag{16.24}$$

If only one of the functions f^+, f^- is finite, f is said to be **quasi-integrable**.

Property 16.2.12 (Absolute integrability). f is integrable if and only if $|f|$ is integrable. Furthermore, $\int_A |f|\, d\mu = \int_A f^+\, d\mu + \int_A f^-\, d\mu$.

Property 16.2.13. 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 16.2.14 (Integrable functions). The set of integrable functions over a set $A \in \mathcal{M}$ forms the vector space $\mathcal{L}^1(A)$.

Property 16.2.15 (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 16.2.16 (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 16.2.17. All continuous functions are locally integrable.

Property 16.2.18 (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 16.5 for a generalization to arbitrary measures.

16.2.4 Convergence theorems

Theorem 16.2.19 (Fatou's lemma). *Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of nonnegative measurable functions.*

$$\int_A \left(\liminf_{n\to\infty} f_n\right) d\mu \leq \liminf_{n\to\infty} \int_A f_n\, d\mu \tag{16.25}$$

Theorem 16.2.20 (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.*

$$\int_A f\, d\mu = \lim_{n\to\infty} \int_A f_n\, d\mu. \tag{16.26}$$

Method 16.2.21. 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 16.2.22 (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 $f_n\to f$ pointwise a.e., then f is integrable over A and*

$$\int_A f\,d\mu = \lim_{n\to\infty}\int_A f_n\,d\mu. \tag{16.27}$$

Property 16.2.23. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of nonnegative measurable functions

$$\int_A \sum_{n=1}^{\infty} f_n\,d\mu = \sum_{n=1}^{\infty}\int_A f_n\,d\mu. \tag{16.28}$$

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 16.2.24 (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

$$\int_A \sum_{i=1}^{\infty} f_n\,d\mu = \sum_{i=1}^{\infty}\int_A f_n\,d\mu. \tag{16.29}$$

Theorem 16.2.25 (Riemann-Lebesgue lemma). *Let $f\in\mathcal{L}^1(\mathbb{R})$. The sequences*

$$s_k = \int_{\mathbb{R}} f(x)\sin(kx)\,dx$$

and

$$c_k = \int_{\mathbb{R}} f(x)\cos(kx)\,dx$$

both converge to 0.

Theorem 16.2.26 (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:*

$$\lim_{n\to\infty}\frac{1}{n+1}\sum_{t=0}^{n} f(T^n(x)) = \int_X f\,d\mu. \tag{16.30}$$

16.2.5 Relation to the Riemann integral

Property 16.2.27. 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 16.2.28. If $f\geq 0$ and the improper Riemann integral 14.7.1 exists, the Lebesgue integral $\int_{\mathbb{R}} f\,d\mu$ exists and the two integrals coincide. Note that positivity of f is required here. Because the Lebesgue integral is absolute 16.2.12, positive and negative parts cannot cancel (Lebesgue integrals can never be conditionally convergent).

[4]Various other theorems and variants of this theorem can be found in the literature under the same name.

The following definition should be compared to 16.2.1 and 17.1.20.

Definition 16.2.29 (Dirac measure). Define the Dirac measure as follows:

$$\delta_a(A) := \begin{cases} 1 & a \in A \\ 0 & a \notin A. \end{cases} \tag{16.31}$$

Integration with respect to the Dirac measure has the following important property:

$$\int_X f\,d\delta_a = f(a). \tag{16.32}$$

16.3 Space of integrable functions

16.3.1 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 one quotients out these degenerate functions:

Definition 16.3.1 (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 16.3.2. L^1 is a Banach space 23.1.7. The norm on L^1 is given by

$$\|f\|_1 := \int_X |f|\,d\mu. \tag{16.33}$$

In particular

$$\int_X |f|\,d\mu = 0 \implies f = 0 \text{ a.e.} \tag{16.34}$$

16.3.2 Hilbert space L^2

Property 16.3.3. L^2 is a Hilbert space 23.2.4. The norm on L^2 is given by

$$\|f\|_2 := \left(\int_X |f|^2\,d\mu\right)^{\frac{1}{2}}. \tag{16.35}$$

This norm is induced by the following inner product:

$$\langle f|g\rangle := \int_X \overline{f}g\,d\mu. \tag{16.36}$$

Formula 16.3.4 (Cauchy-Schwarz inequality). Let $f, g \in L^2(X, \mathbb{C})$. Formula 16.3.7 implies that $fg \in L^1(X, \mathbb{C})$ and

$$\left|\int \overline{f}g\,d\mu\right| \leq \|fg\|_1 \leq \|f\|_2\|g\|_2. \tag{16.37}$$

16.3.3 L^p-spaces

Generalizing the previous two function classes leads to the notion of L^p-spaces with the following norm:

Formula 16.3.5. For all $1 \leq p \leq \infty$, $L^p(X)$ is a Banach space when equipped with the following norm:

$$\|f\|_p := \left(\int_X |f|^p \, d\mu\right)^{\frac{1}{p}}. \tag{16.38}$$

Remark 16.3.6. Note that L^2 is the only L^p-space that is also a *Hilbert space* (Definition 23.2.4). The other L^p-spaces do not have a norm induced by an inner product.

Formula 16.3.7 (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

$$\|fg\|_1 \leq \|f\|_p \|g\|_q. \tag{16.39}$$

This also implies that $fg \in L^1$.

Formula 16.3.8 (Minkowski's inequality). For every $p \geq 1$ and $f, g \in L^p$ one has that

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p. \tag{16.40}$$

This also implies that $f + g \in L^p$.

Property 16.3.9 (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 16.3.10. 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 16.3.11 (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 16.3.12. A norm on L^∞ is given by

$$\|f\|_\infty := \operatorname{ess\,sup} |f|. \tag{16.41}$$

This norm is called the **supremum norm** and it induces the supremum metric 10.3.3.

Property 16.3.13. Equipped with the above norm the space L^∞ becomes a Banach space.

16.4 Product measures

16.4.1 Construction

The general condition for product measures is given by the following equation that should hold for all $A_1 \in \Sigma_1$ and $A_2 \in \Sigma_2$:

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2). \tag{16.42}$$

Definition 16.4.1 (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 2.5.8:

Property 16.4.2. 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 16.4.3. The function $A_{x_2} \mapsto \mu(A_{x_2})$ is a step function:

$$\mu(A_{x_2}) = \begin{cases} \mu_1(A_1) & x_2 \in A_2 \\ 0 & x_2 \notin A_2. \end{cases}$$

Formula 16.4.4 (Product measure). From the previous property it follows that the product measure $\mu(A)$ can be written in the following way:

$$\mu(A) = \int_{X_2} \mu_1(A_{x_2})\, d\mu_2(x_2). \tag{16.43}$$

Property 16.4.5. 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 and

$$\int_{X_2} \mu_1(A_{x_2})\, d\mu_2(x_2) = \int_{X_1} \mu_2(A_{x_1})\, d\mu_1(x_1). \tag{16.44}$$

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.

16.4.2 Fubini's theorem

Property 16.4.6. 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 16.4.7 (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 16.4.8 (Tonelli). *Let $f : X_1 \times X_2 \to \mathbb{R}$ be a nonnegative function. The following equalities hold:*

$$\begin{aligned} \int_{X_1 \times X_2} f\, d\mu &= \int_{X_1} \left(\int_{X_2} f(x_1, x_2) d\mu_2(x_2) \right) d\mu_1(x_1) \\ &= \int_{X_2} \left(\int_{X_1} f(x_1, x_2) d\mu_1(x_1) \right) d\mu_2(x_2). \end{aligned} \tag{16.45}$$

Corollary 16.4.9 (Fubini). Let $f \in L^1(X_1 \times X_2)$. The sections of f are integrable in the appropriate spaces. Furthermore, the functions $x_1 \mapsto \int_{X_2} f(x_1, x_2)\, d\mu_2(x_2)$ and $x_2 \mapsto \int_{X_1} f(x_1, x_2)\, d\mu_1(x_1)$ are in $L^1(X_1)$ and $L^1(X_2)$ respectively and Tonelli's theorem holds.

Remark 16.4.10. 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.

16.5 Radon-Nikodym theorem

Definition 16.5.1 (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

$$\forall A \in \Sigma : \mu(A) = 0 \implies \nu(A) = 0. \tag{16.46}$$

This relation is often denoted by $\nu \ll \mu$.

The following property relates the notion of absolute continuity above with that of Definition 14.3.5:

Property 16.5.2 (Absolute continuity). Let μ, ν be finite measures on a measurable space (X, Σ). Then $\nu \ll \mu$ if and only if

$$\forall \varepsilon > 0 : \exists \delta > 0 : \forall A \in \Sigma : \mu(A) < \delta \implies \nu(A) < \varepsilon. \tag{16.47}$$

Definition 16.5.3 (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 16.5.4 (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 16.5.5 (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 16.5.6 (Radon-Nikodym theorem for dominated measures). *Let μ be a finite measure on a measurable space (X, Σ) and let ν be a measure dominated by μ. There exists a nonnegative, measurable function f such that $\nu(A) = \int_A f\,d\mu$ for all $A \in \Sigma$.*

Definition 16.5.7 (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 $\frac{d\nu}{d\mu}$.

Theorem 16.5.8 (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 $\nu(A) = \int_A f\,d\mu$ for all $A \in \Sigma$.*

Remark 16.5.9. The function f in this theorem is unique up to a μ-null (and thus ν-null) set.

Property 16.5.10 (Integrability). In general the Radon-Nikodym derivative is not integrable (unless the measures are finite). However, it is always locally integrable 16.2.16. Together with Property 16.2.18 this implies that (densities of) absolutely continuous measures are in bijection with locally integrable functions.

Property 16.5.11 (Change of variables). Let μ, ν be finite measures such that $\nu \ll \mu$ and let $\frac{d\nu}{d\mu}$ be the associated Radon-Nikodym derivative. For every ν-integrable function f the following equality holds

$$\int_A f\,d\nu = \int_A f \frac{d\nu}{d\mu}\,d\mu \tag{16.48}$$

for all $A \in \Sigma$.

Property 16.5.12. Let λ, ν and μ be σ-finite measures. If $\lambda \ll \mu$ and $\nu \ll \mu$, the following two properties hold:

- **Linearity**: $\dfrac{d(\lambda + \nu)}{d\mu} = \dfrac{d\lambda}{d\mu} + \dfrac{d\lambda}{d\mu}$.
- **Chain rule**: If $\lambda \ll \nu$, then $\dfrac{d\lambda}{d\mu} = \dfrac{d\lambda}{d\nu}\dfrac{d\nu}{d\mu}$ a.e.

16.6 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. To this end consider a function F that is right-continuous, i.e. $F(x^+) = F(x)$, and monotonically increasing. The length of an interval can be generalized in the following way:

Definition 16.6.1 (F-length). The F-length of an interval $]a, b]$ is defined as follows:

$$l_F\big(]a, b]\big) := F(b) - F(a). \tag{16.49}$$

The restriction to half-open intervals assures that this function is additive when taking unions of intervals. The footnote in Definition 7.1.10 also assures that the σ-algebra generated by these intervals is the Borel σ-algebra on $\mathbb{R}$.

An immediate extension of Definition 16.1.13 gives the outer measure associated to F:

Definition 16.6.2 (F-outer measure). Let $X \subseteq \mathbb{R}$ be a set. The F-outer measure of X is defined as follows:

$$\mu_F^*(X) := \inf\left\{\sum_{n=1}^{\infty} l_F(I_n)\ \middle|\ (I_n)_{n\in\mathbb{N}} \text{ a sequence of half-open intervals that cover } X\right\}. \tag{16.50}$$

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 16.6.3 (Singletons). The F-measure of a singleton $\{x\}$ is equal to the jump of F at x:

$$\mu_F\big(\{x\}\big) = F(x) - F(x^-). \tag{16.51}$$

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 16.6.4. It follows that the Lebesgue-Stieltjes measures having null singletons are exactly those for which F is continuous.

Property 16.6.5 (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 function

$$F(x) = \mu\big(]-\infty, x]\big). \tag{16.52}$$

Example 16.6.6 (Lebesgue measure). The Lebesgue measure is the Lebesgue-Stieltjes measure associated to $F(x) = x$.

Example 16.6.7 (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 16.6.8. If the Lebesgue-Stieltjes measure μ_F is nonatomic, then

$$F_*\mu_F = \lambda, \tag{16.53}$$

for λ the Lebesgue measure.

Property 16.6.9. 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

$$T_*\mu = \nu. \tag{16.54}$$

Triangular means that $Tx \equiv \big(f_1(x_1), f_2(x_1, x_2), \ldots, f_n(x_1, \ldots, x_n)\big)$ 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 16.6.10. 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 dsitribution* of μ:

$$F_\mu(x) := \mu\big(]-\infty, x]\big), \tag{16.55}$$

and let G_ν be the *quantile function* of ν:

$$G_\nu(x) := \inf\{t \in \mathbb{R} \mid F_\nu(t) \geq x\}. \tag{16.56}$$

The required transformation is then given by the composition:

$$T = G_\nu \circ F_\mu. \tag{16.57}$$

16.7 Signed measures

Definition 16.7.1 (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** : $\forall i \neq j : E_i \cap E_j = \emptyset \implies \mu(\bigcup_{n=1}^\infty E_n) = \sum_{i=n}^\infty \mu(E_n)$.

Note that these requirements are the same as for an ordinary measure 16.1.1, 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 16.7.2. An important consequence of this generalization is that signed measures are not necessarily monotonic, i.e. $A \subseteq B \not\Rightarrow \mu(A) \leq \mu(B)$. In fact this is a strict relation. A signed measure is monotonic if and only if it is a genuine measure.

Definition 16.7.3 (Total variation). Consider a signed measure μ on a measurable space (X, Σ). The total variation $|\mu|$ is the measure defined as follows:

$$|\mu|(A) := \sup\left\{\sum_{P \in \mathcal{P}} |\mu(P)| \;\middle|\; \mathcal{P} \subset \Sigma, \mathcal{P} \text{ covers } A\right\}. \tag{16.58}$$

Using this measure one can decompose the signed measure μ as a difference of two genuine measures:

$$\begin{aligned}\mu &= \mu^+ - \mu^- \\ &= \frac{1}{2}(|\mu| + \mu) + \frac{1}{2}(|\mu| - \mu).\end{aligned} \tag{16.59}$$

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 $\mu^- \leq \lambda_2$.

[5] In fact one can relax this to merely being nonatomic.

The following theorem generalizes both the Radon-Nikodym and Lebesgue decomposition theorems to the case of signed measures:

Theorem 16.7.4. *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 $A \in \Sigma$:*

$$\mu(A) = \int_A f\, d\nu + \mu_s(A). \tag{16.60}$$

As before, the function f is called the Radon-Nikodym derivative of μ.

Theorem 16.7.5 (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*

$$\mu^+(B) = \mu(A \cap B) \qquad\qquad \mu^-(B) = \mu(A^c \cap B). \tag{16.61}$$

Definition 16.7.6 (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:

$$\int_A f\, d\mu := \int_A f\, d\mu^+ - \int_A f\, d\mu^-. \tag{16.62}$$

Definition 16.7.7 (Lebesgue-Stieltjes signed measure). Let F be a function of bounded variation. According to Property 14.3.8 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 16.7.8 (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 $\frac{d\mu_F}{d\lambda} = F'$ λ-a.e. Furthermore, for all $x \in [a, b]$ one has*

$$F(x) - F(a) = \mu_F([a, x]) = \int_a^x F'(t)dt. \tag{16.63}$$

Corollary 16.7.9. If F is absolutely continuous and $F' = 0$ λ-a.e., then F is constant.

Chapter 17

Distributions

The main references for this chapter are [9--11]. Although this chapter is technically part of functional analysis and, hence, uses the language of normed spaces (Chapter 23), it is presented in the part on calculus due to its strong relation to measure and integration theory.

17.1 Functionals

Definition 17.1.1 (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.

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 23.3.9 on $\mathcal{D}_K^m(U) := C_K^m(U)$ is constructed using the following family of seminorms:

$$\mathcal{P} = \left\{ \sup_{x\in K} \|f^{(i)}(x)\| \,\middle|\, |i| \leq m \right\}. \tag{17.1}$$

A topology on all of $\mathcal{D}^m(U)$ is then defined as the inductive limit 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 23.3.10. 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 23.1.3 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 17.1.2 (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:

$$p_{K,m}(f) := \sup_{|i|\leq m} \sup_{x\in K} \|f^{(i)}(x)\| \tag{17.2}$$

$$\sup_{x\in K} \sum_{|i|\leq m} \|f^{(i)}(x)\| \tag{17.3}$$

$$\sum_{|i|\leq m} \sup_{x\in K} \|f^{(i)}(x)\|. \tag{17.4}$$

Property 17.1.3. 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

$$|\langle \phi, f\rangle| \leq C_K\, p_{K,m_K}(f) \tag{17.5}$$

for all $f \in \mathcal{D}_K(U)$.

Definition 17.1.4 (Order). The order of a distribution ϕ is the smallest integer m such that

$$|\langle \phi, f\rangle| \leq C_K\, p_{K,m}(f) \tag{17.6}$$

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 17.1.5. A distribution is of order k if and only if it can be (uniquely) extended to a continuous linear functional on $\mathcal{D}^k(U)$.

Theorem 17.1.6 (Riesz-Markov-Kakutani). *The space of positive continuous functionals on $C_c(X)$, the space of continuous functions with compact support on a locally compact Hausdorff space X, is homeomorphic to the space of Radon measures 16.1.10 on X. Every functional Λ can be represented as*

$$\Lambda(f) = \int_X f\, d\mu \tag{17.7}$$

for some Radon measure μ.

*The topological dual of $C_0(X) \equiv C(\widehat{X})$, the continuous functions on the one-point compactification 7.5.27 (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).*

Example 17.1.7 (Ordinary function as generalized function). By Property 16.2.18, every locally integrable function $f \in L^1_{\text{loc}}$ gives rise to a distribution:

$$\langle f, g\rangle = \int_{\mathbb{R}} f(x)g(x)\, dx. \tag{17.8}$$

Distributions of this form are also said to be **regular**. These distributions are of order 0.

Property 17.1.8. The space $\mathcal{D}$ is dense in $\mathcal{D}'$.

Property 17.1.9 (Product with smooth functions). For every smooth function f and every distribution ϕ, the product $f\phi$ is defined as

$$\langle f\phi, g\rangle := \langle \phi, fg\rangle. \tag{17.9}$$

This turns $\mathcal{D}'$ into a C^∞-module.

17.1.1 Support

Definition 17.1.10 (Support). The support of a distribution is defined as the smallest closed set on which it does not vanish.

Property 17.1.11. 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 17.1.12 (Order). Distributions with compact support have finite order.

Property 17.1.13. A distribution that is supported only at 0 can be written as a linear combination of derivatives of the Dirac measure. More generally, a distribution with support at a finite set of points can be written as a linear combination of (shifted) Dirac measures.

Definition 17.1.14 (Singular support). The complement of the largest open set on which a distribution is regular. The singular support of ϕ is denoted by sing supp(ϕ).

17.1.2 Derivatives

Definition 17.1.15 (Derivative of a distribution). The derivative of a distribution ϕ is defined by duality:

$$\left\langle \frac{\partial \phi}{\partial x}, f \right\rangle := -\left\langle \phi, \frac{\partial f}{\partial x} \right\rangle. \tag{17.10}$$

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 $h \in \mathcal{D}$:

$$\langle f, h' \rangle = -\langle g, h \rangle. \tag{17.11}$$

Property 17.1.16 (Smoothness). Every distribution is smooth, i.e. it is infinitely differentiable. Furthermore, it satisfies the conclusion of Schwarz's theorem 14.6.10.

Property 17.1.17 (Constant distributions). If a distribution T satisfies $T' = 0$, then it is a regular distribution induced by a constant function.

Definition 17.1.18 (Fundamental solution). Let D be a differential operator. A fundamental solution for D is a distribution ϕ such that

$$D\phi = \delta. \tag{17.12}$$

17.1.3 Examples

Definition 17.1.19 (Heaviside distribution). The Heaviside function is defined as follows:[1]

$$H(x) := \begin{cases} 0 & x < 0 \\ 1 & x > 0 \end{cases} \tag{17.13}$$

From this definition it follows that for every $f \in \mathcal{D}(U)$:

$$\langle H, f \rangle = \int_0^\infty f(x)\,dx. \tag{17.14}$$

Definition 17.1.20 (Dirac delta distribution). The Dirac delta distribution is defined as the weak derivative of the Heaviside function:

$$\begin{aligned} \langle \delta, f \rangle &:= \langle H', f \rangle \\ &= -\langle H, f' \rangle \\ &= -\int_0^\infty f'(x)\,dx \\ &= f(0). \end{aligned}$$

[1] The case $x = 0$ is often left undefined, but since this function will always enter formulas inside an integral this does not matter.

Property 17.1.21 (Sampling property). The previous definition can be generalized in the following way (whenever $x_0 \in U$):

$$f(x_0) = \int_U f(x)\delta(x - x_0)\,dx, \tag{17.15}$$

where the suggestive notation[2] $\delta(x - x_0)$ was used to denote the Dirac delta distribution with support at x_0.

Definition 17.1.22 (Dirac comb).

$$\text{III}_b(x) := \sum_{n=-\infty}^{\infty} \delta(x - nb) \tag{17.16}$$

Property 17.1.23 (Transformation). Let $f(x) \in C^1(\mathbb{R})$ be a function with n roots x_1 such that $f'(x_i) \neq 0$. The Dirac delta distribution has the following property:

$$\delta\big(f(x)\big) = \sum_{i=1}^{n} \frac{1}{|f'(x_i)|}\delta(x - x_i). \tag{17.17}$$

Formula 17.1.24 (Differentiation across discontinuities). Let f be a piecewise continuous function with discontinuities at $x_1, \ldots, x_n$ and assume that f induces a distribution by integration. Define the jumps of f at its discontinuities by $\sigma_i := f^+(x_i) - f^-(x_i)$. Next, define the (continuous) function

$$f_c(x) := f(x) - \sum_{i=1}^{n} \sigma_i H(x - x_i).$$

Differentiation of this formula gives

$$f'(x) = f_c'(x) + \sum_{i=1}^{n} \sigma_i \delta(x - x_i).$$

It follows that the derivative in the generalized sense of a piecewise continuous function equals the derivative in the classical sense plus a summation of delta functions at the jump discontinuities.

Example 17.1.25 (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:

$$\left\langle \mathcal{P}\frac{1}{x}, f \right\rangle := \lim_{\varepsilon \downarrow 0} \int_{\varepsilon}^{\infty} \frac{f(x) - f(x^-)}{x}\,dx. \tag{17.18}$$

Moreover, this is the distributional derivative of $\ln|x|$.

17.1.4 Growth rates

Definition 17.1.26 (Schwartz space). The Schwartz space of **rapidly decreasing functions** $\mathscr{S}(\mathbb{R}^n)$ is defined as follows:

$$\mathscr{S}(\mathbb{R}^n) := \left\{ f \in C^\infty(\mathbb{R}^n) \,\middle|\, \forall i, j \in \mathbb{N}^n, \forall x \in \mathbb{R}^n : |x^i f^{(j)}(x)| < \infty \right\}, \tag{17.19}$$

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

$$\sup_{x \in \mathbb{R}^n} (1 + \|x\|)^p |f^{(j)}(x)| \leq M_{p,j}(f). \tag{17.20}$$

[2] See the section on *kernels* further on.

This space has the structure of a Fréchet space under the family of seminorms

$$s_{p,N}(f) := \sup_{x\in\mathbb{R}^n} \sup_{|j|\leq N} (1+\|x\|)^p |f^{(j)}(x)|. \tag{17.21}$$

Remark 17.1.27. 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 17.1.28 (Functions of slow growth). The set of functions of slow growth $N(\mathbb{R}^n)$ is defined as follows:

$$N(\mathbb{R}^n) := \left\{ f \in C^\infty(\mathbb{R}^n) \,\middle|\, \forall i \in \mathbb{N}, \exists M_i > 0 : \left|f^{(i)}(x)\right| = O(\|x\|^i) \text{ for } \|x\| \longrightarrow \infty \right\}. \tag{17.22}$$

Property 17.1.29. If $f \in \mathscr{S}(\mathbb{R})$ and $f \in N(\mathbb{R})$, then $fg \in \mathscr{S}(\mathbb{R})$.

17.2 Convolutions and kernels

Definition 17.2.1 (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:

$$\langle \phi \times \psi, f\rangle := \langle \phi, \langle \psi, f\rangle\rangle \tag{17.23}$$

or

$$\langle \phi \times \psi, f\rangle := \langle \psi, \langle \phi, f\rangle\rangle. \tag{17.24}$$

Definition 17.2.2 (Convolution). The convolution of two distributions is defined as follows (if it exists):

$$\langle \phi * \psi, f\rangle := \langle \phi \times \psi, g\rangle \tag{17.25}$$

where $g(x,y) := f(x+y)$. It should be noted that the convolution is commutative.

Example 17.2.3 (Convolution with delta distribution). For every distribution ϕ one has the following property:

$$\delta * \phi = \phi. \tag{17.26}$$

Formula 17.2.4 (Convolution of functions). The convolution of two (locally integrable) functions $f * g$ on $\mathbb{R}^n$ can be defined through Example 17.1.7:

$$(f * g)(x) := \int_{\mathbb{R}} f(y)g(x-y)\,dy. \tag{17.27}$$

Property 17.2.5 (Young inequality). If $f, g \in L^1$, then $f * g$ exists a.e. and

$$\|f * g\|_1 \leq \|f\|_1 \, \|g\|_1. \tag{17.28}$$

This also implies that $f * g \in L^1$. Furthermore, consider p, q and $r \in \,]0,\infty]$ such that

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{r} + 1. \tag{17.29}$$

If $f \in L^p$ and $g \in L^q$, then

$$\|f * g\|_r \leq \|f\|_p \, \|g\|_q. \tag{17.30}$$

This also implies that $f * g \in L^r$. A result similar to 16.3.7 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 ∞.

Theorem 17.2.6 (Schwarz's kernel theorem). *There exists an isomorphism*

$$D'(U \times V) \to D'(V, D'(U)) \tag{17.31}$$

given by

$$\langle f, (x,y)\rangle = \langle T_f y, x\rangle. \tag{17.32}$$

?? COMPLETE (kernels, ...) ??

17.3 Transformations

17.3.1 Fourier series

Definition 17.3.1 (Dirichlet kernel). The Dirichlet kernel is the collection of functions of the form:

$$D_n(x) := \frac{1}{2\pi}\sum_{k=-n}^{n} e^{ikx}. \tag{17.33}$$

Formula 17.3.2 (Sieve property). If $f \in C^1([-\pi,\pi])$, then

$$\lim_{n\to\infty}\int_{-\pi}^{\pi} f(x)D_n(x)dx = 0. \tag{17.34}$$

Formula 17.3.3 (Generalized Fourier series). Let $f \in L^2([-l,l])$ be a $2l$-periodic function. This function can be approximated by the following series:

$$f(x) = \sum_{n=-\infty}^{\infty}\left(\frac{1}{2l}\int_{-l}^{l} e^{-i\frac{n\pi x'}{l}} f(x')dx'\right)e^{i\frac{n\pi x}{l}}. \tag{17.35}$$

Formula 17.3.4 (Fourier coefficients). As seen in the above formula, the Fourier coefficients can be calculated by taking an inner product (23.8):

$$\widetilde{f}(k) = \int_{-l}^{l} e_k^*(x)f(x)\,dx, \qquad \text{where} \qquad e_k := \sqrt{\frac{1}{2l}}e^{i\frac{k\pi x}{l}}. \tag{17.36}$$

Formula 17.3.5. For 2π-periodic functions, the order-n Fourier approximation is given by the following convolution:

$$s_n(x) = \sum_{k=-n}^{n}\widetilde{f}(k)e^{ikx} = (D_n * f)(x). \tag{17.37}$$

Property 17.3.6 (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

$$(D_n * f)(x) \xrightarrow{n\longrightarrow\infty} \frac{f(x+)+f(x-)}{2}. \tag{17.38}$$

Definition 17.3.7 (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:

$$f^{L,\nu}(x) := \frac{f^L(x+)+f^L(x-)}{2}. \tag{17.39}$$

Property 17.3.8. If f is piecewise C^1 on $[-L,L]$, the Fourier series approximation of f converges to $f^{L,\nu}$ on all of $\mathbb{R}$.

17.3.2 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]

$$\mathcal{F}f(\omega) := \frac{1}{\sqrt{2\pi}} \int \mathbb{R} f(t)e^{-i\omega t}dt. \tag{17.40}$$

The inverse Fourier transform, if it exists, is given by

$$f(t) = \mathcal{F}^{-1}(\mathcal{F}f)(t) = \frac{1}{\sqrt{2\pi}}\,\mathcal{P}\int_{-\infty}^{\infty} \mathcal{F}f(\omega)e^{i\omega t}d\omega. \tag{17.41}$$

Equation (17.40) is called the (forward) Fourier transform of f and Equation (17.41) is called the inverse Fourier transform. The pair $(f, \mathcal{F}f)$ is called a **Fourier transform pair**.

Notation 17.3.9. The Fourier transform of a function f is often denoted by $\widetilde{f}$ or $\widehat{f}$.

Property 17.3.10. From the Riemann-Lebesgue lemma 16.2.25 it follows that

$$\mathcal{F}f(\omega) \longrightarrow 0 \qquad \text{if} \qquad |\omega| \longrightarrow 0. \tag{17.42}$$

Theorem 17.3.11 (Parceval). *Let $(f, \widetilde{f})$ and $(g, \widetilde{g})$ be two Fourier transform pairs.*

$$\int_{-\infty}^{\infty} f(x)g(x)dx = \int_{-\infty}^{\infty} \widetilde{f}(k)\widetilde{g}(k)dk \tag{17.43}$$

Corollary 17.3.12 (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:

$$\int_{-\infty}^{\infty} |f(x)|^2dx = \int_{-\infty}^{\infty} |\widetilde{f}(k)|^2dk. \tag{17.44}$$

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 17.3.13. The Fourier transform defines an isomorphism on $\mathscr{S}$.

Definition 17.3.14 (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 17.3.15. $\mathcal{D}$ is dense in $\mathscr{S}$ and, hence, tempered distributions are determined by their values on $\mathcal{D}$.

Property 17.3.16. The Fourier transform of tempered distributions has some nice additional properties:

[3] All functions are required to be Lebesgue integrable to make the integral converge. Weaker conditions are possible (see the literature).

- The Fourier transform defines an isomorphism on $\mathscr{S}^*$.
- 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.)

The Fourier transform also induces the following isomorphisms:

Theorem 17.3.17 (Paley-Wiener[4]). *The space of compactly supported distributions of order N is isomorphic to the space of entire functions satisfying*

$$|F(z)| \leq C(1+|z|)^N e^{b\,|\mathrm{im}(z)|} \tag{17.45}$$

for some constants $b, C \in \mathbb{R}^+$. The distributions have support lies in $\overline{B}(0,b)$.

The space of regular compactly supported distributions is isomorphic to the space of entire functions satisfying

$$\forall N \in \mathbb{N}, \exists C_N \in \mathbb{R}^+ : |F(z)| \leq C_N(1+|z|)^{-N} e^{b\,|\mathrm{im}(z)|}, \tag{17.46}$$

for some constant $b \in \mathbb{R}^+$ such that the distribution has support on $\overline{B}(0,b)$.

17.3.3 Wave front sets

By the Paley-Wiener-Schwartz theorem, 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 17.3.18 (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 $v \in \mathbb{R}^n$ such that its Fourier transform satisfies the Palais-Wiener estimate

$$|\mathcal{F}(\eta\phi)(w)| \leq C_N(1+\|w\|)^{-N} \tag{17.47}$$

for all vectors $w \in \Gamma$ in a conical neighbourhood (Definition 23.3.8) of v and all **cut-off functions** η, i.e. functions that are equal to 1 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 such that the Fourier distribution (localized around the point) becomes singular in the given direction.

Property 17.3.19. The singular fibre is itself a conical subset.

Definition 17.3.20 (Wave front set). Consider a distribution ϕ on $\mathbb{R}^n$. The wave front set of ϕ is defined as follows:

$$\mathrm{WF}(\phi) := \{(x,v) \in \mathbb{R}^n \times \mathbb{R}^n\backslash 0 \mid v \in \Sigma_x(f)\}. \tag{17.48}$$

The projection of the wave front set onto its first argument recovers the singular support.

Property 17.3.21 (Regularity). A compactly supported distribution is regular if and only if its wave front set is empty.

Property 17.3.22 (Derivatives). The wave front set of the derivative of a distribution is contained in the wave front set of the given distribution.

[4]This version, stated in terms of distributions, is actually due to Schwartz.

17.3.4 Laplace transform

Formula 17.3.23 (Laplace transform).

$$\mathcal{L}\{f\}(s) := \int_0^\infty f(t)e^{-st}dt \tag{17.49}$$

Formula 17.3.24 (Bromwich integral).

$$f(t) = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} \mathcal{L}\{f\}(s)e^{st}ds \tag{17.50}$$

17.3.5 Integral representations

Formula 17.3.25 (Mellin transform).

$$\mathcal{M}\{f(x)\}(s) := \int_0^\infty x^{s-1}f(x)dx \tag{17.51}$$

Formula 17.3.26 (Inverse Mellin transform).

$$f(x) = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} \mathcal{M}\{f(x)\}_{(s)}x^{-s}ds \tag{17.52}$$

Formula 17.3.27 (Heaviside step function).

$$\theta(x) = \frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{e^{ikx}}{k-i\varepsilon}dk \tag{17.53}$$

Formula 17.3.28 (Dirac delta distribution).

$$\delta(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx}dk \tag{17.54}$$

17.4 Analysis on groups ♣

Definition 17.4.1 (Haar measure). A left (resp. right) Haar measure on a topological group is a regular Borel measure 16.1.9 that is finite on compact subsets and invariant under the left (resp. right) group action. For locally compact groups this is a Radon measure 16.1.10.

Example 17.4.2 (Lebesgue measure). Consider $\mathbb{R}^n$ as an additive group. Property 16.1.15 implies that the Lebesgue measure is a left (and right) Haar measure.

Theorem 17.4.3 (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 17.4.4 (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:

$$G^\vee := \mathrm{Hom}(G, S^1). \tag{17.55}$$

In general this group is endowed with the compact-open topology. Elements of this group are called **group characters** of G.

Theorem 17.4.5 (Pontryagin duality). *There exists a natural isomorphism $G \mapsto G^{\vee\vee}$.*

[5] A similar theorem holds for right Haar measures.

Construction 17.4.6 (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 $\chi \in G^\vee$:

$$\widehat{f}(\chi) := \int_G f(g)\overline{\chi(g)}\, d\mu(g), \tag{17.56}$$

where the identification $S^1 \cong \mathrm{U}(1)$ is used.

Theorem 17.4.7 (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*

$$f(g) = \int_{G^\vee} \chi(g)\, d\nu(\chi). \tag{17.57}$$

?? COMPLETE ??

Chapter 18

Ordinary differential equations

18.1 Boundary conditions

Unique solutions of a differential equation are obtained by supplying additional conditions. These are called **boundary conditions**.

Definition 18.1.1 (Periodic boundary condition). A boundary condition of the following form:

$$y(x) = y(x + \varphi). \tag{18.1}$$

Definition 18.1.2 (Dirichlet boundary condition). A boundary condition of the form

$$y(x) = f(x) \tag{18.2}$$

for all $x \in \partial\Omega$, where Ω is the domain on which the problem is defined.

Definition 18.1.3 (Neumann boundary condition). A boundary condition of the form

$$\frac{\partial y}{\partial \hat{n}}(x) = f(x) \tag{18.3}$$

for all $x \in \partial\Omega$, where Ω is the domain on which the problem is defined.

?? CHECK (aren't some of these meant for PDEs) ??

18.2 Existence and uniqueness

Theorem 18.2.1 (Picard-Lindelöf). *Consider an ordinary differential equation of the form*

$$\dot{x}(t) = f(t, x(t)), \tag{18.4}$$

where f is defined on a subset $I \times U \subset \mathbb{R} \times \mathbb{R}^n$.[1] *If f is continuous on I and locally Lipschitz on U, then for every point $(t_0, x_0) \in I \times U$ there exists a maximal interval $J \supseteq I$ and a unique solution $x : J \to \mathbb{R}^n$ of the differential equation with initial condition (t_0, x_0).*

[1] Generalizations to arbitrary Banach spaces exist, see e.g. [9].

18.3 First-order ODEs

Definition 18.3.1 (First-order ODE).

$$y'(t) + a(t)y(t) = R(t) \tag{18.5}$$

If the function R is identically zero, the ODE is said to be **homogenous**.

Formula 18.3.2. Let $U \subseteq \mathbb{R}$ be an open set and let the functions $a, R : U \to \mathbb{R}$ be continuous. The solutions $\varphi : U \to \mathbb{R}$ of 18.3.1 are given by:

$$\varphi(t) = e^{-\int a(t)dt}\left(c + \int R(t)e^{\int a(t)dt}dt\right), \tag{18.6}$$

where c is a constant (in general determined by some kind of boundary condition).

18.4 Second-order ODEs

Definition 18.4.1 (Second-order ODE).

$$y''(t) + a(t)y'(t) + b(t)y(t) = R(t) \tag{18.7}$$

If the function R is identically zero, the ODE is said to be **homogenous**.

18.4.1 General solution

Formula 18.4.2. Let $\varphi : U \to \mathbb{R}$ be a nowhere zero solution of the homogeneous equation. The general solution of 18.4.1 is given by

$$y(t) = c_1\varphi + c_2\varphi \int \frac{e^{-\int a}}{\varphi^2} + \psi_0, \tag{18.8}$$

where ψ_0 is a particular solution of 18.4.1.

Property 18.4.3. Let ψ_0 be a solution of 18.4.1. The set of all solutions is given by the affine space

$$\big\{\psi_0 + \chi \mid \chi \text{ is a solution of the homogeneous equation}\big\}. \tag{18.9}$$

Property 18.4.4. Two solutions of the homogeneous equation are independent if the **Wronskian** is nonzero:

$$W(\varphi_1(x), \varphi_2(x)) := \det\begin{pmatrix} \varphi_1(x) & \varphi_2(x) \\ \varphi_1'(x) & \varphi_2'(x) \end{pmatrix} \neq 0. \tag{18.10}$$

Formula 18.4.5 (Abel's identity). An explicit formula for the Wronskian is given by

$$W(x) = W(x_0)\exp\left(-\int_{x_0}^{x} a(x')dx'\right). \tag{18.11}$$

18.4.2 Constant coefficients

Property 18.4.6. A function $\varphi : U \to \mathbb{C}$ is a complex solution of the homogeneous equation if and only if $\mathrm{Re}(\varphi)$ and $\mathrm{Im}(\varphi)$ are real solutions of the homogeneous equation.

Formula 18.4.7 (Characteristic equation). When studying an ODE of the form[2]

$$y''(t) + py'(t) + qy(t) = 0, \tag{18.12}$$

where p and q are constants, the characteristic equation is defined as follows:

$$\lambda^2 + p\lambda + q = 0. \tag{18.13}$$

By teh fundamental theorem of algebra 11.1.4, this polynomial equation has two (complex) roots λ_1 and λ_2. From these roots one can derive the solutions of Equation (18.12) using the following rules (c_1 and c_2 are constants):

- $\lambda_1 \neq \lambda_2$ with $\lambda_1, \lambda_2 \in \mathbb{R}$: $y(t) = c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t}$,
- $\lambda_1 = \lambda_2$: $y(t) = c_1 e^{\lambda t} + c_2 t e^{\lambda t}$, and
- $\lambda_1 = \lambda_2^*$ with $\lambda_1 = a + ib$: $y(t) = c_1 e^{at}\cos(bt) + c_2 e^{at}\sin(bt)$.

18.4.3 Method of Frobenius

Method 18.4.8 (Frobenius). To find a solution of the homogeneous equation one can assume a solution of the form

$$y(x) = \sum_{i=0}^{\infty} a_i (x - x_0)^{i+k}, \tag{18.14}$$

where $k \in \mathbb{Z}$ is a constant.

Definition 18.4.9 (Indicial equation). After inserting the ansatz (18.14) into the homogeneous equation and collecting all terms in x^i, an equation of the form $\sum_{i=n}^{\infty} H_i(k)x^i = 0$ is obtained, where $n \in \mathbb{N}$ and $H_i(k)$ is a polynomial 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 it is called the indicial equation.

Property 18.4.10. The indicial equation generally has two roots k_1, k_2.

- $k_1 = k_2$: Only one solution will be found with the method of Frobenius (another one can be found as in the second term of Formula 18.4.2.
- $k_1 - k_2 \in \mathbb{Z}$: 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 18.4.11 (Fuchs). *If $a(x)$ and $b(x)$ are analytic at $x = x_0$, the general solution $y(x)$ can be expressed as a Frobenius series.*

18.5 Sturm-Liouville theory

Definition 18.5.1 (Sturm-Liouville problem). An ODE of the following form, subject to mixed boundary conditions:

$$\frac{d}{dx}\left(p(x)\frac{dy}{dx}\right) + \Big(g(x) + \lambda r(x)\Big)y(x) = 0, \tag{18.15}$$

where

[2]Any other form of homogeneous, second-order ODEs with constant coefficients can be rewritten in this form.

- $p(x), q(x)$ and $r(x)$ are continuous on $[a, b]$,
- $p(x) \in C^1([a, b])$ with $p(x) < 0$ or $p(x) > 0$ on $[a, b]$,
- $r(x) \geq 0$ or $r(x) \leq 0$ on $[a, b]$, and
- $r(x)$ is not identically zero on any subinterval.

The boundary conditions are given by

$$\alpha_1 y(a) + \beta_1 y'(a) = 0$$
$$\alpha_2 y(b) + \beta_2 y'(b) = 0,$$

where at least one of the constants $\alpha_1, \alpha_2, \beta_1$ or β_2 is nonzero.

Formula 18.5.2. The solutions of a Sturm-Liouville problem are of the form

$$y(x) = c_1 u_1(x; \lambda) + c_2 u_2(x; \lambda). \tag{18.16}$$

Only for certain values of λ will these solutions (u_1, u_2) 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:

$$\det\begin{pmatrix} \alpha_1 u_1(a; \lambda) + \beta_1 u_1'(a; \lambda) & \alpha_1 u_2(a; \lambda) + \beta_1 u_2'(a; \lambda) \\ \alpha_1 u_1(b; \lambda) + \beta_1 u_1'(b; \lambda) & \alpha_1 u_2(b; \lambda) + \beta_1 u_2'(b; \lambda) \end{pmatrix} = 0. \tag{18.17}$$

The independent eigenfunctions can be found by substituting the solutions λ of this condition in the ODE (18.15).

Definition 18.5.3 (Self-adjoint form). A Sturm-Liouville problem can be rewritten as follows:[3]

$$\hat{\mathcal{L}} y(x) = \lambda y(x).$$

The operator

$$\hat{\mathcal{L}} = -\frac{1}{r(x)}\left[\frac{d}{dx}\left(p(x)\frac{d}{dx}\right) + g(x)\right] \tag{18.18}$$

is called the self-adjoint form (because $\hat{\mathcal{L}}$ is a self-adjoint operator). Now, consider the following general linear ODE

$$\left(a_2(x)\frac{d^2}{dx^2} + a_1(x)\frac{d}{dx} + a_0(x)\right)y(x) = 0. \tag{18.19}$$

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 18.5.4. The eigenfunctions corresponding to distinct eigenvalues are orthogonal with respect to the weight function $r(x)$. This can be seen as an instance of Property 20.5.16.

Theorem 18.5.5 (Oscillation theorem). *The n^{th} eigenfunction of a Sturm-Liouville problem has $n - 1$ roots.*

[3] This formulation explains the name "eigenvalue" for the quantity λ.

18.6 Bessel functions

A Bessel's differential equation is an ordinary differential equation of the following form:

$$x^2y''(x) + xy'(x) + (x^2 - n^2)y(x) = 0. \tag{18.20}$$

The solutions of this ODE are the Bessel functions of the first and second kind (also called respectively **Bessel** and **Neumann functions**):

$$J_n(x) = \sum_{m=0}^{\infty} \frac{(-1)^m}{m!(m+n)!}\left(\frac{x}{2}\right)^{2m+n}, \tag{18.21}$$

$$N_n(x) = \lim_{\nu\to n} \frac{\cos(\nu\pi)J_n(x) - J_{-n}(x)}{\sin(\nu\pi)}. \tag{18.22}$$

Remark. Solution (18.21) can be found using the Frobenius method.

Property 18.6.1. For $n \notin \mathbb{N}$ the solutions J_n and J_{-n} are independent.

Remark 18.6.2. For $n \notin \mathbb{N}$ the limiting operation in (18.22) is not necessary because $\sin(n\pi)$ will never become 0 in this case.

Formula 18.6.3 (Generating function). Consider the following function:

$$g(x,t) := \exp\left[\frac{x}{2}\left(t - \frac{1}{t}\right)\right]. \tag{18.23}$$

If this function is expanded as a Laurent series, an expression of the form

$$g(x,t) = \sum_{n=-\infty}^{\infty} J_n(x)t^n \tag{18.24}$$

is obtained. By applying the residue theorem 15.5.20, one can express the functions J_n as follows:

$$J_n(x) = \frac{1}{2\pi i}\oint_C \frac{g(x,t)}{t^{n+1}}dt. \tag{18.25}$$

One can show that these functions are exactly the Bessel functions (18.21). Therefore, $g(x,t)$ is called the generating function of the Bessel functions.

?? SHOW THAT THESE ARE REALLY THE BESSEL FUNCTIONS ??

18.7 Applications

18.7.1 Laplace equation

When solving the Laplace equation in cylindrical coordinates, one obtains Bessel's ODE with integer n, which has the cylindrical Bessel functions (18.21) and (18.22) as solutions.

18.7.2 Helmholtz equation

When solving the Helmholtz equation in spherical coordinates, one obtains a variant of Bessel's ODE for the radial part:

$$x^2y''(x) + 2xy'(x) + \big(x^2 - n(n+1)\big)y(x) = 0, \tag{18.26}$$

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):

$$j_n(r) = \sqrt{\frac{\pi}{2x}} J_{n+\frac{1}{2}}(r). \tag{18.27}$$

?? COMPLETE ??

Chapter 19

Partial differential equations

For a rigorous treatment of partial differential equations, the language of distributions is required. For an introduction, see Chapter 17.

19.1 General linear equations

Definition 19.1.1 (Characteristic curve). A curve along which the highest-order partial derivatives is discontinuous. Equivalently, a curve along which the partial differential equation reduces to an ordinary differential equation.

19.2 First order PDE

Formula 19.2.1 (First-order quasilinear PDE).

$$P(x,y,u)\frac{\partial u}{\partial x} + Q(x,y,u)\frac{\partial u}{\partial y} = R(x,y,u) \tag{19.1}$$

Formula 19.2.2 (Characteristic curve). Nonunique highest-order partial derivatives means that the following *Pfaffian* system has no unique solution:

$$\begin{cases} P\dfrac{\partial u}{\partial x} + Q\dfrac{\partial u}{\partial y} = R \\ \dfrac{\partial u}{\partial x}dx + \dfrac{\partial u}{\partial y}dy = du. \end{cases} \tag{19.2}$$

Algebraically this gives the following condition:

$$\det\begin{pmatrix} P & Q \\ dx & dy \end{pmatrix} = 0. \tag{19.3}$$

By Cramer's rule 20.4.15 the existence of a (nonunique) solution also requires that

$$\det\begin{pmatrix} P & R \\ dx & du \end{pmatrix} = 0. \tag{19.4}$$

The characteristic curves are thus defined by

$$\frac{dx}{P} = \frac{dy}{Q} \tag{19.5}$$

and along these curves the condition

$$\frac{dx}{P} = \frac{du}{R} \tag{19.6}$$

should hold as a consistency condition.

Formula 19.2.3 (Lagrange-Charpit equations). The general solution of (19.1) 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:

$$\frac{dx}{P} = \frac{dy}{Q} = \frac{du}{R}, \tag{19.7}$$

where c_1, c_2 are constants fixed by the boundary conditions.

Remark 19.2.4. Looking at the defining equations of the characteristic curve, it is clear that these fix the general solution of the PDE.

19.3 Method of the characteristics

Formula 19.3.1 (Second order quasilinear PDE). Consider the following pseudolinear differential equation for a function $u : \mathbb{R}^2 \to \mathbb{R}$:

$$R(x, y)u_{xx} + S(x, y)u_{xy} + T(x, y)u_{yy} = W(x, y, u, p, q), \tag{19.8}$$

where $p := u_x$ and $q := u_y$.

Formula 19.3.2 (Characteristic equation). Similar to the case of first-order PDEs, characteristic curves are characterized by the following condition:

$$\det\begin{pmatrix} R & S & T \\ dx & dy & 0 \\ 0 & dx & dy \end{pmatrix} = 0. \tag{19.9}$$

This is equivalent to the following equation:

$$R\left(\frac{dy}{dx}\right)^2 - S\left(\frac{dy}{dx}\right) + T = 0. \tag{19.10}$$

Accordingly this equation is often called the **characteristic equation** of the PDE 19.3.1. The PDE has been reduced to an ODE (as before, Cramer's rule gives rise to additional ODEs).

Definition 19.3.3 (Types of characteristics). Equation (19.10) is quadratic in $\frac{dy}{dx}$. 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 19.3.4 (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 19.3.1 becomes

$$A(\xi_x, \xi_y)\frac{\partial^2\zeta}{\partial\xi^2} + B(\xi_x, \xi_y, \eta_x, \eta_y)\frac{\partial^2\zeta}{\partial\xi\partial\eta} + A(\eta_x, \eta_y)\frac{\partial^2\zeta}{\partial\eta^2} = F(\xi, \eta, \zeta, \zeta_\xi, \zeta_\eta), \tag{19.11}$$

where

- $A(a,b) = Ra^2 + Sab + Tb^2$
- $B(a,b,c,d) = 2Rac + S(bc+ad) + 2Tbd$.

The discriminant Δ of the quadratic equation (19.10) 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 (19.10) 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:

$$\frac{\partial^2\zeta}{\partial\xi\partial\eta} = H(\xi,\eta,\zeta,\zeta_\xi,\zeta_\eta), \tag{19.12}$$

  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:

$$\frac{\partial^2\zeta}{\partial\eta^2} = G(\xi,\eta,\zeta,\zeta_\xi,\zeta_\eta), \tag{19.13}$$

  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:

$$\frac{\partial^2\zeta}{\partial\alpha^2} + \frac{\partial^2\zeta}{\partial\beta^2} = K(\alpha,\beta,\zeta,\zeta_\alpha,\zeta_\beta). \tag{19.14}$$

Theorem 19.3.5 (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.*

19.4 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.

19.4.1 Cartesian coordinates

Method 19.4.1 (Separation of variables). Let D be the operator associated with a partial differential equation such that $Du(x) = 0$, where $x := (x_1, \ldots, x_n)$ denotes the set of variables. A useful method to find solutions is to assume a solution of the form

$$u(x) = \prod_{i=1}^{n} u_i(x_i). \tag{19.15}$$

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 19.4.2. Consider the following PDE:

$$\frac{\partial u}{\partial t} - a\frac{\partial^2 u}{\partial x^2} = 0. \tag{19.16}$$

Substituting a solution of the form $u(x,t) = X(x)T(t)$ gives

$$X(x)\frac{dT(t)}{dt} - aT(t)\frac{d^2X(x)}{dx^2} = 0,$$

which can be rewritten as (the arguments are dropped for convenience)

$$\frac{1}{aT}\frac{dT}{dt} = \frac{1}{X}\frac{d^2X}{dx^2}.$$

Because both sides are independent they must be equal to a constant λ. This results in the following system of ordinary differential equations:

$$\begin{cases} X''(x) = \lambda X(x) \\ T'(t) = a\lambda T(t). \end{cases} \tag{19.17}$$

19.5 Boundary conditions

Formula 19.5.1 (Inhomogeneous boundary condition).

$$\alpha u(a,t) + \beta\frac{\partial u}{\partial x}(a,t) = h(t) \tag{19.18}$$

When h is identically zero, the boundary condition becomes **homogeneous**.

Method 19.5.2 (Steady-state solution). Assume that the function h is constant. In this case it is useful to rewrite the solution as

$$u(x,t) = v(x) + w(x,t).$$

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 19.5.3. 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)[1]. 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:

Method 19.5.4. 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 fitted 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 i^{th} 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 19.5.5 (Inhomogeneous PDE). A possible way to solve inhomogeneous second order partial differential equations of the form

$$Du(x,t) = f(x,t)$$

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

$$\sum_i w_i(t)e_i(x),$$

where the e_i are a complete set of eigenfunctions in the variable x. This solution should satisfy the (homogeneous[2]) boundary conditions.

[1] 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.

[2] Inhomogeneous boundary conditions can be turned into homogeneous ones by the previous two methods.

2. Expand the function f in the same way as u. The coefficients f_n can be found by 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
$$\sum_i \big(\widetilde{D}w_i(t)\big)e_i(x) = 0,$$
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 18.3.2.
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 $w_i(0) = \Psi_i$.
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 19.5.6. The requirement that all involved functions admit a generalized Fourier expansion is restrictive. Not all inhomogeneous PDEs are solvable by this method.

19.5.1 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 19.3.5 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 19.5.7 (Green's function). A fundamental solution 17.1.18 of a Dirichlet problem.

19.6 General

19.6.1 Symbols

Definition 19.6.1 (Symbol). Consider a general k^{th}-order differential operator (multi-indices α are used)
$$D := \sum_{|\alpha|\leq k} c_\alpha(x) D^\alpha. \tag{19.19}$$
The symbol of this operator is defined by replacing the partial derivatives by indeterminates ξ^i:
$$p(D, \xi) := \sum_{|\alpha|\leq k} c_\alpha(x) \xi^\alpha. \tag{19.20}$$

Definition 19.6.2 (Principal symbol). The principal symbol of a k^{th}-order differential operator D is defined as the highest-degree component of $p(D,\xi)$:

$$\sigma_D(\xi) := \sum_{|\alpha|=k} c_\alpha(x)\xi^\alpha. \tag{19.21}$$

For a system of partial differential equations, the functions c_α are replaced by matrix-valued functions $\left(c_i^j\right)_\alpha$.

Property 19.6.3. The principal symbol of a differential operator transforms as a tensor.

Definition 19.6.4 (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.

19.7 Sobolev spaces

Using the theory of L^p-spaces and distributions (Chapters 16 and 17), one can define an important class of function spaces that are ubiquitous in the field of PDEs (and beyond).

Definition 19.7.1 (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 17.1.15 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:

$$\|f\| := \left(\sum_{|\alpha|\leq m} \|f^{(\alpha)}\|_{L^p}\right)^{1/p}. \tag{19.22}$$

Using the fact that the Fourier transform $\mathcal{F}$ defines an (isometric) isomorphism on L^2, one can also define H^m in a different way:

$$H^m(\mathbb{R}^n) := \left\{f \in L^2(\mathbb{R}^n) \,\middle|\, (1+\|x\|^2)^{-m/2}\mathcal{F}f \in L^2(\mathbb{R}^2)\right\}. \tag{19.23}$$

The Sobolev spaces inherit the following property from the L^p-spaces:

Property 19.7.2 (Completeness). Every Sobolev space is a Banach space. Moreover, one can show that the spaces H^m are Hilbert spaces.

The Sobolev spaces also satisfy the following density theorem:

Property 19.7.3. $\mathcal{D}(\mathbb{R}^n)$ 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 19.7.4 (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 19.7.5 (Friedrich). *For all bounded U one can introduce an equivalent norm on $H_0^m(U)$ as follows:*

$$\langle f|g\rangle := \sum_{|\alpha|=m}\int_U f^{(\alpha)}\overline{g^{(\alpha)}}dx. \tag{19.24}$$

Property 16.3.10 allows to define the dual Sobolev spaces:

Definition 19.7.6. 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 (19.23). It can be shown that all elements in $W^{-m,p}(U)$ are of the form

$$T = \sum_{|\alpha|\leq m} f_\alpha^{(\alpha)}, \tag{19.25}$$

where $f_\alpha \in L^{p'}(U)$ with p' the Hölder conjugate of p.

?? COMPLETE (continue in AMP1) ??

Part IV

Linear Algebra

Chapter 20

Linear Algebra

20.1 Vector spaces

Definition 20.1.1 (K-vector space). Let K be a field. A K-vector space V is a set equipped with two operations, **(vector) addition** $V \times V \to V$ and **scalar multiplication** $K \times V \to V$, 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 $\lambda, \mu \in K$ and $v \in V$.
3. The identity of the field K acts as a neutral element for scalar multiplication: $1_K v = v$ for all $v \in V$.
4. Scalar multiplication is distributive with respect to vector addition: $\lambda(v + w) = \lambda v + \lambda w$ for all $\lambda \in K$ and $v, w \in V$.
5. Vector addition is distributive with respect to scalar multiplication: $(\lambda + \kappa)v = \lambda v + \kappa w$ for all $\lambda, \kappa \in K$ and $v \in V$.

From here on the underlying field K will be left implicit unless the results depend on it.

Remark 20.1.2. The above definition can be restated in abstract algebraic terms. A K-vector space is a module 3.6.29 over K.

20.1.1 Linear independence

Definition 20.1.3 (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

$$w = \sum_{i=1}^{n} \lambda_i v_i \tag{20.1}$$

for some $\{\lambda_i\}_{i\leq n} \subset K$. 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 the remark about *Hamel bases* in next section.)

[1] Generalizations are possible in the context of topological vector spaces (see Chapters 7 and 23), where one can define the notion of convergence.

Definition 20.1.4 (Linear independence). A finite set $\{v_i\}_{i\leq n}$ is said to be linearly independent if the following relation holds:

$$\sum_{i=1}^{n} \lambda_i v_i = 0 \iff \forall i \leq n : \lambda_i = 0. \tag{20.2}$$

A general set $S \subset V$ is said to be linearly independent if every finite subset of it is linearly independent.

Definition 20.1.5 (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 20.1.6 (Frame). A k-frame is an ordered set of k linearly independent vectors.

20.1.2 Bases

Definition 20.1.7 (Basis). A subset $\mathcal{B} \subset V$ that is linearly independent and spans V.

Property 20.1.8. Every spanning set contains a basis.

Remark 20.1.9 (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 20.1.10 (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 2.6.2. Zorn's lemma 2.6.11 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 $(a^1, \ldots, a^n, b)$ and a finite sequence of elements $(e_1, \ldots, e_n)$ in S such that:

$$\sum_{i=0}^{n} a^i e_i + bv = 0, \tag{20.3}$$

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]

$$v = -\frac{1}{b} \sum_{i=0}^{n} a^i e_i. \tag{20.4}$$

Because v was randomly chosen, one can conclude that S is a spanning set for V.

[2]It is this step that requires R to be a division ring in Property 3.6.34 because otherwise one would in general not be able to divide by $b \in R$.

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 20.1.11. 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 20.1.12 (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:

$$\dim(V) := n. \tag{20.5}$$

Definition 20.1.13 (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:

$$W \leq V \iff \forall w_1, w_2 \in W, \forall \lambda \in K : \lambda w_1 + w_2 \in W. \tag{20.6}$$

20.1.3 Sum and direct sum

Definition 20.1.14 (Sum). Let V be a vector space and consider a finite collection of subspaces $\{W_1, \ldots, W_k\}$. The sum of these subspaces is defined as follows:

$$W_1 + \cdots + W_k := \left\{ \sum_{i=1}^k w_i \,\middle|\, w_i \in W_i \right\}. \tag{20.7}$$

For an infinite collection of subspaces the linear combinations have to be finite.

Definition 20.1.15 (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 20.1.16 (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 20.1.17. 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:

$$\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2). \tag{20.8}$$

Property 20.1.18. 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 20.1.19 (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 20.1.20 (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.

20.2 Linear maps

Remark 20.2.1. Linear maps are also called **linear transformations** or **linear mappings**.

20.2.1 Homomorphisms

Definition 20.2.2 (Homomorphism space). Let V, W be two vector spaces. The set of all linear maps between V and W is called the homomorphism space from V to W:

$$\mathrm{Hom}_K(V, W) := \big\{f : V \to W \,\big|\, f \text{ is linear}\big\}. \tag{20.9}$$

The collection of K-vector spaces and linear maps between them form a category $\mathbf{Vect}_K$.

Formula 20.2.3. Let V, W be two finite-dimensional vector spaces.

$$\dim\big(\mathrm{Hom}_K(V, W)\big) = \dim(V)\dim(W) \tag{20.10}$$

Definition 20.2.4 (Endomorphism ring). The space $\mathrm{Hom}_K(V, V)$ with composition of maps as multiplication forms a ring, the endomorphism ring. It is denoted by $\mathrm{End}_K(V)$ or $\mathrm{End}(V)$ when the underlying field is clear.

Property 20.2.5 (Commutator). The endomorphism ring $\mathrm{End}(V)$ can also be endowed with the structure of a Lie algebra (see Property 30.2.23) by equipping it with the commutator

$$[A, B] := A \circ B - B \circ A. \tag{20.11}$$

Property 20.2.6. 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.
- f is bijective.

Definition 20.2.7 (Automorphism). An isomorphism from V to V is called an automorphism. The set of all automorphisms on V is denoted by $\mathrm{Aut}(V)$. It forms a group under composition. Often this group is called the general linear group[3] $\mathrm{GL}_K(V)$ or $\mathrm{GL}(V)$ when the underlying field is clear.

Remark 20.2.8. Sometimes automorphisms are also called **linear operators**. However, this terminology is also used for a general linear map in operator theory (Chapter 24) and so this terminology is not adopted in this text.

Definition 20.2.9 (Kernel). Consider a linear map $f : V \to W$. The kernel of f is defined as the following subspace of V:

$$\ker(f) := \{v \in V \mid f(v) = 0\}. \tag{20.12}$$

Property 20.2.10. A linear map $f : V \to W$ is injective if and only if $\ker(f) = 0$.

Definition 20.2.11 (Rank). The dimension of the image of a linear map.

Definition 20.2.12 (Nullity). The dimension of the kernel of a linear map.

Theorem 20.2.13 (Dimension theorem[4]). *Let $f : V \to W$ be a linear map.*

$$\dim(\mathrm{im}(f)) + \dim(\ker(f)) = \dim(V) \tag{20.13}$$

Corollary 20.2.14. Two finite-dimensional vector spaces are isomorphic if and only if they have the same dimension.

[3] It is isomorphic to the general linear group of invertible matrices 20.4.3 (hence the similar name and notation).
[4] Also called the **rank-nullity theorem**.

Definition 20.2.15 (Minimal polynomial). Let $f \in \mathrm{End}(V)$ with V a finite-dimensional vector space. The monic polynomial μ_f of lowest order such that $\mu_f(f) = 0$ is called the minimal polynomial of f.

Property 20.2.16. Let $f \in \mathrm{End}(V)$ with minimal polynomial μ_f. If $\varphi(f) = 0$ for some polynomial φ, the minimal polynomial μ_f divides φ.

Property 20.2.17 (Jordan-Chevalley decomposition). Every endomorphism A can be decomposed as follows:

$$A = A_{ss} + A_n, \tag{20.14}$$

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.

20.2.2 Dual maps

Definition 20.2.18 (Dual space). Let V be a vector space. The (algebraic) dual V^* of V is defined as the following vector space:

$$V^* := \mathrm{Hom}_K(V, K) = \{f : V \to K \mid f \text{ is linear}\}. \tag{20.15}$$

The elements of V^* are called **linear forms** or (linear) **functionals**.

Property 20.2.19 (Dimension). From Theorem 20.2.3 it follows that $\dim(V^*) = \dim(V)$ whenever V is finite-dimensional. If V is infinite-dimensional, this property is <u>never</u> valid. In the infinite-dimensional case $\mathrm{card}(V^*) > \mathrm{card}(V)$ always holds.

Definition 20.2.20 (Dual basis). Let $\mathcal{B} = \{e_1, e_2, \ldots, e_n\}$ be a basis for a finite-dimensional vector space V. One can construct a basis $\mathcal{B}^* = \{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$ for V^*, called the dual basis of $\mathcal{B}$, as follows:

$$\varepsilon_i : \sum_{j=1}^{n} a_i e_i \mapsto a_i. \tag{20.16}$$

The relation between a basis and its associated dual basis can also be expressed as

$$\varepsilon^i(e_j) = \delta^i_j. \tag{20.17}$$

Definition 20.2.21 (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:

$$\langle v, v^* \rangle := v^*(v). \tag{20.18}$$

(See Definition 27.2.1 for a generalization of this map.)

Definition 20.2.22 (Dual map). Let $f : V \to W$ be a linear map. The linear map

$$f^* : W^* \to V^* : \varphi \to \varphi \circ f \tag{20.19}$$

is called the dual map or **transpose** of f. It is also often denoted by f^T.

20.3 Inner product

In this section all vector spaces V will be defined over $\mathbb{R}$ or $\mathbb{C}$.

20.3.1 Inner product space

Definition 20.3.1 (Inner product). A function $\langle\cdot|\cdot\rangle : V \times V \to \mathbb{C}$ 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**: $\langle v|w\rangle = \langle w|v\rangle^*$,
2. **Linearity in the second argument**: $\langle u|\lambda v + w\rangle = \lambda\langle u|v\rangle + \langle u|w\rangle$,
3. **Nondegeneracy**: $\langle v|v\rangle = 0 \iff v = 0$, and
4. **Positive-definiteness**: $\langle v|v\rangle \geq 0$.

Remark 20.3.2. Inner products are special cases of **nondegenerate Hermitian forms** which do not satisfy the positive-definiteness property.

Corollary 20.3.3. The first two properties have the result of conjugate linearity in the first argument:

$$\langle \lambda f + \mu g|h\rangle = \overline{\lambda}\langle f|h\rangle + \overline{\mu}\langle g|h\rangle \tag{20.20}$$

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 20.3.4 (Inner product space). A vector space equipped with an inner product $\langle\cdot|\cdot\rangle$. This is sometimes called a **pre-Hilbert space**.

Definition 20.3.5 (Metric dual). Using the inner product (or any other nondegenerate Hermitian form) one can define the metric dual of a vector by the following map:

$$L : V \to V^* : v \mapsto \langle v|\cdot\rangle. \tag{20.21}$$

(See Equation (34.1) 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 20.3.6 (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

$$\langle A^\dagger v|w\rangle = \langle v|Aw\rangle \tag{20.22}$$

for all $v, w \in V$. Alternatively, one can define the adjoint using the transpose and metric dual as follows:

$$A^\dagger = L^{-1} \circ A^* \circ L. \tag{20.23}$$

If $A = A^\dagger$, A is said to be **Hermitian** or **self-adjoint**. (In Chapter 23 a distinction will be made between these two notions.)

20.3.2 Orthogonality

Definition 20.3.7 (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:

$$\langle v|w\rangle = 0. \tag{20.24}$$

An **orthogonal system** is a collection of vectors, none of them equal to 0, that are mutually orthogonal.

Property 20.3.8. Orthogonal systems are linearly independent.

Definition 20.3.9 (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:

$$\langle v|v\rangle = 1. \tag{20.25}$$

Definition 20.3.10 (Orthogonal complement). Let W be a subspace of an inner product space V. The orthogonal complement of W is defined as the following subspace:

$$W^\perp := \{v \in V \mid \forall w \in W : \langle v|w\rangle = 0\}. \tag{20.26}$$

Remark. $W^\perp$ is pronounced as "W-perp".

Property 20.3.11 (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 20.3.12. Let $W \leq V$ with V a finite-dimensional inner product vector space. Forming orthogonal complements defines an involution:

$$(W^\perp)^\perp = W. \tag{20.27}$$

Definition 20.3.13 (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 $\{w_1, \ldots, w_k\}$ be an orthonormal basis of W. The projections of $v \in V$ on W and $w \in W$ are defined as follows:

$$\mathrm{proj}_W(v) := \sum_{i=1}^{k} \langle v|w_i\rangle w_i \tag{20.28}$$

$$\mathrm{proj}_w(v) := \frac{\langle v|w\rangle}{\langle w|w\rangle} w. \tag{20.29}$$

Property 20.3.14. Orthogonal projections satisfy the following conditions:

$$\forall w \in W : \mathrm{proj}_W(w) = w \qquad \text{and} \qquad \forall u \in W^\perp : \mathrm{proj}_W(u) = 0. \tag{20.30}$$

Method 20.3.15 (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:

$$\begin{aligned} w_1 &= u_1 \\ w_2 &= u_2 - \frac{\langle u_2|w_1\rangle}{\|u_2\|^2} w_1 \\ &\vdots \\ w_n &= u_n - \sum_{i=1}^{n-1} \frac{\langle u_n|w_i\rangle}{\|u_n\|^2} w_i \end{aligned} \tag{20.31}$$

2. Normalization:

$$
\begin{aligned}
e_1 &= \frac{w_1}{\|w_1\|} \\
e_2 &= \frac{w_2}{\|w_2\|} \\
&\vdots \\
e_n &= \frac{w_n}{\|w_n\|}
\end{aligned}
\tag{20.32}
$$

Definition 20.3.16 (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

$$\sigma_v : V \to V : w \mapsto w - 2\frac{\langle w|v\rangle}{\langle v|v\rangle}v. \tag{20.33}$$

This transformation amounts to a reflection in the hyperplane orthogonal to v.

Definition 20.3.17 (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:

$$\cos\theta := \frac{\langle v|w\rangle}{\|v\|\|w\|}. \tag{20.34}$$

The angle between two vectors v, w is sometimes denoted by $\sphericalangle(v, w)$.

20.4 Matrices

Notation 20.4.1. The vector space of all $m \times n$-matrices defined over the field K is denoted by $M_{m,n}(K)$. If $m = n$, the space is denoted by $M_n(K)$ or $M(n, K)$.

Property 20.4.2 (Dimension). The dimension of $M_{m,n}(K)$ is mn.

Definition 20.4.3 (General linear group). The set of invertible matrices is called the general linear group and is denoted by $\mathrm{GL}_n(K)$ or $\mathrm{GL}(n, K)$.

Property 20.4.4. For all $A \in \mathrm{GL}_n(K)$ one has:

- $A^T \in \mathrm{GL}_n(K)$, and
- $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$.

Definition 20.4.5 (Trace). Let $A \equiv (a_{ij}) \in M_n(K)$. The trace of A is defined as follows:

$$\mathrm{tr}(A) := \sum_{i=1}^{n} a_{ii}. \tag{20.35}$$

Property 20.4.6. Let $A, B \in M_n(K)$. The trace satisfies the following properties:

- $\mathrm{tr} : M_n(K) \to K$ is a linear map,
- $\mathrm{tr}(AB) = \mathrm{tr}(BA)$, and
- $\mathrm{tr}(A^T) = \mathrm{tr}(A)$.

Formula 20.4.7 (Hilbert-Schmidt norm). The Hilbert-Schmidt (or **Frobenius**) norm is defined by the following formula:

$$\|A\|_{HS}^2 := \sum_{i,j} |A_{ij}|^2 = \mathrm{tr}\Big(A^\dagger A\Big). \tag{20.36}$$

If one identifies $M_n(\mathbb{C})$ with $\mathbb{C}^{2n}$, this norm equals the standard Hermitian norm.

Formula 20.4.8 (Hadamard product). The Hadamard product of two matrices is defined as the entry-wise product:

$$(A \circ B)_{ij} := A_{ij}B_{ij}. \tag{20.37}$$

Property 20.4.9. Let $A \in M_{m,n}(K)$. Denote the set of columns as $\{A_1, A_2, \ldots, A_n\}$ and the set of rows as $\{R_1, R_2, \ldots, R_m\}$. The set of columns is a subspace of K^m and the set of rows is a subspace of K^n. Their spans satisfy the following property:

$$\dim(\mathrm{span}(A_1, \ldots, A_n)) = \dim(\mathrm{span}(R_1, \ldots, R_m)). \tag{20.38}$$

Definition 20.4.10 (Rank). Using the invariance relation from the previous property, one can define the rank of a matrix $A \in M_{m,n}(K)$ as follows:

$$\mathrm{rk}(A) := \dim(\mathrm{span}(A_1, \ldots, A_n)) = \dim(\mathrm{span}(R_1, \ldots, R_m)). \tag{20.39}$$

Property 20.4.11. Let $A \in M_{m,n}(K), B \in \mathrm{GL}_n(K), C \in M_{n,r}(K)$ and $D \in M_{r,n}(K)$. The ranks of these matrices satisfy the following properties:

- $\mathrm{rk}(AC) \leq \mathrm{rk}(A)$,
- $\mathrm{rk}(AC) \leq \mathrm{rk}(C)$,
- $\mathrm{rk}(BC) = \mathrm{rk}(C)$, and
- $\mathrm{rk}(DB) = \mathrm{rk}(D)$.

Property 20.4.12. Let $A \in M_{m,n}(K)$. The linear map

$$L_A : K^n \to K^m : v \mapsto Av \tag{20.40}$$

satisfies $\mathrm{im}(L_A) = \mathrm{span}(A_1, \ldots, A_n)$.

20.4.1 System of equations

Property 20.4.13. Let $Ax = b$ with $A \in M_{m,n}(K), x \in K^n$ and $b \in K^m$ be a system of m equations in n variables and let L_A be the linear map as defined in Property (20.4.12). The following properties hold:

- The system is inconsistent if and only if $b \notin \mathrm{im}(L_A)$.
- If the system is not inconsistent, the solution set is an affine space. If $x_0 \in K^n$ is a solution, the solution set is given by: $x_0 + \ker(L_A)$.
- If the system is homogeneous, i.e. $b = 0$, the solution set is equal to $\ker(L_A)$.

Property 20.4.14 (Uniqueness). Let $Ax = b$ with $A \in M_n(K)$ be a system of n equations in n variables. If $\mathrm{rk}(A) = n$, the system has a unique solution.

Formula 20.4.15 (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:

$$x_i = \frac{\det(A_i)}{\det(A)}, \tag{20.41}$$

where A_i is the matrix obtained by replacing the i^{th} column of A by the column vector b.

20.4.2 Coordinates and matrix representations

Definition 20.4.16 (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 20.4.17 (Coordinate isomorphism). With the previous definition in mind one can define the coordinate isomorphism induced by $\mathcal{B}$ as follows:

$$\beta : V \to K^n : \sum_{i=1}^n \lambda_i b_i \mapsto (\lambda_1, \dots, \lambda_n)^T. \tag{20.42}$$

Construction 20.4.18 (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

$$\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix} = A_{f,\mathcal{B},\mathcal{C}} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}. \tag{20.43}$$

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 $A_{f,\mathcal{B},\mathcal{C}} \equiv (a_{ij}) \in M_{n,m}(K)$ is called the matrix representation of f. The j^{th} 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 $M_{m,n}(K)$ is isomorphic to the algebra[5] of linear maps $\mathcal{L}(K^n, K^m)$, thereby explaining why the same notation for the space of invertible matrices 20.4.3 and the space of automorphisms 20.2.7 was used:

Property 20.4.19 (Matrices and linear maps). For every matrix $A \in M_{m,n}(K)$ there exists a linear map $f : K^n \to K^m$ such that $A_{f,\mathcal{B},\mathcal{C}} = A$. Conversely, for every linear map $f : K^m \to K^n$ there exists a matrix $A \in M_{n,m}(K)$ such that $f = L_A$ (given by the previous construction).

Corollary 20.4.20. Let $f \in \mathrm{End}(V)$ 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,

$$(A_f)^{-1} = A_{f^{-1}}.$$

In other words, the linear isomorphism $\mathrm{End}(V) \to M_n(K)$ descends to a group isomorphism

$$\mathrm{GL}_K(V) \to \mathrm{GL}_n(K) : f \mapsto A_f, \tag{20.44}$$

where $n = \dim(V)$.

Formula 20.4.21 (Linear forms). Let $V \cong K^n$ and consider a linear form $f \in V^*$. Equation (20.43) can be rewritten as

$$f\big((\lambda_1, \dots, \lambda_n)^T\big) = (f(e_1), \dots, f(e_n))(\lambda_1, \dots, \lambda_n)^T = \sum_{i=1}^n f(e_i)\lambda_i, \tag{20.45}$$

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:

$$f = \sum_{i=1}^n f(e_i)\varepsilon_i. \tag{20.46}$$

[5]The multiplication is given by the composition of linear maps.

Property 20.4.22 (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 20.4.23. The Hermitian adjoint of a linear map 20.3.6 induces the (Hermitian) adjoint of matrices $A \in \mathbb{C}^{m\times n}$. It is given by

$$A^\dagger = \overline{A}^T, \tag{20.47}$$

where $\overline{A}$ denotes the complex conjugate of A.

20.4.3 Coordinate transformations

Definition 20.4.24 (Transition matrix). Let $\mathcal{B} = \{b_1, \ldots, b_n\}$ and $\mathcal{B}' = \{b'_1, \ldots, b'_n\}$ be two bases of V. By definition, every element of $\mathcal{B}'$ can be written as a linear combination of elements in $\mathcal{B}$:

$$b'_j = q_{1j}b_1 + \cdots + q_{nj}b_n. \tag{20.48}$$

The matrix $Q \equiv (q_{ij}) \in M_n(K)$ is called the transition matrix from $\mathcal{B}$ to $\mathcal{B}'$.

Property 20.4.25. 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:

- $Q \in \mathrm{GL}_n(K)$ 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:

  $$B := (\gamma(b_1), \ldots, \gamma(b_n)) \quad \text{and} \quad B' := (\gamma(b'_1), \ldots, \gamma(b'_n)).$$

  In terms of these matrices one finds that $BQ = B'$.
- Consider $v \in V$. Let $(\lambda_1, \ldots, \lambda_n)^T$ be the coordinate vector with respect to $\mathcal{B}$ and let $(\lambda'_1, \ldots, \lambda'_n)^T$ be the coordinate vector with respect to $\mathcal{B}'$, then

  $$Q\begin{pmatrix}\lambda'_1\\ \vdots\\ \lambda'_n\end{pmatrix} = \begin{pmatrix}\lambda_1\\ \vdots\\ \lambda_n\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}\lambda'_1\\ \vdots\\ \lambda'_n\end{pmatrix} = Q^{-1}\begin{pmatrix}\lambda_1\\ \vdots\\ \lambda_n\end{pmatrix}. \tag{20.49}$$

Corollary 20.4.26 (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:

$$A' = P^{-1}AQ. \tag{20.50}$$

Remark 20.4.27. 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:

$$v' = \sum_{i=1}^n \lambda'_i e'_i = \sum_{i,j,k=1}^n Q^{-1}_{ji}\lambda_j Q_{ik}e_k = \sum_{i,j,k=1}^n \delta_{jk}\lambda_j e_k = v.$$

This remark gives a new way to define a vector $v \in V$:

Alternative Definition 20.4.28 (Vector). Consider an n-dimensional vector space V. One can define an equivalence relation on the set $K^n \times FV$, 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 $A \in \mathrm{GL}_n(K)$ 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 20.4.29 (Matrix conjugation). Let $A \in M_n(K)$. The set

$$\left\{Q^{-1}AQ \,\middle|\, Q \in \mathrm{GL}_n(K)\right\} \tag{20.51}$$

is called the conjugacy class of A in accordance with group theory (Definition 3.2.12). Another term for conjugation is **similarity transformation**.

Remark 20.4.30. If A is a matrix representation of a linear operator f, the conjugacy class of A consists of all matrix representations of f.

Property 20.4.31 (Trace). Property 20.4.6 implies that the trace of a matrix is invariant under conjugation:

$$\mathrm{tr}(Q^{-1}AQ) = \mathrm{tr}(A). \tag{20.52}$$

Definition 20.4.32 (Matrix congruence). Let $A, B \in M_n(K)$. The matrices are said to be congruent if there exists a matrix P such that

$$A = P^TBP. \tag{20.53}$$

Property 20.4.33. Every matrix congruent to a symmetric matrix is also symmetric.

Property 20.4.34 (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* (Definition 20.4.56):

$$Q^TQ = \mathbb{1}_n. \tag{20.54}$$

20.4.4 Determinant

Definition 20.4.35 (Minor). The (i, j)-th minor of A is defined as $\det(A_{ij})$ where $A_{ij} \in M_{n-1}(K)$ is the matrix obtained by removing the i^{th} row and the j^{th} column from A.

Definition 20.4.36 (Cofactor). The cofactor α_{ij} of the matrix element a_{ij} is defined as $(-1)^{i+j}\det(A_{ij})$.

Definition 20.4.37 (Adjugate matrix). The adjugate matrix of $A \in M_n(K)$ is defined as follows:

$$\mathrm{adj}(A) := \begin{pmatrix} \alpha_{11} & \alpha_{21} & \cdots & \alpha_{n1} \\ \alpha_{12} & \alpha_{22} & \cdots & \alpha_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1n} & \alpha_{2n} & \cdots & \alpha_{nn} \end{pmatrix}, \tag{20.55}$$

or in terms of the cofactors: $\mathrm{adj}(A) = (\alpha_{ij})^T$, where the transpose is taken after the elements have been replaced by their cofactor.

Formula 20.4.38 (Laplace). The determinant of a matrix $A \equiv (a_{ij}) \in M_n(K)$ can be evaluated as follows:

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+k} a_{ik} \det(A_{ik}). \tag{20.56}$$

Property 20.4.39. Let $A, B \in M_n(K)$ and denote the columns of A by $A_1, \ldots, A_n$. The determinant has the following properties:

- $\det(AB) = \det(A)\det(B)$,
- $\det(A^T) = \det(A)$,
- $\det(A_1, \ldots, A_i + \lambda A'_i, \ldots, A_n) = \det(A_1, \ldots, A_i, \ldots, A_n) + \lambda \det(A_1, \ldots, A'_i, \ldots, A_n)$ for all $A_i, A'_i \in M_{n,1}(K)$, and
- $\det(A_{\sigma(1)}, \ldots, A_{\sigma(n)}) = \mathrm{sgn}(\sigma)\det(A_1, \ldots, A_n)$

Items 2, 3 and 4 further imply that a matrix with two identical rows or columns has a vanishing determinant.

Property 20.4.40. Let $A \in M_n(K)$. The following statements are equivalent:

- $\det(A) \neq 0$,
- $\mathrm{rk}(A) = n$, or
- $A \in \mathrm{GL}_n(K)$.

Property 20.4.41. For all $A \in M_n(K)$ one finds that $A\,\mathrm{adj}(A) = \mathrm{adj}(A)A = \det(A)I_n$.

Corollary 20.4.42. For all $A \in \mathrm{GL}_n(K)$ one finds

$$A^{-1} = \det(A)^{-1}\,\mathrm{adj}(A). \tag{20.57}$$

Alternative Definition 20.4.43 (Minor). Let $A \in M_{m,n}(K)$ and choose $k \leq \min(m,n)$. A $k \times k$-minor of A is the determinant of a $k \times k$-partial matrix obtained by removing $m-k$ rows and $n-k$ columns from A.

Property 20.4.44. Let $A \in M_{m,n}(K)$ and choose $k \leq \min(m,n)$. Then $\mathrm{rk}(A) \geq k$ if and only if A contains a nonzero $k \times k$-minor.

Property 20.4.45 (Invariance of determinant). Let $f \in \mathrm{End}(V)$. The determinant of the matrix representation of f is invariant under basis transformations.

Definition 20.4.46 (Determinant of a linear map). The previous property allows for an unambiguous definition of the determinant of $f \in \mathrm{End}(V)$:

$$\det(f) := \det(A) \tag{20.58}$$

for any matrix representation A of f.

20.4.5 Characteristic polynomial

Definition 20.4.47 (Characteristic polynomial). Consider a linear map $f \in \mathrm{End}(V)$ and denote its matrix representation by A_f. The function

$$\chi_f(x) := \det(x\mathbb{1}_n - A_f) \in K[x] \tag{20.59}$$

is a monic polynomial of degree n in the variable x. The following equation is called the **characteristic equation** or **secular equation** of f:

$$\chi_f(x) = 0. \tag{20.60}$$

Formula 20.4.48. Consider a matrix $A \equiv (a_{ij}) \in M_n(K)$ 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:

$$\begin{cases} c_0 = (-1)^n \det(A), \\ c_{n-1} = -\operatorname{tr}(A) \end{cases} \tag{20.61}$$

Theorem 20.4.49 (Cayley-Hamilton). *Consider a linear map $f \in \operatorname{End}(V)$ with characteristic polynomial χ_f.*

$$\chi_f(f) = f^n + \sum_{i=1}^{n-1} c_i f^i = 0. \tag{20.62}$$

Corollary 20.4.50. From Property 20.2.16 and the Cayley-Hamilton theorem it follows that the minimal polynomial μ_f is a divisor of the characteristic polynomial χ_f.

20.4.6 Matrix groups

Definition 20.4.51 (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 20.4.52 (Invertibility). Elementary matrices have determinant 1 and, accordingly, are elements of $\mathrm{GL}_n(K)$.

Property 20.4.53. Multiplication by an elementary matrix has the following properties:

- Left multiplication by an elementary matrix $E_{ij}(c)$ comes down to replacing the i^{th} row of the matrix with the i^{th} row plus c times the j^{th} row.
- Right multiplication by an elementary matrix $E_{ij}(c)$ comes down to replacing the j^{th} column of the matrix with the j^{th} column plus c times the i^{th} column.
- Left multiplication by an elementary matrix $T_{i,j}$ interchanges the i^{th} and j^{th} rows.

Property 20.4.54. Every invertible matrix can be written as a product of elementary matrices.

Definition 20.4.55 (Special linear group). The subgroup of $\mathrm{GL}_n(K)$ consisting of all matrices with determinant 1:

$$\mathrm{SL}_n(K) := \{A \in \mathrm{GL}_n(K) \mid \det(A) = 1\}. \tag{20.63}$$

Definition 20.4.56 (Orthogonal group). The orthogonal and special orthogonal group are defined as follows:

$$\begin{aligned}\mathrm{O}(n,K) &:= \{A \in \mathrm{GL}_n(K) \mid AA^T = A^TA = \mathbb{1}_n\}\\ \mathrm{SO}(n,K) &:= \mathrm{O}_n(K) \cap \mathrm{SL}_n(K).\end{aligned}$$

Definition 20.4.57 (Unitary group). Consider a field K equipped with an involution $\sigma : \lambda \mapsto \overline{\lambda}$. The unitary and special unitary group are defined as follows:

$$\begin{aligned}\mathrm{U}_n(K,\sigma) &:= \{A \in \mathrm{GL}_n(K) \mid A\sigma(A)^T = \sigma(A)^TA = \mathbb{1}_n\}\\ \mathrm{SU}_n(K,\sigma) &:= \mathrm{U}_n(K) \cap \mathrm{SL}_n(K).\end{aligned}$$

In practice 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 $K = \mathbb{C}$ the notation is further simplified to $\mathrm{U}(n)$ and $\mathrm{SU}(n)$.

Definition 20.4.58 (Unitary equivalence). Let $A, B \in M_n(K)$ over a field K with an involution. The matrices are said to be unitarily equivalent if there exists a unitary matrix U such that

$$A = U^\dagger BU.$$

Property 20.4.59. For orthogonal matrices, conjugacy 20.4.29 and congruency 20.4.32 coincide. More generally, for unitary matrices conjugacy and unitary equivalence coincide.

Definition 20.4.60 (Symplectic group). Consider a vector space V with an antisymmetric nonsingular matrix Ω. The symplectic group $\mathrm{Sp}(V,\Omega)$ is defined as follows:

$$\mathrm{Sp}(V,\Omega) := \{A \in \mathrm{GL}(V) \mid A^T\Omega A = \Omega\}. \tag{20.64}$$

Over the real or complex numbers one can define the canonical **symplectic** matrix

$$\Omega_{st} := \begin{pmatrix} 0 & -\mathbb{1}\\ \mathbb{1} & 0\end{pmatrix}. \tag{20.65}$$

The groups of matrices that preserve this matrix are denoted by $\mathrm{Sp}(n,\mathbb{R})$ and $\mathrm{Sp}(n,\mathbb{C})$.

Remark 20.4.61. Symplectic groups can only be defined on even-dimensional spaces because antisymmetric matrices can only be nonsingular if the dimension n is even.

Definition 20.4.62 (Compact symplectic group). The compact symplectic group is defined as follows (although the notation is confusing, it is standard):

$$\mathrm{Sp}(n) := \mathrm{Sp}(2n,\mathbb{C}) \cap \mathrm{U}(2n). \tag{20.66}$$

This group is in fact isomorphic to the *quaternionic* unitary group in n quaternionic dimensions.

Property 20.4.63.

$$\mathrm{Sp}(1) \cong \mathrm{SU}(2)$$

20.4.7 Matrix decompositions

Method 20.4.64 (QR Decomposition). Every square complex matrix M can be decomposed as

$$M = QR \tag{20.67}$$

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 $R_{ij} = \langle e_i|\text{col}_j(M)\rangle$, where $\text{col}_j(M)$ denotes the j^{th} column of M.
> - a unitary matrix $Q = (e_1, \ldots, e_n)$ constructed by setting the i^{th} column equal to the i^{th} basis vector e_i.

Property 20.4.65. 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, ...) ??

20.5 Eigenvectors

Definition 20.5.1 (Eigenvector). A vector $v \in V\setminus\{0\}$ is called an **eigenvector** of the linear map $f : V \to V$ if it satisfies

$$f(v) = \lambda v \tag{20.68}$$

for some $\lambda \in K$. The scalar λ is called the **eigenvalue** associated to v.

Definition 20.5.2 (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

$$\ker(\lambda\mathbb{1}_V - f). \tag{20.69}$$

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 23.4.5.)

Property 20.5.3 (Characteristic equation). Consider a linear map $f \in \text{End}(V)$. A scalar $\lambda \in K$ is an eigenvalue of f if and only if it satisfies the characteristic equation (20.60).

Property 20.5.4. A linear map $f \in \text{End}(V)$ defined over an n-dimensional vector space V has at most n different eigenvalues.

These property lead to the following method for finding eigenvectors:

Method 20.5.5 (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 (20.60).
2. Find the eigenvector v_i associated to the eigenvalue λ_i by solving

$$(A - \lambda_i\mathbb{1}_V)v_i = 0. \tag{20.70}$$

20.5.1 Diagonalization

Definition 20.5.6 (Diagonalizable map). Let V be a finite-dimensional vector space. A linear map $f \in \mathrm{End}(V)$ is said to be diagonalizable if it admits a diagonal matrix representation.

Property 20.5.7. Every diagonalizable map is semisimple 20.2.17. Conversely, in finite dimensions (and over an algebraically closed field), a semisimple map is diagonalizable.

Theorem 20.5.8. *A matrix $A \in M_n(K)$ is diagonalizable if and only if there exists a matrix $P \in \mathrm{GL}_n(K)$ such that $P^{-1}AP$ is diagonal.*

Corollary 20.5.9 (Trace). Using the Property 20.4.31 that the trace of a linear map is invariant under similarity transformations, the following useful formula can be proven:

$$\mathrm{tr}(f) = \sum_{i=0}^{n} \lambda_i, \tag{20.71}$$

where $\{\lambda_i\}_{i\leq n}$ are the eigenvalues of f.

Property 20.5.10. Let V be an n-dimensional vector space and let $f \in \mathrm{End}(V)$ 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 20.5.11. *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.*

20.5.2 Multiplicity

Definition 20.5.12 (Multiplicity). Let V be a vector space and let $f \in \mathrm{End}(V)$ have characteristic polynomial

$$\chi_f(x) = \prod_{i=1}^{n} (x - \lambda_i)^{n_i}. \tag{20.72}$$

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 20.5.13 (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 K (i.e. if it has "enough' roots" in K). If not, f cannot even be diagonalized. In general there always exists a field f containing K, called a *splitting field*, over which the polynomial can be completely factorized. Note that in general this field is strictly smaller than the algebraic closure of K, which is the *splitting field* of the collection of all polynomials over K.

Property 20.5.14. The algebraic multiplicity is always greater than or equal to the geometric multiplicity.

Theorem 20.5.15. *A linear map $f \in \mathrm{End}(V)$ is diagonalizable if and only if for every eigenvalue the algebraic multiplicity is equal to the geometric multiplicity.*

Property 20.5.16. Every Hermitian linear map $f \in \mathrm{End}(\mathbb{C}^n)$ 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 20.5.17 (Commutator). Let $f, g \in \mathrm{End}(V)$ be two diagonalizable maps. If the commutator $[f, g]$ is zero, the two maps have a common eigenbasis.

Theorem 20.5.18 (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 20.4.59).*

20.6 Euclidean space

A finite-dimensional $\mathbb{R}$-vector space is sometimes called a **Euclidean** or **Cartesian space**.

Notation 20.6.1. When working in a Euclidean space, the inner product $\langle v|w\rangle$ is often written as $v \cdot w$.

Definition 20.6.2 (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 20.6.3 (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 20.6.4. It is convenient to take the standard basis $(e_1, \ldots, e_n)$ to be positively oriented.

20.7 Algebras

Definition 20.7.1 (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 20.7.2 (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 instead of a vector space.

Definition 20.7.3 (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 $\langle a|b\rangle := \frac{1}{2}[q(a+b) - q(a) - q(b)]$ is a nondegenerate inner product (20.3.1).

Theorem 20.7.4 (Frobenius). *There exist three inequivalent finite-dimensional real associative division algebras:* $\mathbb{R}, \mathbb{C}$ *and* $\mathbb{H}$.

Theorem 20.7.5 (Hurwitz). *There exist four inequivalent finite-dimensional real normed division algebras:* $\mathbb{R}, \mathbb{C}, \mathbb{H}$ *and* $\mathbb{O}$.

Example 20.7.6 (Frobenius algebra). An associative algebra A equipped with a nondegenerate bilinear form $\eta : A \times A \to K$ satisfying the following condition for all $a, b, c \in A$:

$$\eta(ab, c) = \eta(a, bc). \tag{20.73}$$

Equivalently, an associative algebra (A, μ) equipped with a linear form $\varepsilon : A \to K$ such that $\varepsilon \circ \mu$ is nondegenerate.[6]

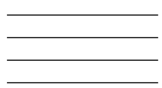
A Frobenius algebra is said to be symmetric if η is symmetric.

Example 20.7.7 (Temperley-Lieb algebra). Let R be a commutative unital ring and fix an element $\delta \in R$. The Temperley-Lieb algebra $\mathrm{TL}_n(\delta)$ is the unital R-algebra with generators $\{U_i\}_{i<n}$ that satisfy the **Jones relations**:

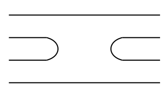
1. $U_i^2 = \delta U_i$,
2. $U_iU_j = U_jU_i$ if $|i - j| \neq 1$, and
3. $U_iU_jU_i = U_i$ if $|i - j| = 1$.

One can represent the elements of a Temperley-Lieb algebra diagrammatically. All elements of $\mathrm{TL}_n(\delta)$ 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 i^{th} input (resp. output) to the $i + 1^{th}$ input (resp. output) and all other inputs are connected to the output directly across the diagram. Multiplication in $\mathrm{TL}_n(\delta)$ is performed diagrammatically by placing two diagrams side by side. Closed loops are replaced by a factor δ.



(a) Unit in $\mathrm{TL}_4(\delta)$.



(b) Generator U_2 in $\mathrm{TL}_4(\delta)$.

Figure 20.1: Temperley-Lieb algebra.

Definition 20.7.8 (Jordan algebra). A nonassociative, commutative algebra A such that

$$(xy)(xx) = x(y(xx)) \tag{20.74}$$

for all $x, y \in A$.

Property 20.7.9 (Power associativity). It can be shown that the Jordan condition implies that powers of elements are well-defined:

$$(xx)x = x(xx) =: x^3 \tag{20.75}$$

for all $x \in A$ and likeiwse for higher-order powers.

[6] A third equivalent definition is given in 26.1.7.

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

$$(x^m y)x^n = x^m(yx^n). \tag{20.76}$$

By commutativity one obtains that the multiplication maps $L_{x^m} : y \mapsto x^m y$ associated to powers commute:

$$L_{x^m} L_{x^n} = L_{x^n} L_{x^m}. \tag{20.77}$$

This leads to the following equivalent definition:

Alternative Definition 20.7.10. A Jordan algebra is a commutative, power-associative algebra A such that Equation (20.77) holds for all $x \in A$.

Property 20.7.11. 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:

$$x \circ y := \frac{1}{2}(xy + yx), \tag{20.78}$$

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**.

20.8 Grassmanians

Definition 20.8.1 (Grassmannian). Let V be a vector space. The set of all subspaces of V of dimension k is called the Grassmannian $\mathrm{Gr}(k, V)$.

Property 20.8.2. $\mathrm{GL}(V)$ acts transitively 3.3.12 on the k-dimensional subspaces of V. Property 3.3.13 implies that the coset space $\mathrm{GL}(V)/H_W$ for the stabilizer H_W of any $W \in \mathrm{Gr}(k, V)$ is isomorphic (as a set) to $\mathrm{Gr}(k, V)$. When V is an n-dimensional real vector space one can show that this quotient is isomorphic to $\mathrm{O}(n)/(\mathrm{O}(k) \times \mathrm{O}(n-k))$. For complex vector spaces the orthogonal groups should be replaced by unitary groups.

Example 20.8.3 (Projective space). Recall Definition 11.2.28. The Grassmannian $\mathrm{Gr}(1, V)$ is given by the projective space $K\mathbb{P}^{\dim(V)-1}$.

Definition 20.8.4 (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 $(\dim(V_1), \ldots, \dim(V_n) = \dim(V))$ 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 20.8.5 (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 (Chapter 29), called the **flag manifold**.

Property 20.8.2 generalizes as follows:

Property 20.8.6 (Parabolic subgroups). Every flag variety has the structure of a homogeneous space: $\mathrm{Fl}_{n,\underline{d}} = \mathrm{GL}(V)/P_{n,\underline{d}}$, 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 21

Vector & Tensor Calculus

References for this chapter are [9, 50]. For a more geometric approach to some of the concepts and results in this chapter, see the content of Chapters 28 and 32 and Section 32.7.

Remark. In this chapter a *vector field* will mean a vector-valued function with smooth projections.

21.1 Nabla-operator

Remark. The geometric approach to this section is summarized in Remark 34.1.9.

Definition 21.1.1 (Gradient). Let $\varphi : \mathbb{R}^3 \to \mathbb{R}$ be a smooth function.

$$\nabla\varphi := \left(\frac{\partial\varphi}{\partial x}, \frac{\partial\varphi}{\partial y}, \frac{\partial\varphi}{\partial z}\right) \tag{21.1}$$

This definition can easily be generalized to arbitrary dimensions.

Property 21.1.2. The gradient of a smooth real-valued function is perpendicular to its level sets 2.3.9.

Formula 21.1.3 (Differential). Let $\varphi : \mathbb{R}^3 \to \mathbb{R}$ be a smooth function. The total differential $d\varphi$ can be rewritten as follows:

$$d\varphi = \nabla\varphi \cdot d\vec{\boldsymbol{r}}. \tag{21.2}$$

Definition 21.1.4 (Directional derivative). Consider a smooth function $\varphi : \mathbb{R}^3 \to \mathbb{R}$ 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 $\hat{a}$:

$$\nabla_{\hat{a}}\varphi := (\hat{a} \cdot \nabla)\varphi. \tag{21.3}$$

Example 21.1.5. Let $\varphi : \mathbb{R}^3 \to \mathbb{R}$ be a smooth real-valued function and let $\frac{d\vec{\boldsymbol{r}}}{ds}$ denote the tangent vector to a curve $\vec{\boldsymbol{r}}(s)$ with natural parameter[1] s. The variation of φ along $\vec{\boldsymbol{r}}(s)$ is given by

$$\frac{\partial\varphi}{\partial s} = \frac{d\vec{\boldsymbol{r}}}{ds} \cdot \nabla\varphi. \tag{21.4}$$

Definition 21.1.6 (Conservative vector field). A vector field that can be expressed as the gradient of a scalar function.

[1]See Definition 28.1.4 for a formal definition.

Definition 21.1.7 (Gradient of a tensor). Let T be a tensor field on $\mathbb{R}^3$ and let $\vec{\boldsymbol{e}}_i$ be the coordinate basis. The gradient of T is defined as follows:

$$\nabla T := \sum_{i=1}^{3} \frac{\partial T}{\partial x^i} \otimes \vec{\boldsymbol{e}}_i. \tag{21.5}$$

Definition 21.1.8 (Divergence). Let $\vec{\boldsymbol{A}}$ be a vector field on $\mathbb{R}^3$.

$$\nabla \cdot \vec{\boldsymbol{A}} := \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \tag{21.6}$$

Definition 21.1.9 (Solenoidal vector field). A vector field $\vec{\boldsymbol{A}}$ that satisfies

$$\nabla \cdot \vec{\boldsymbol{A}} = 0. \tag{21.7}$$

Such a vector field is also said to be **divergence-free** due to Equation (21.14) below.

Definition 21.1.10 (Rotor / curl). Let $\vec{\boldsymbol{A}}$ be a vector field on $\mathbb{R}^3$.

$$\nabla \times \vec{\boldsymbol{A}} := \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}, \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}, \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \right) \tag{21.8}$$

Definition 21.1.11 (Irrotational vector field). A vector field $\vec{\boldsymbol{A}}$ that satisfies

$$\nabla \times \vec{\boldsymbol{A}} = 0. \tag{21.9}$$

Definition 21.1.12 (Laplacian). Let φ and $\vec{\boldsymbol{A}}$ be respectively a smooth function and smooth vector field on $\mathbb{R}^3$.

$$\Delta\varphi := \nabla^2\varphi = \frac{\partial^2\varphi}{\partial x^2} + \frac{\partial^2\varphi}{\partial y^2} + \frac{\partial^2\varphi}{\partial z^2} \tag{21.10}$$

$$\Delta\vec{\boldsymbol{A}} := \nabla^2\vec{\boldsymbol{A}} = \nabla\left(\nabla \cdot \vec{\boldsymbol{A}}\right) - \nabla \times \left(\nabla \times \vec{\boldsymbol{A}}\right) \tag{21.11}$$

$$= (\Delta A_x, \Delta A_y, \Delta A_z) \tag{21.12}$$

The latter is sometimes called the **vector Laplacian**.

Property 21.1.13 (Mixed properties). The differential operators introduced above satisfy the following identities:

$$\nabla \times (\nabla\varphi) = 0 \tag{21.13}$$

$$\nabla \cdot \left(\nabla \times \vec{\boldsymbol{A}}\right) = 0. \tag{21.14}$$

Corollary 21.1.14. All conservative vector fields are irrotational. However, the converse is only true if the domain is simply-connected 8.1.18. (All of this is formalized in the Poincaré lemma 32.8.8.)

Formula 21.1.15 (Helmholtz decomposition). If $\vec{\boldsymbol{A}}$ is a vector field that decays faster than $1/r$ when $r \longrightarrow \infty$, it can be written as

$$\vec{\boldsymbol{A}} = \nabla \times \vec{\boldsymbol{B}} + \nabla\varphi. \tag{21.15}$$

for some smooth vector field $\vec{\boldsymbol{B}}$ and smooth function φ

The differential operators introduced above can also be generalized to curvilinear coordinates. To do this one needs the scale factors as formally defined in Definition 28.2.5. For the remainder of this section the Einstein summation convention will not be used to make everything as explicit as possible.

Formula 21.1.16 (Unit vectors).

$$\frac{\partial \vec{\boldsymbol{r}}}{\partial q^i} = h_i \hat{e}_i \tag{21.16}$$

Formula 21.1.17 (Gradient).

$$\nabla\varphi = \sum_{i=1}^{3} \frac{1}{h_i}\frac{\partial\varphi}{\partial q^i}\hat{e}_i \tag{21.17}$$

Formula 21.1.18 (Divergence).

$$\nabla\cdot\vec{\boldsymbol{A}} = \frac{1}{h_1h_2h_3}\left(\frac{\partial}{\partial q^1}(A_1h_2h_3) + \frac{\partial}{\partial q^2}(A_2h_3h_1) + \frac{\partial}{\partial q^3}(A_3h_1h_2)\right) \tag{21.18}$$

Formula 21.1.19 (Rotor).

$$\left(\nabla\times\vec{\boldsymbol{A}}\right)_i = \sum_{j,k=1}^{3}\frac{\varepsilon_{ijk}}{h_jh_k}\left(\frac{\partial}{\partial q^j}(A_kh_k) - \frac{\partial}{\partial q^k}(A_jh_j)\right), \tag{21.19}$$

where ε_{ijk} is the 3-dimensional Levi-Civita symbol 21.4.8.

Formula 21.1.20 (Laplacian in different coordinate systems). In general the Laplace operator is defined as

$$\Delta f := \nabla\cdot\nabla f. \tag{21.20}$$

The Laplacian can also be expressed in different coordinate systems:

- Cylindrical coordinates (ρ, ϕ, z):

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial}{\partial\rho}\right) + \frac{1}{\rho^2}\frac{\partial^2}{\partial\phi^2} + \frac{\partial^2}{\partial z^2}. \tag{21.21}$$

- Spherical coordinates (r, ϕ, θ):

$$\frac{1}{r^2}\left[\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2} + \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right)\right]. \tag{21.22}$$

21.2 Integration

21.2.1 Line integrals

Formula 21.2.1 (Line integral of a continuous function). Let $f : \mathbb{R}^3 \to \mathbb{R}$ be a continuous function and let Γ be a piecewise smooth curve $\vec{\boldsymbol{\varphi}} : [a, b] \to \mathbb{R}^3$. The line integral of f along Γ is defined as follows:

$$\int_\Gamma f\,ds := \int_a^b f(\vec{\boldsymbol{\varphi}}(t))\,\|\vec{\boldsymbol{\varphi}}'(t)\|dt. \tag{21.23}$$

Formula 21.2.2 (Line integral of a continuous vector field). Let $\vec{F}$ be a continuous vector field on $\mathbb{R}^3$ and let Γ be a piecewise smooth curve with parametrization $\vec{\varphi} : [a, b] \to \mathbb{R}^3$. The line integral of $\vec{F}$ along Γ is defined as follows:

$$\int_\Gamma \vec{F} \cdot d\vec{s} := \int_a^b \vec{F}(\vec{\varphi}(t)) \cdot \vec{\varphi}'(t)dt. \tag{21.24}$$

Property 21.2.3 (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 32.7.23.)

21.2.2 Integral theorems[2]

Theorem 21.2.4 (Fundamental theorem of calculus for line integrals). *Let $\Gamma : \mathbb{R} \to \mathbb{R}^3$ be a piecewise smooth curve defined on the interval $[a, b]$.*

$$\int_\Gamma \nabla f \cdot d\vec{r} = \varphi(\Gamma(b)) - \varphi(\Gamma(a)) \tag{21.25}$$

Theorem 21.2.5 (Kelvin-Stokes theorem). *Let $\vec{A}$ be a vector field defined on $\mathbb{R}^3$ and let S be a smooth surface with boundary ∂S.*

$$\oint_{\partial S} \vec{A} \cdot d\vec{l} = \iint_S \left(\nabla \times \vec{A}\right) dS \tag{21.26}$$

Theorem 21.2.6 (Divergence theorem[3]). *Let $\vec{A}$ be a vector field defined on $\mathbb{R}^3$.*

$$\oiint_{\partial V} \vec{A} \cdot d\vec{S} = \iiint_V \left(\nabla \cdot \vec{A}\right) dV \tag{21.27}$$

Corollary 21.2.7 (Green's identity). Let ϕ, ψ be smooth real-valued functions defined on $\mathbb{R}^3$.

$$\oiint_{\partial V} (\psi\nabla\phi - \phi\nabla\psi) \cdot d\vec{S} = \iiint_V \left(\psi\nabla^2\phi - \phi\nabla^2\psi\right) dV \tag{21.28}$$

21.3 Tensors

21.3.1 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 21.3.1 (Tensor product space). The tensor product of two finite-dimensional vector spaces V and W is defined as[4] 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:

$$(v \otimes w)(g, h) := v(g)w(h). \tag{21.29}$$

In this incarnation the tensor product is sometimes known as the **outer product**. Outer products are also frequently called **pure** or **simple tensors**.

[2] These theorems follow from a more general theorem by Stokes 32.7.23.

[3] Also known as **Gauss's theorem** or the **Gauss-Ostrogradsky theorem**.

[4] "*isomorphic to*" would be better terminology. See the universal property 21.3.3 below.

Definition 21.3.2 (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 $T_{i\ldots j}{}^{k\ldots l} := \mathbf{T}(e_i,\ldots,e_j,e^k,\ldots,e^l)$.

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 21.3.3. 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 $T = f\circ\varphi$, where $\varphi : V\times W\to V\otimes W$ is the bilinear map $(v,w)\mapsto v\otimes w$.

Corollary 21.3.4. The tensor product is unique up to linear isomorphisms. This results in the commutativity of the tensor product:

$$V\otimes W\cong W\otimes V. \tag{21.30}$$

Notation 21.3.5 (Tensor power).

$$V^{\otimes n} := \underbrace{V\otimes\cdots\otimes V}_{n\text{ copies}} \tag{21.31}$$

More generally, the tensor product of r copies of V and s copies of V^* is denoted by

$$\mathcal{T}^r_s(V) = V^{\otimes r}\otimes V^{*\otimes s}. \tag{21.32}$$

Tensors in this space are said to be of **type** (r,s).

Definition 21.3.6 (Scalar). The scalars, i.e. the elements of the underlying field are by definition the $(0,0)$-tensors.

Definition 21.3.7 (Tensor algebra). The tensor algebra over a vector space V is defined as follows:

$$T(V) := \bigoplus_{k\geq 0} V^{\otimes k}. \tag{21.33}$$

The following remark is strongly related to Property 20.2.19:

Remark 21.3.8. For finite-dimensional vector spaces the space $\mathcal{T}^1_1 V$ is isomorphic to $\mathrm{End}(V)$ and the space $\mathcal{T}^1_0 V$ is isomorphic to V itself.

However, when including infinite-dimensional spaces, the space $\mathcal{T}^1_1 V$ is only isomorphic to the endomorphism space $\mathrm{End}(V^*)$ of the dual. This isomorphism is given by the map $\hat{T} : V^*\to V^* : \omega\mapsto \mathbf{T}(-,\omega)$ for every $\mathbf{T}\in\mathcal{T}^1_1 V$. Moreover, in this general setting, the spaces $\mathcal{T}^0_1 V$ and V^* are also isomorphic.

The tensor product space can also be defined as follows:

Alternative Definition 21.3.9 (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 $\mathcal{T}_s^r(V)$ 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 21.3.10 (Dimension). From the previous construction it follows that the dimension of $\mathcal{T}_s^r(V)$ 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 21.3.1 and the pairing of dual vectors (20.17) gives

$$\begin{aligned}\mathbf{T}(\varepsilon^m, \ldots, \varepsilon^n, e_a, \ldots, e_b) &= T_{i\ldots j}{}^{k\ldots l} e_k(\varepsilon^m) \ldots e_l(\varepsilon^n) \varepsilon^i(e_a) \ldots \varepsilon^j(e_b) \\ &= T_{i\ldots j}{}^{k\ldots l} \delta_k^m \ldots \delta_l^n \delta_a^i \ldots \delta_b^j \\ &= T_{a\ldots b}{}^{m\ldots n}.\end{aligned}$$

□

21.3.2 Transformation rules

In this section the behaviour of tensors under basis transformations of the form $e'_i = A^i{}_j\, e_j$ is considered.

Definition 21.3.11 (Contravariant). A tensor component that transforms by the following rule is said to be contravariant:

$$v^i = A^i{}_j\, v'^j. \tag{21.34}$$

Definition 21.3.12 (Covariant). A tensor component that transforms by the following rule is said to be covariant:

$$p'_i = A^j{}_i\, p_j. \tag{21.35}$$

Example 21.3.13 (Mixed tensor). This example gives the transformation rule of a mixed third-order tensor $T \in \mathcal{T}_2^1$:

$$T^k_{ij} = A^k{}_w (A^{-1})^u{}_i (A^{-1})^v{}_j T'^w{}_{uv}. \tag{21.36}$$

Method 21.3.14 (Quotient rule). Assume that an equation such as $Q_i{}^j A_{jl}{}^k = B_{il}{}^k$ is given, with A and B two known tensors. The quotient rule asserts the following: "*If the equation holds under all transformations, then Q is a tensor of the indicated type.*" Note that this rule does not necessarily hold when $B = 0$ because transformation rules are not well-defined for the zero tensor.

Remark. This rule is a useful substitute for the "illegal" division of tensors.

21.3.3 Tensor operations

Definition 21.3.15 (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 20.2.21.

Definition 21.3.16 (Direct product). Let A and B be two tensors. The tensor constructed by the componentwise multiplication of A and B is called the direct product of A and B. This is a generalization of the Hadamard product 20.4.8.

Example 21.3.17. Let $A^i{}_k$ and $B^j{}_{lm}$ represent two tensors. Their direct product is equal to

$$C^{i\ j}_{\ k\ lm} = A^i{}_k B^j{}_{lm}.$$

Formula 21.3.18 (Operator product). It is also possible to combine operators acting on different vector spaces to make them act on the tensor product space:

$$(A \otimes B)(v \otimes w) := Av \otimes Bw. \tag{21.37}$$

Notation 21.3.19 (Abuse of notation). Consider an operator A acting on a vector space V_1. When working with a tensor product space $V_1 \otimes V_2$, the operator A can be extended to the product as $A \otimes \mathbb{1}$. However, it is often still denoted by A.

Notation 21.3.20 (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 sometimes denoted by

$$T_{(ij)} = \frac{1}{2}(T_{ij} + T_{ji}) \tag{21.38}$$

and

$$T_{[ij]} = \frac{1}{2}(T_{ij} - T_{ji}). \tag{21.39}$$

This notation is easily generalized to other types of tensors.

Property 21.3.21 (Gradient of tensor products). The gradient of an outer product is defined through the Leibniz rule:

$$\nabla \cdot (v \otimes w) := (\nabla \cdot v)w + (v \cdot \nabla)w. \tag{21.40}$$

Definition 21.3.22 (Complexification). Let V be a real vector space. The complexification of V is defined as the following tensor product:

$$V^{\mathbb{C}} := V \otimes \mathbb{C}. \tag{21.41}$$

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 as follows for all $\alpha \in \mathbb{C}$:

$$\alpha(v \otimes \beta) := v \otimes (\alpha\beta). \tag{21.42}$$

Property 21.3.23. 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

$$V^{\mathbb{C}} \cong V \oplus iV. \tag{21.43}$$

21.4 Exterior algebra

21.4.1 Antisymmetric tensors

Definition 21.4.1 (Antisymmetric tensor). A tensor that changes sign under the interchange of any two indices.

Notation 21.4.2 (Symmetric tensors). The space of symmetric $(n,0)$-tensors is denoted by $S^n(V)$. The space of symmetric $(0,n)$-tensors is denoted by $S^n(V^*)$.

Notation 21.4.3 (Antisymmetric tensors). The space of antisymmetric $(n,0)$-tensors is denoted by $\Lambda^n(V)$. The space of antisymmetric $(0,n)$-tensors is denoted by $\Lambda^n(V^*)$.

Property 21.4.4. Let $n = \dim(V)$. The space $\Lambda^r(V)$ equals the zero space for all $r \geq n$.

21.4.2 Determinant

Definition 21.4.5 (Form). An n-form is a totally antisymmetric element of $\mathcal{T}^0_n V$.

Definition 21.4.6 (Volume form). A form of rank $\dim(V)$ is also called a **top form** or **volume form**.

Definition 21.4.7 (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:

$$\det(\varphi) := \frac{\omega(\varphi(e_1),\ldots,\varphi(e_n))}{\omega(e_1,\ldots,e_n)}. \tag{21.44}$$

This definition is well-defined, i.e. it is independent of the choice of volume form and basis. Furthermore, it coincides with Definition 20.4.46.

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 principal bundles (see Section 33).

21.4.3 Levi-Civita symbol

Definition 21.4.8 (Levi-Civita symbol). In n dimensions the Levi-Civita symbol is defined as follows:

$$\varepsilon_{i_1\ldots i_n} = \begin{cases} 1 & \text{if } (i_1\ldots i_n) \text{ is an even permutation of } (1\ldots n) \\ -1 & \text{if } (i_1\ldots i_n) \text{ is an odd permutation of } (1\ldots n) \\ 0 & \text{if any of the indices occurs more than once.} \end{cases} \tag{21.45}$$

Remark 21.4.9 (Pseudotensor). The Levi-Civita symbol is not a tensor, it is a *pseudotensor*. This means that the sign changes under reflections or any transformation with determinant -1. (To turn it into a proper tensor, one should multiply it by a factor $\sqrt{g}$, where g is the determinant of the metric.)

Definition 21.4.10 (Cross product). Using the Levi-Civita symbol, one can define the i^{th} component of the cross product as follows:

$$(v\times w)_i = \sum_{j,k=1}^{3} \varepsilon_{ijk} v_j w_k. \tag{21.46}$$

The previous remark implies that the cross product is in fact not a vector, instead it is a "pseudovector".

Remark 21.4.11 (Generalization and 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, one $\mathbb{R}^3$ 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 20.7.5, 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 ordinary 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 21.4.12 (Exceptional Lie group G_2). Using the vector product, one can define an associative 3-form, the **triple product**:

$$u \otimes v \otimes w \mapsto u \cdot (v \times w). \tag{21.47}$$

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.)

21.4.4 Wedge product

Definition 21.4.13 (Antisymmetrization). Let $S_k =$ denote the permutation group 3.2.25 on k elements. The antisymmetrization operator is defined as follows:

$$\mathrm{Alt}(e_1 \otimes \cdots \otimes e_k) := \sum_{\sigma \in S_k} \mathrm{sgn}(\sigma) e_{\sigma(1)} \otimes \cdots \otimes e_{\sigma(k)}. \tag{21.48}$$

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 21.4.16 below should be modified.

Definition 21.4.14 (Wedge product). Let $\{e_i\}_{i \leq \dim(V)}$ be a basis for V. The wedge product of basisvectors is defined as follows:

$$e_1 \wedge \ldots \wedge e_k = \mathrm{Alt}(e_1 \otimes \cdots \otimes e_k) \tag{21.49}$$

From this definition it immediately follows that the wedge product is (totally) antisymmetric.

Construction 21.4.15. 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

$$\{e_{i_1} \wedge \ldots \wedge e_{i_r} \mid \forall k \leq r : 1 \leq i_k \leq \dim(V)\}.$$

Accordingly, the dimension of this space is given by

$$\dim \Lambda^r(V) = \binom{n}{r}. \tag{21.50}$$

For $r = 0$ this construction would be vacuous, so one just defines $\Lambda^0(V) := \mathbb{R}$.

Formula 21.4.16. Let $v \in \Lambda^r(V)$ and $w \in \Lambda^m(V)$. The wedge product 21.4.14 can be generalized as follows:

$$v \wedge w = \frac{1}{r!m!} \operatorname{Alt}(v \otimes w). \tag{21.51}$$

Definition 21.4.17 (Blades). 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 21.4.18 (Cross product). In dimension 3 there exists an important isomorphism $J : \Lambda^2(\mathbb{R}^3) \to \mathbb{R}^3$:

$$J(\lambda)^i = \frac{1}{2}\varepsilon^i{}_{jk}\lambda^{jk}, \tag{21.52}$$

where $\lambda \in \Lambda^2(\mathbb{R}^3)$. (See also the Hodge $*$-operator 21.4.26 further below.)

Looking at the definition of the cross product 21.4.10, 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.

21.4.5 Exterior algebra

Definition 21.4.19 (Exterior power). In the theory of tensor calculus, the space $\Lambda^k(V)$ is often called the k^{th} exterior power of V. As mentioned before, its elements are called (exterior) k-**forms**.

Definition 21.4.20 (Exterior algebra). One can define a graded vector space 27.1.1 as follows:

$$\Lambda^\bullet(V) := \bigoplus_{k \geq 0} \Lambda^k(V). \tag{21.53}$$

This graded vector space can be turned into a graded algebra by taking the wedge product as the multiplication:

$$\wedge : \Lambda^k(V) \times \Lambda^l(V) \to \Lambda^{k+l}(V). \tag{21.54}$$

This algebra is called the exterior algebra or **Grassmann algebra** over V. Elements of the space $\otimes_{k\text{even}}\Lambda^k(V)$ are said to be **Grassmann-even** and elements of $\otimes_{k\text{odd}}\Lambda^k(V)$ are said to be **Grassmann-odd**.

Alternative Definition 21.4.21. Let $T(V)$ be the tensor algebra 21.3.7 over the vector space V, i.e.

$$T(V) = \bigoplus_{k \geq 0} V^{\otimes k}. \tag{21.55}$$

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 of equivalence. Consider the equality

$$(u + v) \otimes (u + v) - u \otimes u - v \otimes v = u \otimes v + v \otimes u \tag{21.56}$$

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 21.4.14 because it imposes the relation $u \wedge v = -v \wedge u$. □

However, one should pay attention to one little detail. As mentioned in 21.4.21 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 21.4.22 (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 27.1.5.

Property 21.4.23 (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 a pure k-form.

21.4.6 Hodge star

Equation (21.50) 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 20.3.2.

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

$$\mathrm{Vol} = e_1 \wedge \ldots \wedge e_n. \tag{21.57}$$

This convention will also be adopted in the remainder of this section.

Definition 21.4.24 (Orientation). Let $\mathrm{Vol}(V)$ be the standard 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 $\mathrm{Vol}(V)$. Denote this number by r. This also implies that a choice of volume form induces an equivalence relation on top-dimensional forms. An equivalence class under this relation is called an orientation on V. If $r > 0$, the orientation is said to be **positive** and, if $r < 0$, the orientation is said to be **negative**.

Formula 21.4.25 (Inner product). Let V be equipped with an inner product $\langle\cdot|\cdot\rangle$. One can extend this to an inner product on $\Lambda^k(V)$ by first defining it on decomposable forms and extending it by linearity to all of $\Lambda^k(V)$:

$$\langle v_1 \wedge \ldots \wedge v_k | w_1 \wedge \ldots \wedge w_k\rangle_k := \det(\langle v_i|w_j\rangle). \tag{21.58}$$

For an orthogonal basis this formula factorizes as follows:

$$\langle v_1 \wedge \ldots \wedge v_k | w_1 \wedge \ldots \wedge w_k\rangle_k = \langle v_1|w_1\rangle \cdots \langle v_k|w_k\rangle. \tag{21.59}$$

Definition 21.4.26 (Hodge star). The Hodge star $* : \Lambda^k(V) \rightarrow \Lambda^{n-k}(V)$ 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:

$$\omega \wedge \rho = \langle *\omega|\rho\rangle_{n-k}\,\mathrm{Vol}(V), \tag{21.60}$$

where $\langle\cdot|\cdot\rangle_{n-k}$ is the inner product (21.58) 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 $\mathrm{Vol}(V)$. This implies that it can be written as

$$c_\omega(\rho)\,\mathrm{Vol}(V).$$

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 23.2.7 can be applied to identify c_ω with a unique element $*\omega \in \Lambda^{n-k}(V)$ such that

$$c_\omega(\rho) = \langle *\omega|\rho\rangle_{n-k}.$$

$\square$

Formula 21.4.27. 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 $\{i_1,\ldots,i_k\}$ and $\{j_1,\ldots,j_{n-k}\}$ be two ordered, complementary index sets and consider an element $\omega = e_{i_1}\wedge\ldots\wedge e_{i_k} \in \Lambda^k(V)$.
>
> $$*\omega = \mathrm{sgn}(\tau)\prod_{m=1}^{n-k}\langle e_{j_m}|e_{j_m}\rangle e_{j_1}\wedge\ldots\wedge e_{j_{n-k}}, \tag{21.61}$$
>
> where τ is the permutation that maps $e_{i_1}\wedge\ldots\wedge e_{i_k}\wedge e_{j_1}\wedge\ldots\wedge e_{j_{n-k}}$ to $\mathrm{Vol}(V)$.

Using this formula one can easily prove the following important property:

Property 21.4.28. Consider an inner product space V. The Hodge dual is involutive up to a factor:

$$**\omega = (-1)^{k(n-k)}\omega. \tag{21.62}$$

Taking the defining relation of the Hodge star operator together with the above property implies the following formula (which is often found in the literature as the defining relation):

Formula 21.4.29. For all $\omega,\rho\in\Lambda^k(V)$ the Hodge star operator satisfies the following formula:

$$\omega\wedge *\rho = \langle\omega|\rho\rangle\,\mathrm{Vol}(V). \tag{21.63}$$

Corollary 21.4.30. Consider three vectors $u,v,w\in\mathbb{R}^3$.

$$*(v\wedge w) = v\times w \tag{21.64}$$
$$*(v\times w) = v\wedge w \tag{21.65}$$
$$*(u\wedge v\wedge w) = u\cdot(v\times w) \tag{21.66}$$

Remark 21.4.31. Formula (21.52) is an explicit evaluation of the first equation (21.64).

Proof. The signs $\mathrm{sgn}(\sigma)$ in the definition of wedge products can be written using the Levi-Civita symbol ε_{ijk} as defined in 21.4.8. The factor $\frac{1}{2}$ is introduced to correct for the double counting due to the contraction over both the indices j and k.

Definition 21.4.32 (Self-dual form). Let V be a 4-dimensional inner product space and consider an element $\omega\in\Lambda^2(V)$. Then ω is said to be self-dual if

$$*\omega = \omega. \tag{21.67}$$

Furthermore, every element $\rho\in\Lambda^2(V)$ can be uniquely decomposed as the sum of a self-dual and an anti-self-dual two-form:

$$\rho = \frac{1}{2}\big((\rho + *\rho) + (\rho - *\rho)\big). \tag{21.68}$$

Chapter 22

Representation Theory

References for this chapter are [50, 62]. Sections 3.2 and 3.3 can be visited for an introduction to groups and group actions.

22.1 Group representations

Group actions on vector spaces are so important that they receive their own name:

Definition 22.1.1 (Representation). A representation of a group G on a vector space V is a group morphism $\rho : G \rightarrow \mathrm{GL}(V)$ from G to the automorphism group 20.2.7 of V.

Property 22.1.2 (Never free). Because every linear map takes the zero vector to itself, a representation can never be free.

Definition 22.1.3 (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 22.1.4 (Permutation representation). Consider a vector space V with basis $\{e_i\}_{i\leq n}$ and let $G = S_n$ be the symmetric group on n elements. Based on Definition 3.3.3, one can consider the action of G on the index set $\{1,\ldots,n\}$. This induces a representation given by

$$\rho(g) : \sum_{i=1}^{n} v_i e_i \mapsto \sum_{i=1}^{n} v_i e_{g\cdot i}. \tag{22.1}$$

Example 22.1.5 (Contragredient representation). For every representation ρ on V, there exists a natural representation on the dual space V^*:

$$\rho^*(g) := \rho^T(g^{-1}) : V^* \rightarrow V^*, \tag{22.2}$$

where ρ^T is the transpose 20.2.22. It is implicitly defined by requiring

$$\Big\langle \rho^*(g)(v^*), \rho(g)(v) \Big\rangle = \langle v^*, v\rangle \tag{22.3}$$

for all $v \in V$ and $v^* \in V^*$, where $\langle\cdot,\cdot\rangle$ is the natural pairing.

Example 22.1.6 (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:

$$\rho_{V\otimes W}(g)(v \otimes w) := \rho_V(g)(v) \otimes \rho_W(g)(w). \tag{22.4}$$

More generally, consider two representations $\phi : G \to \mathrm{GL}(V)$ and $\psi : H \to \mathrm{GL}(W)$. A representation of the direct product $G \times H$ on $V \otimes W$ is given by the tensor product of representations:

$$\rho(g,h)(v \otimes w) := \phi(g)(v) \otimes \psi(h)(w). \tag{22.5}$$

The former case can be obtained as a subrepresentation induced by the diagonal subgroup inclusion $\Delta_G : G \hookrightarrow G \times G$.

Definition 22.1.7 (Intertwiner). If one views G-representations as G-modules, the natural morphisms are the intertwiners 3.3.5.

22.2 Irreducible representations

Definition 22.2.1 (Irreducibility). A representation is said to be irreducible if there exist no proper nonzero subrepresentations.

Example 22.2.2 (Standard representation). Consider the action of S_n on a vector space V with basis $\{e_i\}_{i\leq n}$. The line generated by $e_1 + e_2 + \ldots + e_n$ 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

$$W = \left\{ \sum_{i=1}^n \lambda_i e_i \,\middle|\, \sum_{i=1}^n \lambda_i = 0 \right\} \tag{22.6}$$

forms an irreducible representation. It is called the standard representation of S_n on V.

Schur's lemma 4.8.29 and its corollary are usually found in the following form:

Theorem 22.2.3 (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 22.2.4 (Complementary representation). If W is a subrepresentation of V, 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 associated projection map $\pi_0 : V \to W$. Averaging over G gives the intertwiner

$$\pi(v) := \sum_{g\in G} g \circ \pi_0(g^{-1}v). \tag{22.7}$$

On W it is given by multiplication by $|G|$. Its kernel is an invariant subspace of V, complementary to W.

Theorem 22.2.5 (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*

$$V = V_1^{\oplus a_1} \oplus \cdots \oplus V_k^{\oplus a_k}, \tag{22.8}$$

where all V_k's are distinct irreducible representations.

22.3 Classification by Young tableaux

For an introduction to Young tableaux, see Section 2.7.1.

Definition 22.3.1 (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.

Definition 22.3.2 (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[1] on M^λ. For every Young tableau y, it has a subrepresentation induced by the subgroup Q_t of permutations that leave the columns invariant, which is spanned by the following elements (called **polytabloids**):

$$e_t := \sum_{\sigma\in Q_t} \mathrm{sgn}(\sigma)\sigma\cdot\{t\}, \tag{22.9}$$

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 22.3.3. 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 22.3.4 (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:

$$a_\lambda := \sum_{p\in P_\lambda} p \tag{22.10}$$

$$b_\lambda := \sum_{q\in Q_\lambda} \mathrm{sgn}(q)q. \tag{22.11}$$

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 22.3.5 (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 $\mathrm{GL}(V)$. For all $n\in\mathbb{N}$ there is an induced (diagonal) representation on $V^{\otimes n}$ by $\mathrm{GL}(V)$. 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 22.3.6 (Schur-Weyl). *The above representation of* $\mathrm{GL}(V)\times S_n$ *can be decomposed as follows:*

$$V^{\otimes n} \cong \bigoplus_{|\lambda|=n} V_\lambda \otimes S_\lambda V, \tag{22.12}$$

where

[1] This can be generalized to an action of the group algebra $K[S_n]$.

- *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 $\mathrm{GL}(V)$ of the form $S_\lambda V \equiv \mathrm{Hom}_{S_n}(V_\lambda, V^{\otimes n})$.*

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 22.3.7 (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 34. 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 ??

22.4 Tensor operators

Definition 22.4.1 (Representation operators). An intertwiner $\psi : (\rho, V_0) \to \mathrm{End}(V)$ 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 $\{\hat{O}_i\}_{i\in I} \subset \mathrm{End}(V)$ 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:

$$\rho(g)\hat{O}_i\rho(g)^{-1} = \sum_{j\in I} R(g)_{ij}\hat{O}_j. \tag{22.13}$$

Example 22.4.2 (Tensor operators). Consider $G = \mathrm{SO}(3)$ 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 22.4.3 (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 $\hat{O}$ be a representation operator indexed by a vector space V. For all $v \in V, w_i \in W_i$ one has

$$\langle w_1|\hat{O}(v)w_2\rangle = 0, \tag{22.14}$$

unless $V \otimes W_2$ contains a subrepresentation equivalent to W_1.

Theorem 22.4.4 (Wigner-Eckart). *Consider two irreducible* $\mathrm{SU}(2)$*-subrepresentations* W_j *and* $W_{j'}$ *of some unitary representation* $\mathcal{H}$*, together with two degree-q spherical tensors* $\hat{O}, \hat{O}' : V_0 \to \mathrm{End}(\mathcal{H})$*. 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*

$$\langle v'|\hat{O}_k v\rangle \neq 0,$$

then for all indices $k \geq 0$ *and pairs* $(v, v') \in W_j, W_{j'}$ *the following equality holds*

$$\langle v'|\hat{O}'_k v\rangle = C\langle v'|\hat{O}_k v\rangle \tag{22.15}$$

for some constant C *that only depends on* q, j *and* j'*.*

By noting that the Clebsch-Gordan coefficients 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 22.4.5. 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:

$$\langle j', m'|R^{(q)}|j, m\rangle = \langle j'\|R_k^{(q)}\|j\rangle\langle j', m'|q, j; k, m\rangle. \tag{22.16}$$

The factor $\langle j'\|R^{(q)}\|j\rangle$ is sometimes called the **reduced matrix element**.

Chapter 23

Functional Analysis

The main references for this chapter are [9, 10]. For a revision of topological spaces and inner product spaces, see Chapter 7 and Section 20.3, respectively.

In this chapter the term "linear operator", which was previously reserved for vector space automorphisms, is now used instead of "linear map". This is to keep the terminology in sync with that of the standard literature on Banach spaces and operator spaces. In this chapter "dual space" will indicate the topological/continuous dual and not just the algebraic/linear dual (unless stated otherwise).

23.1 Banach spaces

Definition 23.1.1 (Topological vector space). A vector space for which addition and scalar multiplication are continuous. This is abbreviated as **TVS**.

Definition 23.1.2 (Weak topology). The initial topology 7.2.2 on a TVS with respect to its dual, i.e. a net $(v_\alpha)_{\alpha\in I}$ in V converges to v if and only if $\lambda(v_\alpha) \longrightarrow \lambda(v)$ for all $\lambda \in V^*$.

Definition 23.1.3 (Weak-* topology). Every TVS admits a canonical embedding into its double dual:

$$\iota : V \to V^{**} : v \mapsto \mathrm{ev}_v, \tag{23.1}$$

where the evaluation map ev_v is defined as

$$\mathrm{ev}_v : V^* \to K : \lambda \mapsto \lambda(v). \tag{23.2}$$

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 23.1.4 (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:

1. **Nondegeneracy**: $\|v\| = 0 \iff v = 0$,
2. **Homogeneity**: for all scalars $\lambda \in K : \|\lambda v\| = |\lambda|\,\|v\|$, and
3. **Triangle equality (subadditivity)**: $\|v + w\| \leq \|v\| + \|w\|$.

Remark 23.1.5 (Metric). A norm $\|\cdot\|$ induces a metric 10.1.1 by defining $d(v, w) := \|v - w\|$. The metric topology induced in this way is called the **norm topology** or **strong topology**.

Definition 23.1.6 (Normed vector space). A TVS equipped with a norm $\|\cdot\|$.

Definition 23.1.7 (Banach space). A normed vector space that is complete 10.4.2 in the norm topology.

Property 23.1.8 (Duals). The dual of a Banach space is also a Banach space.

Definition 23.1.9 (Reflexive space). A Banach space V for which the canonical inclusion $V \hookrightarrow V^{**}$ is an (isometric) isomorphism.

Property 23.1.10. Every finite-dimensional Banach space is reflexive.

Property 23.1.11. On a reflexive space, the weak and weak-* topologies coincide.

Property 23.1.12 (Weak duals). Consider a Banach space V. The dual of $(V^*, \tau_{\text{weak}^*})$ is isomorphic to V.

Property 23.1.13 (Continuity). Every linear map $\varphi : V \to W$ of topological vector spaces with V finite-dimensional is continuous. Moreover, if V is normed, every linear isomorphism $\varphi : V \to W$ is a homeomorphism.

Corollary 23.1.14. Two finite-dimensional normed vector spaces with the same dimension are homeomorphic. It follows that all metrics on a finite-dimensional normed vector space are equivalent.

Theorem 23.1.15 (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 23.1.16 (Banach-Alaoglu[2]). *The closed unit ball in the dual of a normed space is compact in the weak-* topology.*

Definition 23.1.17 (Bracket). Let X be a set and consider a normed subset $(\mathcal{F}, \|\cdot\|)$ of $X^{\mathbb{R}}$. For every two functions $u, l : X \to \mathbb{R}$, the bracket $[l, u]$ on $\mathcal{F}$ consists of all functions $f \in \mathcal{F}$ such that

$$\forall x \in X : l(x) \leq f(x) \leq u(x). \tag{23.3}$$

A ε-bracket, for $\varepsilon > 0$, 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 $\mathcal{F}$.

23.2 Hilbert spaces

Remark 23.2.1. Let V be an inner product space 20.3.1. A norm on V can be induced by the inner product in the following way:

$$\|v\|^2 = \langle v|v\rangle. \tag{23.4}$$

However, the converse is not true: not every norm induces an inner product. Only norms that satisfy the **parallelogram law**

$$\|v + w\|^2 + \|v - w\|^2 = 2(\|v\|^2 + \|w\|^2) \tag{23.5}$$

can be used to define an inner product. This inner product can be recovered through the **polarization identity**:

$$4\langle v|w\rangle = \|v + w\|^2 - \|v - w\|^2 + i\big(\|v + iw\|^2 - \|v - iw\|^2\big). \tag{23.6}$$

[1] Sometimes called the **Banach-Schauder** theorem.
[2] Apparently at least 12 different mathematicians should be named in this theorem.

Property 23.2.2 (Cauchy-Schwarz inequality).

$$|\langle v|w\rangle| \leq \|v\|\,\|w\| \tag{23.7}$$

The equality holds if and only if v and w are linearly dependent.

Corollary 23.2.3 (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.

Definition 23.2.4 (Hilbert space). A Banach space where the norm is induced by an inner product.

Example 23.2.5. Consider two square-integrable functions $f, g \in L^2([a,b],\mathbb{C})$. As mentioned in Section 16.3.2, the inner product of f and g is defined as follows:

$$\langle f|g\rangle = \int_a^b \overline{f(x)}g(x)\,dx. \tag{23.8}$$

It is possible to generalize this inner product with respect to a weight function $\phi \in L^2([a,b],\mathbb{C})$:

$$\langle f|g\rangle_\phi := \int_a^b \overline{f(x)}g(x)\phi(x)\,dx. \tag{23.9}$$

Formula 23.2.6 (Pythagoras). In an inner product space the triangle equality reduces to the well-known Pythagorean theorem for orthogonal vectors:

$$\langle v|w\rangle = 0 \implies \|v+w\|^2 = \|v\|^2 + \|w\|^2. \tag{23.10}$$

This formula can be extended to any set of orthogonal vectors $v_1, \ldots, v_n$ as follows:

$$\left\|\sum_{i=1}^n v_i\right\|^2 = \sum_{i=1}^n \|v_i\|^2. \tag{23.11}$$

Theorem 23.2.7 (Riesz 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*

$$\rho(h) = \langle v_0|h\rangle \tag{23.12}$$

for all $h \in \mathcal{H}$. This implies that $\mathcal{H}$ and $\mathcal{H}^*$ are isometrically isomorphic.[3] *Furthermore, the operator norm of ρ is equal to the norm of v_0.*

Remark 23.2.8. 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 23.2.9 (Relation to Riesz-Markov theorem). Recall the Riesz-Markov theorem 17.1.6. Every continuous functional on $C(\widehat{X})$ can be written as the integration againt some Radon measure. By using the theorem that every Hilbert space is isomorphic to some function space $L^2(X, \mu_{\text{count}})$, together with Equation (23.8), one can obtain the representation theorem above.

[3]Anti-isomorphic in the convention with conjugate linearity in the second argument.

23.2.1 Generalized Fourier series

Property 23.2.10 (Bessel's inequality). The following general equality holds for all orthonormal vectors $v_1, \ldots, v_n$ and scalars $\lambda_1, \ldots, \lambda_n$:

$$\left\| v - \sum_{i=1}^{n} \lambda_i v_i \right\|^2 = \|v\|^2 - \sum_{i=1}^{n} |\langle v|v_i\rangle|^2 + \sum_{i=1}^{n} |\langle v|v_i\rangle - \lambda_i|^2. \tag{23.13}$$

This expression is minimized when the last term vanishes. This leads to Bessel's inequality

$$\sum_{i=1}^{n} |\langle v|v_i\rangle|^2 \leq \|v\|^2, \tag{23.14}$$

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 23.2.11. The sum in (23.14) 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 23.2.12. *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 expansion $\sum_{i=1}^{\infty} \lambda_i v_i$ converges in $\mathcal{H}$ if and only if $(\lambda_n)_{n\in\mathbb{N}} \in l^2$. Furthermore, the expansion satisfies the following equality:*

$$\left\| \sum_{i=1}^{\infty} \lambda_i v_i \right\|^2 = \sum_{i=1}^{\infty} |\lambda_i|^2. \tag{23.15}$$

Bessel's inequality implies that the sequence $(\langle v, v_n\rangle)_{n\in\mathbb{N}}$ belongs to l^2, so the generalized Fourier series of $v \in \mathcal{H}$ converges in $\mathcal{H}$.

Remark 23.2.13. 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 23.2.14 (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:

$$v = \sum_{i\in I} \langle e_i|v\rangle e_i. \tag{23.16}$$

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 23.2.15. 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 to it satisfying

$$\forall v_i \in S : \langle v_i|w\rangle = 0. \tag{23.17}$$

Theorem 23.2.16 (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.*

23.2.2 Orthogonality and projections

The basic notions on orthogonality in inner product space can be found in Section 20.3.2.

Property 23.2.17. 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 23.2.18. The previous property implies that the orthogonal complemement of some arbitrary subset of a Hilbert space is a Hilbert space itself.

Theorem 23.2.19 (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 23.2.20. An equivalent definition for this projection is the vector v' satisfying

$$\|v - v'\| = \inf\{\|v - w\| \mid w \in S\}. \tag{23.18}$$

It is often denoted by $\mathrm{proj}_S(v)$.

Corollary 23.2.21. 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 23.2.22 (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:

$$\mathrm{tr}(S) := \sum_{i\in I} \langle e_i | S e_i \rangle. \tag{23.19}$$

23.2.3 Separable Hilbert spaces

The definition of separable spaces in the sense of point-set topology is given in 7.5.23. An equivalent definition for Hilbert spaces is the following one (provided that one accepts Zorn's lemma 2.6.11):

Alternative Definition 23.2.23 (Separable space). A Hilbert space that contains a Hilbert basis.

Example 23.2.24 (Finite dimensions). By the Gram-Schmidt method it follows that every finite-dimensional Hilbert space is separable.

The following theorem shows that (up to an isomorphism) there are only two distinct types of separable Hilbert spaces:

Theorem 23.2.25. *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, the space of square-summable sequences.*

Property 23.2.26. Every orthogonal subset of a separable Hilbert space is countable.

23.3 Seminorms

Definition 23.3.1 (Seminorm). Let V be a K-vector space, where K is a normed field. 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 $v \in V$, and

2. **Triangle equality (subadditivity)**: $p(v + w) \leq p(v) + p(w)$ for all $v, w \in V$.

Theorem 23.3.2 (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*

$$|F(V)| \leq p(v) \tag{23.20}$$

for all $v \in V$.

23.3.1 Topology

In this subsection $\mathscr{P}$ denotes a family of seminorms defined on a TVS V with index family I.

Definition 23.3.3 ($\mathscr{P}$-open ball). A $\mathscr{P}$-open ball centered on v_0 is a subset $W \subseteq V$ such that all points $w \in W$ satisfy the following condition for a finite number of seminorms $p_i \in \mathscr{P}$:

$$p_i(w - v_0) \leq \varepsilon_i, \tag{23.21}$$

where $\varepsilon_i > 0$.

Property 23.3.4. The set of $\mathscr{P}$-open balls generates a topology on V. This topology is often called the **$\mathscr{P}$-topology**.

Definition 23.3.5 (Separated family). A family of seminorms $\mathscr{P}$ is said to be separated if for every point $v \in V, v \neq 0$ there exists a seminorm $p \in \mathscr{P}$ such that $p(v) \neq 0$. If $\mathscr{P}$ is separated, then $\sum_i p_i$ is a norm.

Property 23.3.6. 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 metric, which is translation-invariant, is given by

$$d(v, w) := \sum_{i \in I} \frac{1}{2^i} \frac{p_i(v - w)}{1 + p_i(v - w)}. \tag{23.22}$$

Although the Hahn-Banach theorem 23.3.2 does not imply that the linear extension is unique, one can refine the statement in the case of dense subspaces:

Corollary 23.3.7. 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.

23.3.2 Locally convex spaces

Definition 23.3.8 (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 23.3.9 (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 23.3.10 (Fréchet space). A locally convex topological vector space that admits a complete translation-invariant metric.

By Property 23.3.6 there exists an equivalent formulation:

Alternative Definition 23.3.11 (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 23.3.12 (Gâteaux derivative). The Gâteaux differential of a continuous map of locally convex spaces $f : V \to W$ is defined as follows:

$$df(v;h) := \lim_{t\to 0}\frac{f(v+th)-f(v)}{t} = \left.\frac{d}{dt}f(v+th)\right|_{t=0}. \tag{23.23}$$

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:

$$d^{(k)}f(v;h_1\otimes\cdots\otimes h_k) := \left.\frac{d^k}{dt^k}f(v+th_1+\cdots+th_k)\right|_{t=0}. \tag{23.24}$$

Now, it should be noted that the map $df(v;-)$ is not necessarily linear. If it is linear, the function $\delta_v f : V \to W : h \mapsto df(v;h)$ 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 23.3.13 (Fundamental theorem of calculus). If $f \in C^1(V,W)$, then

$$f(v+h) = f(v) + \int_0^1 df(v+th;h)dt, \tag{23.25}$$

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

$$\langle\varphi, e_A\rangle = \int_A \varphi\circ f\,d\mu, \tag{23.26}$$

then f is said to be **Gelfand-Pettis integrable** and v_A is called the **Gelfand-Pettis integral** of f over A.

One can also introduce an alternative notion of differentiability:

Definition 23.3.14 (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 Df_v such that

$$\lim_{\|h\|\to 0} \frac{\|f(v+h) - f(v) + Df_v(h)\|}{\|h\|} = 0. \tag{23.27}$$

If the linear operator Df 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 $V \to \mathcal{B}(V,W) : v \mapsto Df_v$ is continuous, then f is said to be of class C^1.

The relation between Gâteaux and Fréchet derivatives is clarified by the following property:

Property 23.3.15. If a function $f : V \to W$ 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 *df* and Fréchet derivative Df coincide.

Although one can extend functional analysis to Fréchet spaces (or even locally convex spaces), they are less well-behaved than Banach spaces:

Property 23.3.16. 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 of linear maps between Fréchet spaces $\mathcal{L}(V,W)$ is Fréchet if and only if W is Banach.

Theorem 23.3.17 (Krein-Milman). *Every compact, convex subset of a locally convex Hausdorff space is equal to the convex hull of its extreme points.*

The following theorem gives an extension

Theorem 23.3.18 (Choquet). *For every convex, compact subset C of a normed vector space there exists an assignment*

$$\mu : C \to \mathbb{P}(\mathrm{Extr}(C)) : c \mapsto \mu_c \tag{23.28}$$

such that

$$f(c) = \int_{\mathrm{Extr}(C)} f \, d\mu_c \tag{23.29}$$

for all affine functions f.

Remark 23.3.19. The extension of Choquet's theorem to locally convex topological vector spaces is called the **Choquet-Bishop-de Leeuw theorem**.

Definition 23.3.20 (Strong topology). Let X be a topological vector space. The strong topology (or **topology of uniform convergence on bounded sets**) on X'^* is the locally convex topology defined by the seminorms

$$p_B(\lambda) := \sup_{x\in B} |\lambda(x)|, \tag{23.30}$$

where B is a bounded set of X. This is an example of a *polar topology.* A dual space equipped with the strong topology is often called the **strong dual**.

23.3.3 Tensor products

When moving from finite-dimensional vector spaces to general topological vector spaces, the algebraic tensor product from Section 21.3 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 23.3.21 (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

$$\langle v_1 \otimes w_1 | v_2 \otimes w_2\rangle_{V\otimes W} := \langle v_1|v_2\rangle_V \, \langle w_1|w_2\rangle_W \tag{23.31}$$

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 23.3.22 (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:

$$\|x\|_{\text{proj}} := \inf\left\{\sum_{i=1}^n \|a_i\|\|b_i\| \,\middle|\, x = \sum_{i=1}^n a_i \otimes b_i\right\} \tag{23.32}$$

and

$$\|x\|_{\text{inj}} := \sup\left\{|(\mu\otimes\nu)(x)| \,\middle|\, \mu \in V^*, \nu \in W^* : \|\mu\| = \|\nu\| = 1\right\}. \tag{23.33}$$

These two norms are called the **projective** and **injective** norms, respectively. Accordingly, the completions $V \otimes_\pi W$ and $V \otimes_\varepsilon W$ of the algebraic tensor product $V \otimes W$ with respect to these norms are called the **projective** and **injective** tensor products.

Definition 23.3.23 (Tensor products of locally convex spaces). Let V, W be locally convex spaces. Definition 23.3.9 gives rise to a family of projective seminorms as in the definition above. These define the projective tensor products $V \otimes_\pi W$. Note that in general the projective tensor product is not complete, even when both V and W are.[4] The completion is denoted by $V \widehat{\otimes}_\pi W$.

The injective tensor product can be extended to locally convex spaces as follows. Equip the dual V' with the finest locally convex topology that coincides with the weak topology on equicontinuous sets 10.5.7. The, equip the space of continuous linear maps $L_{\text{cont}}(V', W)$ with the topology of uniform convergence on compact equicontinuous sets. $V \otimes W$ with the induced topology, as a subset of $L_{\text{cont}}(V', W)$, is called the injective tensor product $V \otimes_\varepsilon W$.

Alternative Definition 23.3.24. Let V, W be locally convex TVSs. The projective tensor product $V \otimes_\pi W$ carries the finest locally convex topology with respect to the canonical injection $V \times W \to V \otimes W : (v, w) \mapsto v \otimes w$.

Definition 23.3.25 (Nuclear space). A locally convex space such that the injective and projective tensor products with any other locally convex space are isomorphic.

[4] In fact, if both V and W are infinite-dimensional Banach spaces, their tensor product (in this sense) will never be complete.

23.3.4 Measure theory

Definition 23.3.26 (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 17.1.6:

Remark 23.3.27. For finite-dimensional spaces the distributions are in a one-to-one correspondence with the regular Borel measures.

?? COMPLETE ??

23.4 Operators

23.4.1 Operator topologies

Definition 23.4.1 (Weak operator topology). The topology generated by the seminorms $\big\{T \to |\lambda(Tv)| \,\big|\, v \in V, \lambda \in V^*\big\}$. 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 $T_\alpha x \longrightarrow Tx$ for all x in the weak topology.

In the case of Hilbert spaces one can simplify the above definition using Riesz's representation theorem 23.2.7. The weak operator topology on a Hilbert space is generated by the seminorms $\big\{T \mapsto |\langle Tv|w\rangle| \,\big|\, v, w \in \mathcal{H}\big\}$.

Definition 23.4.2 (Strong operator topology). The topology generated by the seminorms $\big\{T \to \|Tv\| \,\big|\, v \in V\big\}$. 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 the norm (strong) topology.

Definition 23.4.3 (Operator norm). The operator norm of L is defined as follows:

$$\|L\|_{\text{op}} = \inf\big\{M \in \mathbb{R} \,\big|\, \forall v \in V : \|Lv\|_W \leq M\|v\|_V\big\}. \tag{23.34}$$

Equivalent definitions of the operator norm are:

$$\|L\|_{\text{op}} = \sup_{\|v\|\leq 1} \|L(v)\| = \sup_{\|v\|=1} \|L(v)\| = \sup_{v\neq 0} \frac{\|L(v)\|}{\|v\|}. \tag{23.35}$$

Definition 23.4.4 (Norm topology[5]). 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 7.1.26.)

23.4.2 Bounded operators

Definition 23.4.5 (Bounded operator). Let $L : V \to W$ be a linear operator between two normed spaces. The linear operator is said to be bounded if it satisfies

$$\|L\|_{\text{op}} < \infty. \tag{23.36}$$

Notation 23.4.6. The space of bounded linear operators from V to W is denoted by $\mathcal{B}(V, W)$.

[5] Also called the **uniform (operator) topology**.

Property 23.4.7. If V is a Banach space, $\mathcal{B}(V)$ is also a Banach space.

The following property reduces the problem of continuity to that of boundedness (or vice versa):

Property 23.4.8. Consider a linear operator $f \in \mathcal{L}(V, W)$. The following statements are equivalent:

- f is bounded.
- f is continuous at 0.
- f is continuous on V.
- f is uniformly continuous.
- f maps bounded sets to bounded sets.

Property 23.4.9 (Eigenvalue bound). Let A be a bounded linear operator. The eigenvalues of A are bounded by its operator norm. Furthermore, every bounded linear operator on a Banach space has at least one eigenvalue.

Property 23.4.10 (BLT theorem[6]). 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 23.4.11 (Schatten class operator). Consider the space of bounded linear operators on a Hilbert space $\mathcal{H}$. The **Schatten p-norm** is defined as

$$\|T\|_p = \text{tr}\left(\sqrt{T^*T}^{\ p}\right)^{1/p}. \tag{23.37}$$

Linear operators for which this norm is finite form the p^{th} Schatten class $\mathcal{I}_p$.

Property 23.4.12. The Schatten classes are Banach spaces with respect to the associated Schatten norms.

Example 23.4.13 (Trace class operator). The space of trace class operators on a Hilbert space $\mathcal{H}$ is defined as follows:

$$\mathcal{B}_1(\mathcal{H}) := \{S \in \mathcal{B}(\mathcal{H}) \mid \text{tr}(|S|) < \infty\}, \tag{23.38}$$

where the trace functional was defined in 23.2.22 and $|S| := \sqrt{S^*S}$.

The following theorem can be seen as the analogue of Riesz's theorem for trace class operators:

Property 23.4.14. 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

$$\rho(S) = \text{tr}(ST) \tag{23.39}$$

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.

The previous property allows for the following definition:

Definition 23.4.15 (Weak-* operator topology). The 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 $\mathcal{B}(\mathcal{H})$.

[6] BLT stands for ''bounded linear transformation''.

Example 23.4.16 (Hilbert-Schmidt operator). Consider the Hilbert-Schmidt norm $\|\cdot\|_2$ from Definition 20.4.7. A linear operator $T \in \mathcal{B}(\mathcal{H})$ is said to be a Hilbert-Schmidt operator if it satisfies

$$\|T\|_2 < \infty. \tag{23.40}$$

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 23.4.17 (Closed operator). A linear operator $f : V \to W$ such that for every sequence $(v_n)_{n\in\mathbb{N}}$ in $\mathrm{dom}(f)$ converging to $v \in V$, where $f(v_n)$ converges to $w \in W$, one finds that $v \in \mathrm{dom}(f)$ and $f(v) = w$.

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 23.4.18 (Closure). Let $f : V \to W$ be a linear operator. Its closure (if it exists) is the closed linear operator $\overline{f}$ such that the graph of $\overline{f}$ is the closure of the graph of f in $V \oplus W$.

Theorem 23.4.19 (Closed graph theorem). *A linear operator on a Banach space is closed if and only if it is bounded.*

23.4.3 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 all linear operators are allowed to be unbounded.

Definition 20.3.6 for finite-dimensional spaces can be generalized as follows:

Definition 23.4.20 (Adjoint). Let A be a linear operator on a Hilbert space $\mathcal{H}$. A linear operator A^* is said to be the adjoint of A if the following conditions are satisfied:

1. $\langle v|Aw\rangle = \langle A^*v|w\rangle$ for all $v \in \mathrm{dom}(A^*)$ and $w \in \mathrm{dom}(A)$.
2. Every other linear operator B satisfying this property is a restriction of A^* (i.e. the domain of A^* is maximal with respect to the above property).

Property 23.4.21. Let A be a bounded linear operator. Its adjoint A^* is also bounded and $\|A\|_{\mathrm{op}} = \|A^*\|_{\mathrm{op}}$.

Definition 23.4.22 (Symmetric operator). A linear operator A on a Hilbert space $\mathcal{H}$ such that $\mathrm{dom}(A) \subseteq \mathrm{dom}(A^*)$ and $A = A^*|_{\mathrm{dom}(A)}$.

Definition 23.4.23 (Self-adjoint operator). A linear operator A on a Hilbert space $\mathcal{H}$ such that $\mathrm{dom}(A)$ is dense in $\mathcal{H}$ and $A = A^*$.

The notion of Hermitian operator is the one where almost nobody agrees upon its definition. Here the definition from [12] is chosen:

Definition 23.4.24 (Hermitian operator). A bounded symmetric operator.

Theorem 23.4.25 (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 23.4.26 (Stone). *Consider a strongly continuous unitary one-parameter group, i.e. a family of unitary operators $U : \mathbb{R} \to \mathrm{U}(\mathcal{H})$ such that*

- *U is continuous in the strong operator topology:*

$$\lim_{t\to t_0} U(t)x = U(t_0)x$$

for all $t_0 \in \mathbb{R}, x \in \mathcal{H}$.

- *U forms a one-parameter group in the sense of Definition 30.1.9.*

There exists a self-adjoint operator A such that $U(t) = e^{itA}$. Furthermore, the linear operator A is bounded if and only if U is continuous in the norm topology.

Definition 23.4.27 (Generator). The linear operator A is called the (infinitesimal) generator of the family U. It can be obtained through a formal derivative:

$$A = \left.\frac{dU(t)}{dt}\right|_{t=0}. \tag{23.41}$$

23.4.4 Compact operators

Definition 23.4.28 (Compact operator). Let V, W be Banach spaces. A linear operator $A : V \to W$ is said to be compact if the image of any bounded set in V is relatively compact 7.5.11.

Alternative Definition 23.4.29 (Compact operator). Let V, W be Banach spaces. A linear operator $A : V \to W$ is compact if for every bounded sequence $(v_n)_{n\in\mathbb{N}}$ in V the sequence $(Av_n)_{n\in\mathbb{N}} \subset W$ has a convergent subsequence.

Notation 23.4.30. 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 23.4.31. $\mathcal{B}_0(V)$ is a two-sided ideal in the (Banach) algebra $\mathcal{B}(V)$.

Property 23.4.32 (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 23.4.33. Every compact operator is bounded.

Corollary 23.4.34. Every linear map between finite-dimensional Banach spaces is bounded.

Property 23.4.35. If A is a compact self-adjoint operator on a Hilbert space, then $-\|A\|$ or $\|A\|$ are an eigenvalue of A. Furthermore, the set of nonzero eigenvalues is either finite or converges to 0.

Definition 23.4.36 (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 23.4.37 (Fredholm operator). A bounded linear operator $F \in \mathcal{B}(V, W)$ for which the kernel and cokernel are finite-dimensional. The space of Fredholm operators is denoted by $\mathfrak{F}(V)$.

By a theorem of Atkinson one can characterize Fredholm operators using the Calkin algebra:

Property 23.4.38 (Atkinson). A bounded linear operator $F : V \to W$ is a Fredholm operator if and only if it is invertible in the Calkin algebra, i.e. there exists a bounded linear operator $G : W \to V$ and compact operators C_1, C_2 such that $\mathbb{1}_W - FG = C_1$ and $\mathbb{1}_V - GF = C_2$. G is called the **parametrix** of F.

Definition 23.4.39 (Fredholm index). The index of a Fredholm operator F is defined as follows:

$$\mathrm{ind}(F) := \dim\ker(F) - \dim\mathrm{coker}(F). \tag{23.42}$$

Property 23.4.40. The induced function

$$\mathrm{ind} : \pi_0(\mathfrak{F}(\mathcal{H})) \to \mathbb{Z} \tag{23.43}$$

is a group isomorphism:

- $\mathrm{ind}(F^*) = -\mathrm{ind}(F)$, and
- $\mathrm{ind}(FG) = \mathrm{ind}(F) + \mathrm{ind}(G)$.

This theorem is generalized in K-theory by the Atiyah-Jänich theorem 39.2.17.

23.4.5 Spectrum

Definition 23.4.41 (Resolvent operator). Let A be a bounded linear operator on a normed space V. The resolvent operator of A_λ for some $\lambda \in \mathbb{C}$ is defined as the linear operator $(A - \lambda\mathbb{1}_V)^{-1}$.

Definition 23.4.42 (Resolvent set). The resolvent set $\rho(A)$ 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 A.

Definition 23.4.43 (Spectrum). The set of scalars $\mu \in \mathbb{C} \setminus \rho(A)$ is called the spectrum $\sigma(A)$.

Remark 23.4.44. From Remark 20.5.2 it is clear that every eigenvalue of A belongs to the spectrum of A. The converse, however, is not true. This is remedied by introducing the following concepts:

Definition 23.4.45 (Point spectrum). The set of scalars $\mu \in \mathbb{C}$ for which $A - \mu\mathbb{1}_V$ fails to be injective is called the point spectrum $\sigma_p(A)$. This set coincides with the set of eigenvalues of A.

Definition 23.4.46 (Continuous spectrum). The set of scalars $\mu \in \mathbb{C}$ for which $A - \mu\mathbb{1}_V$ is injective with dense image but fails to be surjective is called the continuous spectrum of A.

Definition 23.4.47 (Residual spectrum). The set of scalars $\mu \in \mathbb{C}$ for which $A - \mu\mathbb{1}_V$ is injective but fails to have a dense image is called the residual spectrum $\sigma(A)$.

Definition 23.4.48 (Essential spectrum). The set of scalars $\mu \in \mathbb{C}$ for which $A - \mu\mathbb{1}_V$ is not a Fredholm operator is called the essential spectrum $\sigma_{\text{ess}}(A)$.

From Atkinson's theorem[7] one can derive the following result:

Property 23.4.49. Let A be a bounded linear operator and let T be a compact operator. The essential spectra of A and $A + T$ coincide.

Property 23.4.50. 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 $\mathbb{R}^+$.

[7]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.

23.4.6 Spectral theorem

This section focuses on the algebra of bounded operators $\mathcal{B}(\mathcal{H})$ on a (complex) Hilbert space $\mathcal{H}$.

Property 23.4.51 (Closed subspaces). There exists a bijection between the set of closed subspaces of $\mathcal{H}$ and the set of 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 23.4.52 (Projection-valued measure). Consider a topological space X and let Σ be a σ-algebra 2.5.2 on X. A projection-valued measure (PVM) or **spectral measure**[8] on X is a map $P : \Sigma \to \mathcal{B}(\mathcal{H})$ satisfying the following conditions:[9]

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 $(E_n)_{n\in\mathbb{N}} \subset \Sigma$:

$$\sum_{n\in\mathbb{N}} P_{E_n} = P_{\bigcup_{n\in\mathbb{N}} E_n}. \tag{23.44}$$

Property 23.4.53. Let P be a spectral measure on (X, Σ). For every two elements $v, w \in \mathcal{H}$ the map

$$E \mapsto \mu^P_{v,w}(E) := \langle v|P_E w\rangle \tag{23.45}$$

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 23.4.54. 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

$$\int_X f(\lambda)\,dP(\lambda) : \Delta_f \to \mathcal{H} \tag{23.46}$$

defined by

$$\left\langle v \,\middle|\, \int_X f(\lambda)\,dP(\lambda) w \right\rangle := \int_X f(\lambda)\,d\mu^P_{v,w}(\lambda) \tag{23.47}$$

is closed and normal. It also satisfies the following two equalities:

$$\left(\int_X f(\lambda)\,dP(\lambda)\right)^* = \int_X \overline{f(\lambda)}\,dP(\lambda) \tag{23.48}$$

$$\left\|\int_X f(\lambda)\,dP(\lambda)v\right\|^2 = \int_X |f(\lambda)|^2 d\mu^P_{v,v}(\lambda). \tag{23.49}$$

If f is bounded, the above operator is bounded by the supremum norm of f:

$$\left\|\int_X f(\lambda)dP(\lambda)\right\| \leq \|f\|_\infty. \tag{23.50}$$

[8]Sometimes also called a **resolution of the identity**.

[9]The third property can in fact be shown to follow from the others.

Theorem 23.4.55 (Spectral decomposition). *Let A be a self-adjoint operator on a Hilbert space $\mathcal{H}$. There exists a unique spectral measure $P_A : \Sigma \to \mathcal{B}(\mathcal{H})$ on the Borel σ-algebra of the real line such that*

$$A = \int_{\mathbb{R}} \lambda \, dP_A(\lambda). \tag{23.51}$$

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 an orthonormal basis of $\mathcal{H}$ consisting of eigenvectors of A and, moreover, the eigenvalues are necessarily real:

$$A = \sum_{i=1}^{\infty} \lambda_i e_i^* \otimes e_i. \tag{23.52}$$

Property 23.4.56 (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 23.4.57 (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 23.4.58 (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

$$\Phi : \mathrm{Meas}(X, \mathbb{C}) \to \mathcal{B}_0(\mathcal{H}) \tag{23.53}$$

satisfying the following conditions:

1. **Identity**: $\Phi(1) = \mathbb{1}_{\mathcal{H}}$.
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**:[10] $\Phi(f)$ is densely defined.
5. **Involutivity**: $\Phi(f^*) = \overline{\Phi(f)}$.
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.

Property 23.4.59. If $(\Phi, \mathcal{H})$ is a measurable functional calculus on (X, Σ), then

$$P_\Phi : \Sigma \to \mathcal{B}(\mathcal{H}) : E \mapsto \Phi(\mathbb{1}_E) \tag{23.54}$$

is a projection-valued measure. Conversely, every proejction-valued measure P gives rise to a measurable functional calculus Φ such that $P = P_\Phi$.

Definition 23.4.60 (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 $\Phi(\mathbb{1}_{\mathbb{C}}) = A$.

[10]This condition is actually redundant.

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 23.4.61 (Borel functional calculus). 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 $\sigma(A)$.

$$f(A) := \int_{\sigma(A)} f(\lambda)\,dP_A(\lambda) = \int_{\mathbb{R}} g(\lambda)\,dP_A(\lambda) =: g(A). \tag{23.55}$$

23.4.7 Dixmier trace

Definition 23.4.62 (Dixmier ideal). For every compact operator one can sort the singular values in decreasing order. One can then define the operator

$$\sigma_n(A) := \sum_{i=1}^n s_i(A). \tag{23.56}$$

The Dixmier ideal is defined as follows:

$$\mathfrak{D}(\mathcal{H}) := \{A\in \mathrm{End}_0(\mathcal{H}) \mid \sigma_n(A) = O(\ln(n))\}. \tag{23.57}$$

The following functional gives an alternative for the ordinary trace on operators that are not trace-class.

Definition 23.4.63 (Dixmier trace). For every element in $A\in\mathfrak{D}(\mathcal{H})$, the Dixmier trace is defined as follows:

$$\mathrm{tr}_{\mathfrak{D}}(A) := \omega\left(\left\{\frac{\sigma_n(A)}{1+\log(n)} \mid n\in\mathbb{N}\right\}\right), \tag{23.58}$$

where $\omega:\ell^\infty(\mathbb{N})$ is a linear functional satisfying the 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:

$$D:(s_1,s_2,\ldots)\mapsto(s_1,s_1,s_2,s_2,\ldots). \tag{23.59}$$

The state ω is dilation-invariant if $\omega(s)=\omega(D(s))$ 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 23.4.64. 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):

$$A\in\mathcal{B}_1(\mathcal{H}) \implies \mathrm{tr}_{\mathfrak{D}}(A)=0. \tag{23.60}$$

Chapter 24

Operator Algebras

The main reference for this chapter is [36].

24.1 C^*-algebras

24.1.1 Involutive algebras

Definition 24.1.1 (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 24.1.2 (C^*-algebra). A C^*-algebra is an involutive Banach algebra 23.1.7 such that the **C^*-identity**

$$\|a^* a\| = \|a\| \|a^*\| \tag{24.1}$$

is satisfied.

The Artin-Wedderburn theorem 3.6.39 implies the following decomposition theorem:

Theorem 24.1.3. *Let C be a finite-dimensional C^*-algebra. There exist unique integers N and $d_1, \ldots, d_N$ such that*

$$C \cong \bigoplus_{i=1}^{N} M_{d_i}(K). \tag{24.2}$$

This implies that every C^*-algebra can be represented using block matrices.

Example 24.1.4 (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 24.1.5. Every norm-closed $*$-subalgebra of a C^*-algebra is a C^*-algebra.

Definition 24.1.6 (Normal element). An element of a $*$-algebra that commutes with its adjoint:

$$a^*a = aa^*. \tag{24.3}$$

Property 24.1.7. Every element in a $*$-algebra can be decomposed as the sum of two normal elements:

$$a = \frac{1}{2}\big((a + a^*) + (a - a^*)\big). \tag{24.4}$$

This implies that a linear morphism defined on normal elements extends uniquely to the whole algebra.

Definition 24.1.8 (H^*-algebra). A Hilbert space $\mathcal{H}$ equipped with a unital $*$-algebra structure that satisfies the following conditions for all $a, b, c \in \mathcal{H}$:

1. $\langle ab|c\rangle = \langle b|a^*c\rangle$, and
2. $\langle ab|c\rangle = \langle a|cb^*\rangle$.

Example 24.1.9 (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 20.4.7 (up to a factor $k > 0$):

$$\langle f|g\rangle_{\text{HS}} := k\ \text{tr}(f^*g). \tag{24.5}$$

The resulting space is denoted by $L^2(\mathcal{H}, k)$. A result, analogous to the Artin-Wedderburn theorem 3.6.39, states that every H^*-algebra can be decomposed as an orthogonal direct sum of finitely many algebras of the form $L^2(\mathcal{H}_i, k_i)$.

24.1.2 Positive maps

Definition 24.1.10 (Positive element). A self-adjoint element of a C^*-algebra A for which its spectrum is contained in $[0, +\infty[$. The cone of all positive elements in A is often denoted by A^+.

Property 24.1.11 (Positive decomposition). Every positive element a can be written as $a = b^*b$ for some element b. Moreover, every positive element admits a positive square-root.

Definition 24.1.12 (Cuntz algebra). The n^{th} Cuntz algebra $\mathcal{O}_n$ is defined as the (universal) unital C^*-algebra generated by n isometric elements s_i under the additional relation

$$\sum_{i=1}^{n} s_i^* s_i = 1, \tag{24.6}$$

where 1 is the unit element.

Definition 24.1.13 (Positive map). A morphism of C^*-algebras is called positive if every positive element is mapped to a positive element.

Definition 24.1.14 (Completely positive map). A morphism of C^*-algebras $T : A \rightarrow B$ is called completely positive if for all $k \in \mathbb{N}$ the following map is positive:

$$\mathbb{1}_k \otimes T : \mathbb{C}^{k\times k} \otimes A \rightarrow \mathbb{C}^{k\times k} \otimes B. \tag{24.7}$$

If T satisfies this condition only up to an integer n, it is said to be n**-positive**.

Definition 24.1.15 (State). A positive linear functional of unit norm on a C^*-algebra.

Definition 24.1.16 (Positivity-improving map). A positive map ϕ that satisfies

$$a \geq 0, a \neq 0 \implies \phi(a) > 0. \tag{24.8}$$

Definition 24.1.17 (Ergodic map). A positive map ϕ that satisfies

$$\forall a > 0 : \exists t \in \mathbb{R}_0 : \exp(t_a\phi)a > 0. \tag{24.9}$$

Property 24.1.18 (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**.

Example 24.1.19 (Function algebras). 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 17.1.6 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).

The following is some kind of Jordan algebra-theoretic analogue of the Gel'fand-Naimark theorem 24.1.41:

Theorem 24.1.20 (Alfsen-Schutz). *The state spaces of two C^*-algebras are isomorphic if and only if the algebras are isomorphic as (special) Jordan algebras 20.7.8.*

24.1.3 Traces

Definition 24.1.21 (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 24.1.22. 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 24.1.23 (Adjoint map). Assume that a trace functional tr is given on a C^*-algebra and consider a continuous linear map ϕ defined on the Schatten class $\mathcal{I}_p$. One can define the adjoint map ϕ^* on $\mathcal{I}_q$ whenever p, q are Hölder conjugate. This adjoint is given by the following equation:

$$\mathrm{tr}\left((\phi^*(A))^*B\right) = \mathrm{tr}\left(A^*\phi(B)\right), \tag{24.10}$$

where $A \in \mathcal{I}_q, B \in \mathcal{I}_p$.

Definition 24.1.24 (Trace-preserving map). A map ϕ is said to be trace-preserving if it satisfies

$$\mathrm{tr}(\phi(A)) = \mathrm{tr}(A) \tag{24.11}$$

for all trace-class elements A. Using the above definition it is easily seen that on a unital C^*-algebra this is equivalent to

$$\phi^*(1) = 1. \tag{24.12}$$

Property 24.1.25. A completely positive, trace-preserving map ϕ satisfies:

$$\|\phi\|_1 = 1, \tag{24.13}$$

where the subscript 1 indicates that this operator is defined on trace-class elements.

Property 24.1.26 (States). Whenever a C^*-algebra is commutative, every state defines a trace and, possibly after a suitable normalization, every trace defines a state. However, for noncommutative C^*-algebras only the latter implication holds.

24.1.4 Representations

Definition 24.1.27 (C^*-algebra representation). A representation of a C^*-algebra A is a unital $*$-morphism $A \to \mathcal{B}(\mathcal{H})$.

Definition 24.1.28 (Normal state). Consider a Hilbert space $\mathcal{H}$ and a C^*-representation $\pi : A \to \mathcal{B}(\mathcal{H})$. A normal state ω is a state such that there exists a trace-class operator $\rho \in \mathcal{B}_1(\mathcal{H})$ with the following property:

$$\omega(a) = \frac{\mathrm{tr}(\rho\pi(a))}{\mathrm{tr}(\rho)}. \tag{24.14}$$

In case where $A = \mathcal{B}(\mathcal{H})$, the normal states are exactly the σ-weakly continuous states 23.4.15.

Definition 24.1.29 (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 24.1.30 (Cyclic vector). A cyclic vector for a C^*-algebra representation $\rho : A \to \mathcal{B}(\mathcal{H})$ is a vector $\xi \in \mathcal{H}$ such that $\{\rho(a)\xi \mid a \in A\}$ is (norm) dense in $\mathcal{H}$.

Theorem 24.1.31 (Stinespring). *Consider a linear operator $\Phi : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$. It is completely positive if and only if there exists a C^*-representation $\pi : \mathcal{B}(\mathcal{H}_2) \to \mathcal{B}(\mathcal{K})$ and a bounded linear operator $V : \mathcal{H}_1 \to \mathcal{K}$ such that*

$$\Phi^\dagger(a) = V^\dagger\pi(a)V \tag{24.15}$$

or, equivalently,

$$\Phi(\rho) = \pi^\dagger(V\rho V^\dagger). \tag{24.16}$$

Moreover, V is an isometry if and only if Φ is trace-preserving.

Remark 24.1.32. Because the adjoint of a completely positive map is again completely positive, the above two characterizations can be used interchangeably. The former is the mostly used by mathematicians, while the latter is better known in the physics literature.

The most common situation is where $\mathcal{H} := \mathcal{H}_1 = \mathcal{H}_2$ and $\pi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H} \otimes \mathcal{K}) : a \mapsto a \otimes b$, 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:

$$\Phi^\dagger(a) = V^\dagger(a \otimes b)V \tag{24.17}$$

and, equivalently,

$$\Phi(\rho) = \mathrm{tr}_\mathcal{K}(V\rho V^\dagger). \tag{24.18}$$

24.1.5 Gel'fand duality

Definition 24.1.33 (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 23.1.3). This space is compact if and only if the algebra is unital.

Definition 24.1.34 (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:

$$\hat{a}(\lambda) = \langle \lambda, a \rangle \tag{24.19}$$

where $\langle \cdot, \cdot \rangle$ denotes the pairing between A and Φ_A. By definition of the topology on 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 24.1.35 (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 on its Gel'fand spectrum.*

Remark 24.1.36. 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 24.1.37. A compact Hausdorff space is connected if and only if its algebra of continuous functions has no nontrivial projections.

Property 24.1.38. A commutative C^*-algebra is separable if and only if its Gel'fand spectrum is metrizable.

Formula 24.1.39 (Spectrum). Recall Definition 23.4.43 of the spectrum of an operator. This is related to the spectrum of a unital, commutative C^*-algebra as follows:

$$\sigma(a) = \{a(x) \mid x \in \Phi_A\}, \tag{24.20}$$

for all $a \in A$.

Construction 24.1.40 (Gel'fand-Naimark-Segal). Let A be a C^*-algebra. Given a state ω on A there exists a C^*-representation $\rho : A \to \mathcal{B}(D)$, 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 distinguished cyclic unit vector ξ such that $D = \{\rho(a)\xi \mid a \in A\}$.
- For all elements $a \in A$ the following equality holds:

$$\omega(a) = \langle \rho(a)\xi|\xi\rangle. \tag{24.21}$$

?? COMPLETE CONSTRUCTION ??

Theorem 24.1.41 (Gel'fand-Naimark: General case). *Every C^*-algebra is isometrically $*$-isomorphic to a norm closed (C^*-)algebra of bounded operators on a Hilbert space $\mathcal{H}$.*

24.1.6 Hilbert modules ♣

Definition 24.1.42 (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

$$\langle \cdot \mid \cdot \rangle_A : H \times H \to A, \tag{24.22}$$

sometimes called the A**-inner product**, satisfying the following conditions:

1. **Conjugate linearity**: For all $x, y, z \in H$:

$$\langle x|y + z\rangle_A = \langle x|y\rangle_A + \langle x|z\rangle_A \qquad \langle x|y\rangle_A = \langle y|x\rangle_A^*. \tag{24.23}$$

2. **Nondegeneracy**: For all $x \in H$:

$$\langle x|x\rangle_A \geq 0 \qquad \langle x|x\rangle_A = 0 \iff x = 0. \tag{24.24}$$

3. **Equivariance**: For all $x, y \in H$ and $a \in A$:

$$\langle x|y\rangle_A a = \langle x|y \cdot a\rangle_A. \tag{24.25}$$

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 24.1.43 (Hilbert C^*-bimodule). Let A, B be C^*-algebras and consider a Hilbert C^*-module H over B. H is a Hilbert (A, B)-bimodule if it admits a left $*$-representation of A such that

$$\langle a^* \cdot x|y\rangle_B = \langle x|a \cdot y\rangle_A \tag{24.26}$$

for all $a \in A$, i.e. adjoints in A correspond to adjoints with respect to the B-inner product.

Example 24.1.44. Consider $A = \mathbb{C}$ and let $\pi : \mathbb{C} \to \mathrm{End}(\mathcal{H})$ be the unique unital representation. Fredholm modules with this data correspond to (essentially self-adjoint) Fredholm operators on $\mathcal{H}$.

Definition 24.1.45 (Fredholm module). Let $\pi : A \to \mathrm{End}(\mathcal{H})$ be a $*$-representation of a unital C^*-algebra on a Hilbert space $\mathcal{H}$. When equipped with an operator $F \in \mathrm{End}(\mathcal{H})$ this gives an (**odd**) Fredholm module if

$$F = F^* \qquad \text{and} \qquad F^2 = \mathbb{1}_{\mathcal{H}} \qquad \text{and} \qquad [F, \pi(A)] \subseteq \mathcal{B}_0(\mathcal{H}). \tag{24.27}$$

The Fredholm operator is said to be **even** if $\mathcal{H}$ is a super-Hilbert space with grading $\Gamma \in \mathrm{End}(\mathcal{H})$ satisyfing

$$\Gamma^* = \Gamma \qquad \text{and} \qquad \Gamma^2 = \mathbb{1}_{\mathcal{H}} \qquad \text{and} \qquad [\Gamma, \pi(A)] = 0 \qquad \text{and} \qquad \{\Gamma, F\}_+ = 0. \tag{24.28}$$

F has the form

$$\begin{pmatrix} 0 & F_+ \\ F_- & 0 \end{pmatrix}, \tag{24.29}$$

where F_+ is Fredholm.

Remark 24.1.46 (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 24.1.47 (Kasparov bimodule). Let A, B be two C^*-algebras and let H be a super-Hilbert (A, B)-bimodule. H is 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, then the bimodule is said to be **normalized**.

Definition 24.1.48 (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 24.1.49. Every Kasparov bimodule is homotopic to a normalized bimodule. Hence, for Kasparov K-theory, one can restrict to normalized bimodules.

24.2 von Neumann algebras

Definition 24.2.1 (von Neumann algebra). A $*$-subalgebra of a C^*-algebra equal to its double commutant: $M = M''$.

Definition 24.2.2 (Concrete von Neumann algebra). A weakly closed unital $*$-algebra of bounded operators on some Hilbert space.

Theorem 24.2.3 (Double Commutant theorem[1]). *The above definitions are equivalent.*

Definition 24.2.4 (Projection). An element p of a von Neumann algebra is called a projection if it satisfies

$$p = p^2 = p^*. \tag{24.30}$$

This terminology reflects the property that if a von Neumann algebra is regarded as an algebra of bounded operators, the projections are exactly the operators associated to an orthogonal projection.

Property 24.2.5. Any von Neumann algebra is generated by its projections.

Definition 24.2.6 (Murray-von Neumann equivalence). Two closed subspaces are said to be Murray-von Neumann equivalent if one is mapped isomorphically onto the other by 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 $q = u^*u$.

Definition 24.2.7 (Finite projection). The collection of projections inherits the structure of a partial order from the partial order on the corresponding subspaces. A projection p is said to be finite if there exists no smaller projection q that is equivalent to p.

Theorem 24.2.8 (Gleason). *Consider a **finitely additive measure**[2] 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*

$$\mu\left(\sum_{i=1}^n p_i\right) = \sum_{i=1}^n \mu(p_i) \tag{24.31}$$

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.

24.2.1 Factors

Definition 24.2.9 (Factor). Consider a von Neumann algebra M. A $*$-subalgebra A is called a factor of M if its center $Z(A)$ is given by the scalar multiples of the idenity.

Definition 24.2.10 (Type-I factor). A factor is of type I if it contains a *minimal projection*.

Property 24.2.11 (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 of this Hilbert space (which may be ∞) one sometimes uses the subclassification of type I_n factors.

Definition 24.2.12 (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.

[1] Often called **von Neumann's double commutant theorem**.

[2] If this also holds for countable sequences, it is called **σ-additive** in analogy with Definition 16.1.1.

Definition 24.2.13 (Type-II factor). A factor is of type II if it 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 of type II_∞.

Definition 24.2.14 (Type-III factor). A factor is of type III if it does not contain any nonzero finite projections.

Chapter 25

Clifford Algebra

The main references for this chapter are [9, 10, 111]. 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 [111] for more information.

25.1 Clifford algebra

Definition 25.1.1 (Clifford algebra). Consider a unital associative 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:

$$a \cdot a = Q(a)1, \tag{25.1}$$

where 1 is the unit element in A. This condition implies that the square of a vector is a scalar.

Notation 25.1.2. The Clifford algebra corresponding to A and Q is often denoted by $C\ell(A, Q)$.

Construction 25.1.3. The previous definition can be given an explicit construction. First, construct the tensor algebra of A:

$$T(A) = \bigoplus_{k\in\mathbb{N}} A^{\otimes k}. \tag{25.2}$$

Then, construct a two-sided ideal I of A generated by $\{a \otimes a - Q(a)1 \mid a \in A\}$ as defined in 3.6.20. The Clifford algebra $C\ell(A, Q)$ can then be constructed as the quotient algebra $T(A)/I$.

Remark 25.1.4. Looking at Definition 21.4.21 it can be seen that the exterior algebra $\Lambda^\bullet(A)$ coincides with the Clifford algebra $C\ell(A, 0)$. If $Q \neq 0$, the two algebras are still isomorphic as vector spaces (if[1] $\mathrm{char}(A) \neq 2$).

Property 25.1.5 (Dimension). If A has dimension n, then $C\ell(A, Q)$ has dimension 2^n.

Example 25.1.6 (Inner product spaces). The classical example of a Clifford algebra is given by an inner product space with Lorentzian signature (p, q), i.e. a vector space with a semidefinite form $g(\cdot, \cdot)$ admitting a basis $\{e_i\}_{i\leq p+q}$ such that

$$\begin{cases} g(e_i, e_i) = 1 & 1 \leq i \leq p \\ g(e_i, e_i) = -1 & p < i \leq p + q. \end{cases} \tag{25.3}$$

[1]This condition will often come back in this chapter.

The Clifford algebra $C\ell_{p,q}(V)$ or $V_{p,q}$ is then defined as the Clifford algebra generated under the relation $v \cdot v = -g(v, v)1$. In physics this convention would correspond to the ‘‘mostly pluses’’-convention, which is mainly adopted in general relativity.

Formula 25.1.7 (Dimensional reduction).

$$\mathbb{R}_{p+1,q+1} \cong \mathbb{R}_{p,q} \otimes M_2(\mathbb{R}) \tag{25.4}$$

Formula 25.1.8.

$$\mathbb{R}_{p+1,q} \cong \mathbb{R}_{q+1,p} \tag{25.5}$$

Formula 25.1.9.

$$\mathbb{R}_{p,q+2} \cong \mathbb{R}_{q,p} \otimes \mathbb{H} \tag{25.6}$$

25.2 Geometric algebra

Definition 25.2.1 (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 25.1.1. This identification is unique as long as $\mathrm{char}(V) \neq 2$.

Definition 25.2.2 (Inner and exterior product). Analogous to the inner product in linear algebra and the wedge product in exterior algebra, one can define a(n) (a)symmetric product on the 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:

$$ab = \frac{1}{2}(ab + ba) + \frac{1}{2}(ab - ba). \tag{25.7}$$

One can then define the inner product as the symmetric part:

$$a \cdot b := \frac{1}{2}(ab + ba) = \frac{1}{2}\big((a + b)^2 - a^2 - b^2\big) = g(a, b). \tag{25.8}$$

Analogously, one can define the exterior (outer) product as the antisymmetric part:

$$a \wedge b := \frac{1}{2}(ab - ba). \tag{25.9}$$

These definitions allow to rewrite Equation (25.7) as follows:

$$ab = a \cdot b + a \wedge b. \tag{25.10}$$

Remark. Looking at the last equality in the definition of the inner product (25.8), it can be seen that condition (25.1) is indeed satisfied when $a = b$.

Definition 25.2.3 (Multivector). Any element of the GA over V is called a multivector. The simple multivectors of grade k, i.e. elements of the form $v_1v_2 \ldots v_k$ with $v_i \in V$ for all i, are called **k-blades**. (This should again remind the reader of the content of Section 21.4.4.) Sums of multivectors of different grades are called **mixed** multivectors (even though these elements do not represent a geometric structure).

Let $n = \dim(V)$. Multivectors of grade n are also called **pseudoscalars** and multivectors of grade $n - 1$ are also called **pseudovectors**.

Definition 25.2.4 (Grade projection operator). Let a be a general multivector. The grade (projection) operator $\langle\cdot\rangle_k : \mathcal{G} \to \mathcal{G}_k$ is defined as the projection of a on the k-vector part of a.

Using these projection operators one can extend the inner and exterior product to the complete GA as follows:

Formula 25.2.5. Let A, B be two multivectors of respectively grades m and n. Their inner product is defined as

$$A \cdot B := \langle AB\rangle_{|m-n|} \tag{25.11}$$

and their exterior product is defined as

$$A \wedge B := \langle AB\rangle_{m+n}. \tag{25.12}$$

An explicit calculation for $A \in \mathcal{G}_1, B \in \mathcal{G}_k$ gives us:

$$A \cdot B = \frac{1}{2}\Big(AB - (-1)^k BA\Big) \tag{25.13}$$

$$A \wedge B = \frac{1}{2}\Big(AB + (-1)^k BA\Big). \tag{25.14}$$

25.3 Bott periodicity ♣

The following theorem has deep implications in K-theory (Chapter 39). It is also (through K-theory) related to the *tenfold way* of *Altland* & *Zirnbauer* in condensed matter physics.

Theorem 25.3.1 (Bott periodicity). *The classification of (real) Clifford algebras is periodic modulo 8:*

$$\mathbb{R}_{p,q+8} \cong \mathbb{R}_{p+8,q} \cong \mathbb{R}_{p,q} \otimes M_{16}(\mathbb{R}). \tag{25.15}$$

For complex Clifford algebras a similar statement exists, but with periodicity 2.

Bott periodicity also has some implications in category theory. Here the language of Section 4.8 will be used.

Definition 25.3.2 (Banach category). An additive category enriched over (real) Banach spaces.

For every Banach category $\mathbf{C}$ and finite-dimensional $\mathbb{R}$-algebra A, the category $\mathbf{C}^A$ is defined as follows:

1. **Objects**: The pairs (X, ρ) where $X \in \mathrm{ob}(\mathbf{C})$ and $\rho : A \to \mathrm{End}(X)$ is an A-representation on X.
2. **Morphism**: The A-equivariant morphisms/intertwiners.

Notation 25.3.3. For brevity some specific notations for the cases $A = \mathbb{R}_{p,q}$ and $A = M_n(\mathbb{R})$ are introduced: $\mathbf{C}^{p,q}$ and $\mathbf{C}(n)$.

Property 25.3.4 (Morita equivalence). Every pseudo-Abelian Banach category $\mathbf{C}$ is equivalent to $\mathbf{C}(n)$ for some $n \in \mathbb{N}$.

Property 25.3.5. For A, B two finite-dimensional $\mathbb{R}$-algebras and $\mathbf{C}$ a pseudo-Abelian Banach category, the following equivalences of categories exist:

$$\mathbf{C}^{A\oplus B} \cong \mathbf{C}^A \times \mathbf{C}^B, \tag{25.16}$$

$$\mathbf{C}^{A\otimes B} \cong \left(\mathbf{C}^A\right)^B. \tag{25.17}$$

Property 25.3.6 (Bott periodicity). For $\mathbf{C}$ a pseudo-Abelian real Banach category, the equivalence classes of categories $\mathbf{C}^{p,q}$ are determined by $p-q$ mod 8.

Construction 25.3.7 (Grothendieck group). Consider a **Banach functor** $\varphi : \mathbf{C} \to \mathbf{C}'$ 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 C' 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 $X, Y \in \mathrm{ob}(\mathbf{C})$ 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 $\mathbf{C}'$ 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 $\mathrm{Aut}(\varphi(X))$. The group $K(\varphi)$ is defined as the quotient of $\mathscr{V}(\varphi)$ by the following equivalence relation:
>
> $$v \sim v' \iff \exists e, e' \in \mathscr{E}(\varphi) : v + e \cong_{\mathscr{V}} v' + e'. \tag{25.18}$$

Definition 25.3.8 ($K^{p,q}$). 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 $p-q$ mod 8.

Example 25.3.9 ($K^{0,0}$). 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 25.3.10 (Complex spaces). The above constructions can also be done in the setting of complex Banach spaces and complex algebras. However, Bott periodicity will then give a $p-q$ mod 2 classification.

25.4 Pin group

25.4.1 Clifford group

Definition 25.4.1 (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:

$$\cdot^t : e_i \otimes e_j \otimes \cdots \otimes e_k \mapsto e_k \otimes \cdots \otimes e_j \otimes e_i. \tag{25.19}$$

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 $C\ell(V)$.

Definition 25.4.2 (Main involution). Let V_0, V_1 be respectively the grade-0 and 1 components of the Clifford algebra $C\ell(V, Q)$. Consider the following operator:

$$\hat{v} = \begin{cases} v & v \in V_0 \\ -v & v \in V_1. \end{cases} \tag{25.20}$$

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 27.1.6.

Definition 25.4.3 (Twisted conjugation). Let $v \in V$ be a vector and let $s \in C\ell(V, Q)$ be a unit of the Clifford algebra over V, i.e. $Q(s) \neq 0$. The twisted conjugation of v by s is defined as the following map:

$$\chi : C\ell(V, Q) \to \mathrm{Aut}(C\ell(V, Q)) \qquad \text{with} \qquad \chi(s)v = sv\hat{s}^{-1}. \tag{25.21}$$

Definition 25.4.4 (Clifford group[2]). The Clifford group $\Gamma(V,Q)$ is defined as follows:

$$\Gamma(V,Q) := \left\{s \in C\ell_{\text{hom}}(V,Q) \,\middle|\, s \text{ is invertible and } v \in V \implies sv\hat{s}^{-1} \in V\right\} \tag{25.22}$$

Because the units of $C\ell(V,Q)$ form a group, $\Gamma(V,Q)$ also forms a group.

Property 25.4.5. When restricting to the units of $C\ell(V)$ that belong to V itself, the twisted conjugation is given by a Householder transformation 20.3.16.

Property 25.4.6. Restrict to the case where V is finite-dimensional and Q is nondegenerate. If one interprets the condition $\chi_s(v) \in V$ as stating the existence of a linear transformation[3] $L \in \text{End}(V)$ such that

$$se_i\hat{s}^{-1} = L_i^j e_j, \tag{25.23}$$

it can be seen that L preserves the norm on V and, accordingly, that the map $s \mapsto L$ defines a surjective homomorphism[4]

$$\rho : \Gamma(V,Q) \to \mathrm{O}(V,Q) : s \mapsto L. \tag{25.24}$$

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 3.1.10 it follows that $\mathrm{O}(V,Q)$ is isomorphic to $\Gamma(V,Q)/\ker\chi$, where $\ker\chi = \mathbb{R}_0$. This isomorphism also implies[5] that the Clifford group is given by the set of finite products of invertible elements $v \in V$:

$$\Gamma(V,Q) = \left\{\prod_i^n s_i \,\middle|\, s_i \text{ invertible in } V, n \in \mathbb{N}\right\}. \tag{25.25}$$

Corollary 25.4.7. By noting that pure rotations can be decomposed as an even number of reflections one finds that

$$\Gamma^+(V,Q)/\mathbb{R}_0 \cong \mathrm{SO}(V,Q), \tag{25.26}$$

where Γ^+ is the intersection of the even Clifford algebra and the Clifford group.

Remark 25.4.8. 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[6]

$$\mathrm{ad}_s : v \mapsto svs^{-1}. \tag{25.27}$$

Although the difference between these maps seems to be rather subtle, the implications are important. If the conjugation ad would have been chosen for the definition of the Clifford group, only a surjective homomorphism would have been found in the case of $\dim(V)$ being odd. Moreover, the action by a degree-1 element would not be given by a Householder transformation anymore, but instead it would be the negative of this operation. This distinction is in particular important for the next section.

[2]Sometimes called the **Lipschitz group**.

[3]Here the isomorphism between the degree-1 subspace of $C\ell(V)$ and V itself is used.

[4]In $\text{char}(K) \neq 2$, the surjectiveness of the map χ follows from the *Cartan-Dieudonné theorem*. For characteristic 2 one can prove that the surjectiveness holds using different methods.

[5]Again using the *Cartan-Dieudonné theorem*, valid only when $\text{char}(K) \neq 2$. In fact this statement is more or less the *Cartan-Dieudonneé* theorem in terms of geometric algebra.

[6]The notation ad_s comes from the fact that this map resembles the adjoint action of a group.

25.4.2 Pin and Spin

Formula 25.4.9 (Spinor norm). On $\Gamma(V, Q)$ one can define the spinor norm[7]

$$\mathcal{N}(x) : \Gamma(V, Q) \to K^\times : x \mapsto x^t x, \tag{25.28}$$

where x^t is the transposition 25.4.1. On V, $\mathcal{N}$ coincides with the norm induced by Q.

Definition 25.4.10 (Pin and spin groups). Using the spinor norm $\mathcal{N}$ one can now define the pin and spins groups as follows:

$$\mathrm{Pin}(V) := \{s \in \Gamma(V, Q) \mid \mathcal{N}(s) = \pm 1\} \tag{25.29}$$

and

$$\mathrm{Spin}(V) := \mathrm{Pin}(V) \cap \Gamma^+(V, Q). \tag{25.30}$$

Remark 25.4.11. In the literature one can sometimes find the following alternative definition of the spinor norm:

$$\mathcal{N}(x) := \hat{x}^t x. \tag{25.31}$$

Alternative Definition 25.4.12. 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 25.4.13. The Pin group satisfies the following isomorphism:

$$\mathrm{Pin}(V, Q)/\mathbb{Z}_2 \cong \mathrm{O}(V, Q). \tag{25.32}$$

An analogous relation holds for the Spin group and $\mathrm{SO}(V, Q)$. These relations imply that the Pin and Spin groups form a double cover 7.2.13 of respectively the orthogonal and special orthogonal groups.

Notation 25.4.14. The Pin-groups associated to $\mathrm{O}(n, 0)$ and $\mathrm{O}(0, n)$ are often denoted by $\mathrm{Pin}^+(n)$ and $\mathrm{Pin}^-(n)$, where the signs refer to the sign of squares in the Clifford condition 25.1.

Definition 25.4.15 (Spinor). Consider a vector space V equipped with a (faithful) representation of the group $\mathrm{Spin}(m, n)$. 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 3.6.39) 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 relation (25.1) 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 $C\ell_{m,n}(\mathbb{R})$.[8] 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]This map can be generalized to the full Clifford algebra, but then the image will not just be the underlying field anymore.

[8]Note, however, that these matrices themselves are still complex-valued in general.

Example 25.4.16. The following table gives some group isomorphisms for the Spin group in dimension n:

n	$\mathrm{Spin}(n)$
1	$\mathrm{O}(1)$
2	$\mathrm{U}(1)$
3	$\mathrm{SU}(2)$
4	$\mathrm{SU}(2) \times \mathrm{SU}(2)$.

For quadratic forms of signature (p, q) the following table is found:

$(1, n)$	$\mathrm{Spin}(1, n)$
$(1, 1)$	$\mathrm{GL}(1, \mathbb{R})$
$(1, 2)$	$\mathrm{SL}(2, \mathbb{R})$
$(1, 3)$	$\mathrm{SL}(2, \mathbb{C})$
$(1, 5)$	$\mathrm{SL}(2, \mathbb{H})$
$(1, 9)$	$\mathrm{SL}(2, \mathbb{O})$.[9]

Formula 25.4.17. Consider the basis of $\mathfrak{su}(2)$ given by the Pauli matrices 55.2.3. An explicit (double) covering map $\rho : \mathrm{Spin}(3) \cong \mathrm{SU}(2) \rightarrow \mathrm{SO}(3)$ is given by:

$$\rho : U \mapsto \frac{1}{2}\,\mathrm{tr}(U\sigma_i U^\dagger \sigma^j). \tag{25.33}$$

Property 25.4.18. For all $m, n \in \mathbb{N}$ the following isomorphism exists:

$$\mathrm{Spin}(m, n) \cong \mathrm{Spin}(n, m). \tag{25.34}$$

Remark 25.4.19 (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 [82]).

Definition 25.4.20 (Semispin group). By definition every Spin group has a canonical $\mathbb{Z}_2$-subgroup. However, for $n \in 4\mathbb{N}_0$, $\mathrm{Spin}(n)$ also contains a noncanonical central $\mathbb{Z}_2$-subgroup $\iota : \mathbb{Z}_2 \hookrightarrow \mathrm{Spin}(n)$. The quotient $\mathrm{Spin}(n)/\iota$ is called the semispin group $\mathrm{SemiSpin}(n)$.

Property 25.4.21. For $n = 8$, the $\mathbb{Z}_2$-subgroups are isomorphic and, as a consequence

$$\mathrm{SemiSpin}(8) \cong \mathrm{SO}(8). \tag{25.35}$$

[9] 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 dimension do not even agree as vector spaces. A suitable definition of $\mathrm{SL}(2, \mathbb{O})$ falls outside the scope of this compendium (see e.g. a paper by *Hitchin*.)

Chapter 26

Noncommutative Algebra

References for this chapter are [26, 61].

26.1 Coalgebras

Dual (in the categorical sense) to the definition of a (unital associative) algebra we have:

Definition 26.1.1 (Coalgebra). A vector space C together with two linear maps $\Delta : C \to C \otimes C$ and $\varepsilon : C \to K$, 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 26.1.2. 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:

$$\Delta(e_i) := e_i \otimes e_i \tag{26.1}$$

and

$$\varepsilon(e_i) := 1. \tag{26.2}$$

By linearity these maps can be extended to all of V. Important cases are the tensor algebra and exterior algebra over a vector space (Definitions 21.3.7 and 21.4.20).

Remark 26.1.3. 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 below).

Definition 26.1.4 (Group-like element). An element c in a coalgebra (C, Δ, ε) that satisfies $\Delta(c) = c \otimes c$ and $\varepsilon(c) = 1$.

Remark. The name "group-like" 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 26.1.5 (Unital coalgebra). A coalgebra (C, Δ, ε) is said to be unital if it comes equipped with a coalgebra morphism $\eta : K \to C$. The element $\eta(1)$ is often also denoted by 1.

Definition 26.1.6 (Primitive element). An element c in a unital coalgebra (C, Δ, ε) that satisfies $\Delta(c) = c \otimes 1 + 1 \otimes c$.

Definition 26.1.7 (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:

$$(\mathbb{1}_A \otimes \nabla) \circ (\Delta \otimes \mathbb{1}_A) = \Delta \circ \nabla = (\nabla \otimes \mathbb{1}_A) \circ (\mathbb{1}_A \otimes \Delta). \tag{26.3}$$

If $\nabla \circ \Delta = \mathbb{1}_A$, the algebra is said to be **special**.

Remark 26.1.8. It is this definition that can be generalized to arbitrary symmetric monoidal categories.

Notation 26.1.9 (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:

$$\Delta(c) = \sum_{(c)} c_{(1)} \otimes c_{(2)} \tag{26.4}$$

or even

$$\Delta(c) = c_{(1)} \otimes c_{(2)}. \tag{26.5}$$

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 26.1.10 (Coideal). Let (C, Δ, ε) be a coalgebra. A subcoalgebra I of C is a (left) coideal if

$$\Delta(I) \subseteq C \otimes I, \tag{26.6}$$

i.e. it is a comodule with respect to the comultiplication.

$$
\begin{tikzcd}
A \otimes A \arrow[r, "\nabla"] \arrow[d, "\Delta \otimes \Delta"'] & A \arrow[r, "\Delta"] & A \otimes A \arrow[d, "\nabla \otimes \nabla"] \\
A \otimes A \otimes A \otimes A \arrow[rr, "\mathbb{1}_A \otimes \sigma_{A,A} \otimes \mathbb{1}_A"'] & & A \otimes A \otimes A \otimes A
\end{tikzcd}
$$

$$
\begin{tikzcd}
& K \arrow[dl, "\eta"'] \arrow[dr, "\eta \otimes \eta"] & \\
A \arrow[rr, "\Delta"'] & & A \otimes A
\end{tikzcd}
\qquad
\begin{tikzcd}
A \otimes A \arrow[rr, "\nabla"] \arrow[dr, "\varepsilon \otimes \varepsilon"'] & & A \arrow[dl, "\eta"] \\
& K &
\end{tikzcd}
\qquad
\begin{tikzcd}
& A \arrow[dr, "\varepsilon"] & \\
K \arrow[ur, "\eta"] \arrow[rr, "\mathbb{1}_K"'] & & K
\end{tikzcd}
$$

Figure 26.1: Bialgebra conditions.

26.2 Hopf algebras

Definition 26.2.1 (Bialgebra). Let A be a vector space over a field 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 Figure 26.1. These diagrams state that ∇, η are coalgebra morphisms and Δ, ε are algebra morphisms.

Definition 26.2.2 (Convolution). Let $(A, \nabla, \eta, \Delta, \varepsilon)$ be a bialgebra. The convolution of two operators $f, g : A \to A$ is defined as follows:

$$f * g := \nabla \circ (f \otimes g) \circ \Delta. \tag{26.7}$$

When equipped with the convolution as multiplication, the space of operators on a bialgebra also becomes an algebra.

Definition 26.2.3 (Hopf algebra). A bialgebra $(A, \nabla, \eta, \Delta, \varepsilon)$ equipped with a linear map $S : A \to A$ that satisfies

$$\nabla \circ (\mathbb{1}_A \otimes S) \circ \Delta = \nabla \circ (S \otimes \mathbb{1}_A) \circ \Delta = \eta \circ \varepsilon \tag{26.8}$$

or, using the convolution on A,

$$\mathbb{1}_A * S = S * \mathbb{1}_A = \eta \circ \varepsilon. \tag{26.9}$$

The map S is called the **antipode** or **coinverse**.

Remark 26.2.4. Some authors require the antipode to be invertible. (In finite dimensions this is always the case as noted below.)

Property 26.2.5. 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.

Property 26.2.6 (Finite-dimensional bialgebras). Any finite-dimensional bialgebra admits an invertible antipode and, in particular, is a Hopf algebra.

[1]Sometimes the notation $\Delta(c) = \Delta_1(c) \otimes \Delta_2(c)$ is used.

Definition 26.2.7 (Quasi-triangular 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 i^{th} and j^{th} position and the unit element $1 \in H$ in the other position, i.e. $(a \otimes b)_{13} = a \otimes 1 \otimes b$.

The element R is often called the **universal R-matrix**[3]. Any Hopf algebra admitting such an element is said to be **quasi-cocommutative**.

Example 26.2.8 (Universal enveloping algebra). Let $\mathfrak{g}$ be a Lie algebra and consider its universal enveloping algebra $U(\mathfrak{g})$ from Section 30.4.6. This algebra admits a Hopf algebra structure with the following operations:

$$\Delta(e_i) := e_i \otimes 1 + 1 \otimes e_i \tag{26.10}$$

$$\varepsilon(e_i) := 0 \tag{26.11}$$

$$S(e_i) := -e_i. \tag{26.12}$$

It becomes quasi-triangular when equipped with the trivial R-matrix $\nVdash_{\mathfrak{g}} \otimes \nVdash_{\mathfrak{g}}$.

Remark 26.2.9 (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**Mod**. Now one would like to define a monoidal structure on A**Mod** induced by the tensor product on A. However, this monoidal structure should be compatible with the action of A.

The intuitive (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:

$$A \otimes (M \otimes_A N) \to M \otimes_A N : a \otimes m \otimes n \mapsto (a_{(1)}m) \otimes (a_{(2)}n). \tag{26.13}$$

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**Mod** will be a monoidal category (this is in fact an equivalence known as **Tannaka duality**).

[2]Sometimes called a **braided Hopf algebra**.

[3]This name is in general used for all elements $R \in H \otimes H$ satisfying the first condition above.

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 map $S : A \to A$ one could define a general action as follows:

$$(af)(v) := f\big(S(a)v\big). \tag{26.14}$$

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 quasi-triangular (this also explains why these Hopf algebras are sometimes said to be braided).

26.2.1 Drinfel'd double

One can easily generalize Definition 3.2.46 (when written in terms of group algebras) to the case of bialgebras by replacing the group comultiplication by a general comultiplication:

Definition 26.2.10 (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:

$$S(a \otimes b) = S_B(b_{(2)}) \cdot S_A(a_{(2)}) \otimes S_B(b_{(1)})^{S_A(a_{(1)})}. \tag{26.15}$$

Example 26.2.11 (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 26.2.12 (Drinfel'd double[4]). Consider a Hopf algebra H (with invertible antipode). It can be shown that H and $(H^{op})^*$ form a matched pair of bialgebras. The left and right actions are induced by pullback:

$$(a \cdot f)(b) = f\big(S^{-1}(a_{(2)})ba_{(1)}\big) \tag{26.16}$$

$$a^f = f\big(S^{-1}(a_{(3)})a_{(1)}\big)a_{(2)} \tag{26.17}$$

The resulting bicrossed product $(H^{op})^* \bowtie H$ is called the Drinfel'd double $D(H)$.

[4] Also known as the **quantum double** (especially in physics).

Example 26.2.13 (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:

$$\Delta(g) = g \otimes g, \tag{26.18}$$
$$\varepsilon(g) = 1. \tag{26.19}$$

On the other hand one can also put a Hopf algebra structure on the dual $\mathbb{C}[G]^*$:

$$\Delta(P_g) = \sum_{hh'=g} P_h \otimes P_{h'}, \tag{26.20}$$
$$\varepsilon(P_g) = \delta_{g,e}, \tag{26.21}$$

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 ??

26.3 Quantum groups

This section heavily builds upon the theory presented in Chapter 30. The content is partially based on talks by *André Henriques*.

Construction 26.3.1 (Jimbo-Drinfeld). Consider a Lie algebra $\mathfrak{g}$ together with its universal enveloping algebra $U(\mathfrak{g})$ constructed using the Chevalley-Serre relations 30.4.57:

1. $[H_i, H_j] = 0$,
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]:

$$K_i := q^{d_iH_i}, \tag{26.22}$$

where $d_i := \frac{\langle\alpha_i,\alpha_i\rangle}{2}$ is related to the norm of the i^{th} 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:

$$f \cdot E_i = E_i\tau_{\alpha_i}(f),$$
$$f \cdot F_i = F_i\tau_{-\alpha_i}(f),$$

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*. $K_iE_j = q^{d_ia_{ij}}E_jK_i$, and

[5]To be complete one should also add generators K_i^{-1} that act formally as inverses of the generators K_i.

3*. $K_iF_j = q^{-d_ia_{ij}}F_jK_i$.

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]:

$$[n]_q := \frac{q^n - q^{-n}}{q - q^{-1}}. \tag{26.23}$$

Using this definition Serre relation 4 becomes

4*. $[E_i, F_j] = \delta_{ij}[H_i]_{q^{d_i}} = \delta_{ij}\frac{K_i - K_i^{-1}}{q^{d_i} - q^{-d_i}}$,

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*. $\sum_{k=1}^{1-a_{ij}}(-1)^k \begin{bmatrix} 1-a_{ij} \\ k \end{bmatrix}_{q^{d_i}} E_i^{1-a_{ij}-k}E_jE_i^k = 0$, and

6*. $\sum_{k=1}^{1-a_{ij}}(-1)^k \begin{bmatrix} 1-a_{ij} \\ k \end{bmatrix}_{q^{d_i}} F_i^{1-a_{ij}-k}F_jF_i^k = 0$.

26.4 Differential structures

26.4.1 Differential calculi

Definition 26.4.1 (First-order differential calculus). Let A be an associative algebra and let Γ be an A-bimodule. Together with an A-bimodule morphism $d : A \rightarrow \Gamma$ this structure is called a first-order differential calculus (FODC) if it satisfies the following two conditions:

1. **Leibniz rule**: $d(ab) = (da)b + a(db)$.
2. **Standard form**: Every element $g \in \Gamma$ can be written as

$$g = \sum_{i=1}^{n} a_i(db_i)$$

for some $n \in \mathbb{N}$ and (not necessarily unique) elements $\{a_i, b_i\}_{i\leq n}$.

If $\ker(d) \cong K$, where 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 $da = [\theta, a]$ for all $a \in A$.

An algebra morphism $\phi : A \rightarrow B$ is said to be **differentiable**, with respect to a choice of FODCs on A and B, if there exists an A-bimodule morphism $\phi_* : \Gamma_A \rightarrow \Gamma_B$ (A inherits an action on Γ_B through ϕ) such that $d \circ \phi = \phi_* \circ d$. If such a morphism exists, it is given by $\phi_*(adb) = \phi(a)d\phi(b)$.

Remark 26.4.2. The second condition can be rewritten in terms of a right action using the Leibniz rule.

Definition 26.4.3 (Cotangent dimension). If Γ is free over A with a basis of cardinality 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

[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.

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 26.4.4 (Universal FODC). Consider an algebra A with multiplication μ. The bimodule $\Omega_{\text{uni}} := \ker(\mu)$, equipped with the operator $da := 1 \otimes a - a \otimes 1$, 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 26.4.5 (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

$$d(ab) = adb + bda. \tag{26.24}$$

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 26.4.6 (Differential calculus). Consider an algebra A. A differential calculus over A is an A-bimodule with the structure of differential graded algebra $(\Gamma^\bullet, d)$ such that $\Gamma^0 = A$. If $\Gamma^\bullet$ is freely generated by A and dA, then $(\Gamma^\bullet, d)$ is often called an **exterior algebra** over A.

Property 26.4.7 (Prolongation). Every FODC (Γ, d) admits a maximal extension (or **prolongation**) to an exterior algebra. In degree k it is obtained by taking

$$\Gamma^k := \underbrace{\Gamma \otimes_A \cdots \otimes_A \Gamma}_{k\text{ times}}, \tag{26.25}$$

and modding out the ideal generated by the following relations:

$$\left\{ \sum_i da_i \otimes db_i + du_i \otimes dv_i = 0 \,\middle|\, \sum_i a_i db_i = \sum_i (du_i) v_i \right\}. \tag{26.26}$$

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.

26.4.2 Covariance

Definition 26.4.8 (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

$$\Phi(adb) = \Delta(a)(\mathbb{1}_H \otimes d)\Delta(b). \tag{26.27}$$

Chapter 27

Higher-dimensional Algebra ♣

The main reference for this chapter is the series of papers carrying the same name by *Baez et al.* [76, 77]. References for the section on Berezin calculus are [10, 103]. For Kapranov-Voevodsky 2-vector spaces the reader is referred to the original paper [40]. The section about higher Lie theory is mainly based on [96]. For fusion and modular categories the main reference is [42].

27.1 Graded vector spaces

Definition 27.1.1 (Graded vector space). A vector space V that can be decomposed as

$$V = \bigoplus_{i\in I} V_i \tag{27.1}$$

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 27.1.2 (Finite type). A graded vector space is said to be of finite type if it is finite-dimensional in each degree.

Example 27.1.3 (Super vector space). A $\mathbb{Z}_2$-graded vector space.

Definition 27.1.4 (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 $k, l \in \mathbb{Z}$.[1]

Definition 27.1.5 (Graded-commutative algebra). A graded algebra $(V, \star)$ such that

$$v \star w = (-1)^{\deg(v)\deg(w)} w \star v \tag{27.2}$$

holds for all homogeneous elements $v, w \in V$.

Example 27.1.6 (Superalgebra). A $\mathbb{Z}_2$-graded algebra

$$A = A_0 \oplus A_1, \tag{27.3}$$

such that for all i, j:

$$A_i \star A_j \subseteq A_{i+j \bmod 2}. \tag{27.4}$$

[1]The grading can be relaxed to any commutative monoid.

Definition 27.1.7 (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:

$$(\Pi V)_0 := V_1, \tag{27.5}$$
$$(\Pi V)_1 := V_0. \tag{27.6}$$

A more general construction exists in $\mathbb{Z}$-**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):

$$V[k]_i := V_{i-k}. \tag{27.7}$$

Example 27.1.8 (Free GCA). Let V be a graded vector space. The free GCA $\mathrm{Sym}^\bullet V$ on V is defined as the quotient of the tensor algebra $T(V)$ by the relations

$$x \otimes y - (-1)^{\deg(x)\deg(y)} y \otimes x \tag{27.8}$$

ranging over all homogeneous elements $x, y \in V$. (The notation stems from the fact that it is inherited from the symmetric monoidal structure on $\mathbf{Ch}_\bullet(\mathbf{Vect})$.) This algebra can equivalently be obtained as the tensor product

$$\mathrm{Sym}^\bullet V = \mathrm{Sym}(V_{\text{even}}) \otimes \mathrm{Alt}(V_{\text{odd}}), \tag{27.9}$$

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 (for this reason it is sometimes also denoted by $\mathrm{Sym}^\bullet V$). A similar definition gives a graded alternating algebra:

$$\mathrm{Alt}^\bullet V = T(V)/(x \otimes y - (-1)^{\deg(x)\deg(y)} y \otimes x). \tag{27.10}$$

Note that both of these algebras actually carry a bigrading, the total degree coming from V and the **word length**:

$$\deg(v_1 \cdots v_n) := \deg(v_1) + \cdots + \deg(v_n) \tag{27.11}$$
$$\mathrm{wl}(v_1 \cdots v_n) := n. \tag{27.12}$$

In general, only the word length is made explicit when writing down the space, i.e. $v \in \mathrm{Alt}^{\mathrm{wl}(v)} V$.

Remark 27.1.9 (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.[2] 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 the Grassmann algebra over an ordinary vector space V to the free GCA on the shifted space $V[1]$, i.e. $\mathrm{Alt}^\bullet V \cong \mathrm{Sym}^\bullet V[1]$. However, at this point, the **décalage isomorphism**

$$\mathrm{dec}_k : \Lambda^k V[k] \cong \mathrm{Sym}^k V[1] \tag{27.13}$$

is only a linear isomorphism. There are two ways to see that it can be extended to an algebra isomorphism.

[2]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 first one defines the suspension functor as an intertwiner between the symmetrization and antisymmetrization operations to define Sym and Alt. Define the symmetric and antisymmetric Koszul signs of a permutation $\sigma \in S_n$ as follows:

$$\varepsilon(\sigma; v_1, \ldots, v_n) := (-1)^{\#\text{ odd-odd neighbour transpositions in }\sigma} \tag{27.14}$$
$$\chi(\sigma; v_1, \ldots, v_n) := \mathrm{sgn}(\sigma)\varepsilon(\sigma; v_1, \ldots, v_n). \tag{27.15}$$

Décalage then says that the suspension functor should satisfy

$$\varepsilon(\sigma; v_1, \ldots, v_n)\sigma \circ [1]^{\otimes n} = [1]^{\otimes n} \circ \chi(\sigma; v_1, \ldots, v_n)\sigma. \tag{27.16}$$

Since both Sym and Alt can be defined in terms of the projectors

$$p_{\mathrm{Sym}} := \sum_{\sigma \in S_n} \varepsilon(\sigma)\sigma \qquad \text{and} \qquad p_{\mathrm{Alt}} := \sum_{\sigma \in S_n} \chi(\sigma)\sigma, \tag{27.17}$$

décalage interchanges symmetric and antisymmetric tensors. The most common choice is the following one:

$$[1] : V^{\otimes n} \to V[1]^{\otimes n} : v_1 \otimes \cdot \otimes v_n \mapsto (-1)^{\sum_{i=1}^n (n-i)\deg(v_i)} v_1[1] \otimes \cdots \otimes v_n[1]. \tag{27.18}$$

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):

$$[1] : \mathbb{Z}\text{-}\mathbf{Vect}_k \to \mathbb{Z}\text{-}\mathbf{Vect}_k : V \mapsto k[1] \otimes V. \tag{27.19}$$

This definition also directly induces an algebra isomorphism in the following way. Consider two homogeneous elements $v, w \in V$. In $\mathrm{Sym}^2 V$ their product satisfies

$$vw = (-1)^{\deg(v)\deg(w)} wv.$$

After applying the suspension functor, the product on the left-hand side becomes:

$$v[1]w[1] \equiv (\underline{1} \otimes v)(\underline{1} \otimes w) \cong (-1)^{\deg(w)} \underline{\underline{1}} \otimes (vw) \equiv (-1)^{\deg(w)} vw[2].$$

To calculate the suspension of the right-hand side, the braiding in $\mathbb{Z}\text{-}\mathbf{Vect}_k$ is used:

$$\begin{aligned} v[1]w[1] \equiv (v \otimes \underline{1})(w \otimes \underline{1}) &\mapsto (-1)^{\deg(v)+\deg(w)+\deg(v)\deg(w)+1}(w \otimes \underline{1})(v \otimes \underline{1}) \\ &= (-1)^{\deg(w)+\deg(v)\deg(w)+1}(wv) \otimes \underline{\underline{1}} \\ &\equiv (-1)^{\deg(w)+\deg(v)\deg(w)+1} wv[2]. \end{aligned}$$

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 $\mathrm{Sym}^2 V$. The most thorough review of these issues was found in [108].

27.1.1 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 27.1.10 (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 27.1.11 (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:

$$\mathrm{str}(X) := \mathrm{tr}(A) - \mathrm{tr}(D). \tag{27.20}$$

Property 27.1.12. 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:

$$\mathrm{str}(XY) = (-1)^{\deg(X)\deg(Y)}\mathrm{str}(YX). \tag{27.21}$$

Definition 27.1.13 (Berezinian). The Berezinian or **superdeterminant** generalizes the determinant of an ordinary matrix. It is (uniquely) defined through the following two conditions:

1. $\mathrm{Ber}(XY) = \mathrm{Ber}(X)\mathrm{Ber}(Y)$, and
2. $\mathrm{Ber}(e^X) = e^{\mathrm{str}(X)}$.

An explicit formula is given by

$$\mathrm{Ber}(X) = \det(A - BD^{-1}C)\det(D)^{-1} = \det(A)\det(D - CA^{-1}B)^{-1}, \tag{27.22}$$

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.

27.1.2 Berezin calculus

This section is an application of the concept of exterior algebras 21.4.20. Grassmann numbers/variables are used in quantum field theory when performing calculations in e.g. the fermionic sector or *Faddeev-Popov quantization.*

Definition 27.1.14 (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 ommitted when writing the product:

$$\theta_i \wedge \theta_j \equiv \theta_i\theta_j.$$

Remark 27.1.15. From the (anti)commutativity it follows that one can regard the Grassmann variables as being nonzero square roots of zero.

Notation 27.1.16 (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 27.1.17 (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

$$p(\theta) = a + b\theta. \tag{27.23}$$

Definition 27.1.18. One can equip the exterior algebra Λ with Grassmann variables θ_i with an involution:

$$(\theta_i\theta_j \ldots \theta_k)^* := \theta_k \ldots \theta_j\theta_i. \tag{27.24}$$

Elements $z \in \Lambda$ that satisfy $z^* = z$ are said to be **(super)real** and elements thar 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 27.1.19 (Derivative of Grassmann variables). Consider the polynomial algebra $\mathbb{C}[\theta_1, \ldots, \theta_n]$ 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:

$$\frac{\partial}{\partial\theta_j}\theta_i = \delta_i^j \qquad\qquad \theta_i\frac{\partial}{\partial\theta_j} + \frac{\partial}{\partial\theta_j}\theta_i = 0 \tag{27.25}$$

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

$$\theta_i\frac{\partial^R}{\partial\theta^j} = \delta_i^j. \tag{27.26}$$

The left and right derivatives are also sometimes denoted by

$$\frac{\overrightarrow{\partial}}{\partial\theta^i} \qquad \text{and} \qquad \frac{\overleftarrow{\partial}}{\partial\theta^i},$$

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 27.1.20 (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 27.1.21. 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 27.1.22 (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:[3]

$$\int_B (a + b\theta)d\theta := b. \tag{27.27}$$

This means that the integral is equal to the coefficient of the highest-degree term. As a simple generalization, define

$$\int_B f(\theta_1, \ldots, \theta_n)d\theta := \text{coefficient of } \theta_1 \cdots \theta_n. \tag{27.28}$$

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 27.1.23. 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 27.1.24 (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:

$$\int_B f(\xi)d\xi = \int_B f(\theta)(\det J)^{-1}d\theta. \tag{27.29}$$

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 27.1.13.

[3]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.

27.2 Monoidal categories II: Duality

The general theory of monoidal categories was introduced in Section 4.6.

Definition 27.2.1 (Dual object). Let $(\mathbf{C}, \otimes, \mathbf{1})$ be a monoidal category and consider an object $x \in \mathrm{ob}(\mathbf{C})$. A left dual[4] of x is an object in $x^* \in \mathrm{ob}(\mathbf{C})$ 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 Figure 27.1 commute. x is said to be **dualizable** if the object x^* and the morphisms η, ε exist.

$$\begin{tikzcd}
 & x & \\
\mathbf{1} \otimes x \arrow[ur, "\lambda_x"] \arrow[d, "\eta \otimes \mathbb{1}_x"'] & & x \otimes \mathbf{1} \arrow[ul, "\rho_x"'] \\
(x \otimes x^*) \otimes x \arrow[rr, "\alpha_{x,x^*,x}"] & & x \otimes (x^* \otimes x) \arrow[u, "\mathbb{1}_x \otimes \varepsilon"']
\end{tikzcd}$$

(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 27.1: Dualizable objects.

Definition 27.2.2 (Rigid category). A monoidal category that admits 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 27.2.3 (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 27.2.4 (Compact closed category). A symmetric rigid category.

Example 27.2.5 (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":

$$\eta : \mathbf{1} \to V \otimes V^* : 1 \mapsto \sum_{i=1}^{\dim(V)} e_i \otimes \phi^i, \tag{27.30}$$

where $\{e_i\}$ and $\{\phi^i\}$ are bases of V and V^*, respectively.

It should be noted that the category **Vect** of all vector spaces is not rigid. By Property 27.2.3 above, left and right duals coincide in any braided monoidal category (such as **Vect**). However for infinite-dimensional vector spaces it is known that $A \cong (A^*)^*$ never holds and as such rigidity cannot be extended to **Vect**.

Property 27.2.6 (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 $\mathcal{V}$:[6]

$$\int^{V \in \mathcal{V}} V^* \otimes V \cong K. \tag{27.31}$$

[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.

[6] 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:

$$\int^{V\in\mathcal{V}} \mathcal{V}(V,-)\otimes \mathrm{id}_{\mathcal{V}}V \cong \mathrm{id}_{\mathcal{V}}. \tag{27.32}$$

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 27.2.7. This property can also be generalized by replacing $\mathcal{V}$ by a category of modules A**Mod** for some finite-dimensional algebra A. The end and coend give respectively the algebra A and its dual A^*.

The trace on **FinVect** can be generalized as follows:

Definition 27.2.8 (Trace). Let $(\mathbf{C},\otimes,\mathbf{1})$ be a rigid category and let $f \in \mathbf{C}(x,x^{**})$. The left (**categorical** or **quantum**) trace of f is defined as the following morphism in $\mathrm{End}_{\mathbf{C}}(\mathbf{1})$:

$$\mathrm{tr}^L(f) := \varepsilon_{x^*}\circ(f\otimes \mathbb{1}_{x^*})\circ\eta_x. \tag{27.33}$$

For $f \in \mathbf{C}(x, {}^{**}x)$, the right trace is defined similarly:

$$\mathrm{tr}^R(f) := \varepsilon_{{}^{**}x}\circ(\mathbb{1}_x\otimes f)\circ\eta_{{}^*x}. \tag{27.34}$$

Property 27.2.9. The following linear algebra-like properties hold for the categorical trace:

- $\mathrm{tr}^L(f) = \mathrm{tr}^R(f^*)$,
- $\mathrm{tr}^L(f\otimes g) = \mathrm{tr}^L(f)\,\mathrm{tr}^L(g)$, and
- for additive categories: $\mathrm{tr}^L(f\oplus g) = \mathrm{tr}^L(f) + \mathrm{tr}^L(g)$.

The second and third property can be stated analogously for the right trace.

Definition 27.2.10 (Pivotal category). Let $\mathbf{C}$ be a rigid monoidal category. A pivotal structure on $\mathbf{C}$ is a monoidal natural isomorphism $a_x : x \cong x^{**}$.

Definition 27.2.11 (Dimension). Let $(\mathbf{C},a)$ be a pivotal category and consider an object $x\in\mathrm{ob}(\mathbf{C})$. The dimension of x is defined as follows:

$$\dim_a(x) := \mathrm{tr}^L(a_x). \tag{27.35}$$

Definition 27.2.12 (Spherical category). Let $(\mathbf{C},a)$ be a pivotal category. If the left and right traces with respect to a coincide in $\mathbf{C}$, i.e. $\dim_a(x) = \dim_a(x^*)$, the pivotal structure is said to be spherical.

Definition 27.2.13 (Symmetric monoidal dagger category). A symmetric monoidal category $(\mathbf{C},\otimes,\mathbf{1})$ that also carries the structure of a dagger category 4.3.1 such that

$$(f\otimes g)^\dagger = f^\dagger\otimes g^\dagger \tag{27.36}$$

and such that the coherence and braiding morphisms are unitary.

Definition 27.2.14 (Dagger-compact category). A symmetric monoidal dagger category that is also a compact closed category such that the following diagram commutes for all objects:

$$\begin{tikzcd}
 & \mathbf{1} \arrow[dl, "\eta"'] \arrow[dr, "\varepsilon^\dagger"] & \\
x^*\otimes x & & x\otimes x^* \arrow[ll, "\sigma_{x,x^*}"]
\end{tikzcd}$$

Definition 27.2.15 (Calabi-Yau category). A **Vect**-enriched category $\mathbf{C}$ equipped with a **trace** functional

$$\mathrm{tr}_c : C(c,c) \to K \tag{27.37}$$

for each object $c \in \mathrm{ob}(\mathbf{C})$ such that the induced pairing

$$\langle\cdot,\cdot\rangle : C(c,d) \otimes C(d,c) \to K : f \otimes g \mapsto \mathrm{tr}_c(g \circ f) \tag{27.38}$$

is symmetric and nondegenerate.

Example 27.2.16. A one-object Calabi-Yau category is the (pointed) monoid delooping of a Frobenius algebra 20.7.6.

27.3 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 (often this will be $\mathbb{C}$).

Definition 27.3.1 (Tensor category). A monoidal category with the following properties:

1. it is rigid,
2. it is Abelian,
3. it is K-linear (and it is so in a way compatible with the Abelian structure),
4. $\mathrm{End}(\mathbf{1}) \cong K$, and
5. $-\otimes-$ is bilinear on morphisms.

Some authors (such as [42]) also add "locally finite" as a condition (Definition 4.8.27).

Remark 27.3.2. If K is not algebraically closed, one should exchange the last condition by the condition that $\mathbf{1}$ is a simple object. However, if K is algebraically closed, these statements are equivalent.

Definition 27.3.3 (Pointed tensor category). A tensor category where all of the simple objects are (weakly) invertible.

Definition 27.3.4 (Fusion category). A semisimple finite tensor category.

Property 27.3.5. Let $\mathbf{C}$ be a fusion category. There exists a natural isomorphism $\mathbb{1}_{\mathbf{C}} \cong **$.

Remark 27.3.6. 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 27.3.7 (Categorical dimension). Consider a fusion category $\mathbf{C}$ and choose a natural isomorphism $a : \mathbb{1}_{\mathbf{C}} \cong **$. For every simple object x one can define a dimension function, sometimes called the **norm squared**, in the following way:

$$|x|^2 := \mathrm{tr}(a_x)\,\mathrm{tr}((a_x^{-1})^*). \tag{27.39}$$

If $\mathbf{C}$ is pivotal, this becomes $|x|^2 = \dim_a(x)\dim_a(x^*)$. In particular, when $\mathbf{C}$ is spherical, this becomes $|x|^2 = \dim_a(x)^2$.

The categorical dimension, sometimes called the **Müger dimension**, is then defined as follows:

$$\dim(\mathbf{C}) := \sum_{x\in\mathcal{O}(\mathbf{C})} |x|^2, \tag{27.40}$$

where $\mathcal{O}(\mathbf{C})$ denotes the set of isomorphism classes of simple objects.

Remark 27.3.8. It should be noted that the above quantities do not depend on the choice of isomorphism $a_x : x \cong x^{**}$ since all of them only differ by a scale factor.

Property 27.3.9 (Nonzero dimension). For any fusion category $\mathbf{C}$ one has that $\dim(\mathbf{C}) \neq 0$. In particular, if $K = \mathbb{C}$, then $\dim(\mathbf{C}) \geq 1$ (since the norm squared of any simple object is then also positive).

Definition 27.3.10 (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:

$$\mathbf{C} \cong \bigoplus_{g\in G} \mathbf{C}_g, \tag{27.41}$$

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 $\mathbf{C}_g \otimes \mathbf{C}_h \subset \mathbf{C}_{gh}$ for all $g, h \in G$.

Example 27.3.11 (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 27.3.12. 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 27.3.13 (Tannaka duality). *The category of modules of a weak Hopf algebra has the structure of a fusion category. Conversely, any fusion category can be obtained as the category modules of a weak Hopf algebra.*

27.4 Ribbon and modular categories

Definition 27.4.1 (Ribbon structure). Consider a braided monoidal category $(\mathbf{C}, \otimes, \mathbf{1})$ with braiding σ. A **twist** or **balancing** is a natural transformation θ such that the following equation is satisfied for all $x, y \in \mathrm{ob}(\mathbf{C})$:

$$\theta_{x\otimes y} = (\theta_x \otimes \theta_y) \circ \sigma_{y,x} \circ \sigma_{x,y}. \tag{27.42}$$

If in addition $\mathbf{C}$ is rigid and the twist satisfies $\theta_{x^*} = (\theta_x)^*$ for all $x \in \mathrm{ob}(\mathbf{C})$, one speaks of a ribbon category.

Definition 27.4.2 (Drinfel'd morphism). Let $(\mathbf{C}, \otimes, \mathbf{1})$ be a rigid braided monoidal category with braiding σ. This structure admits a canonical natural isomorphism $x \cong x^{**}$ defined as follows:

$$x \xrightarrow{\mathbb{1}_x\otimes\eta_{x^*}} x \otimes x^* \otimes x^{**} \xrightarrow{\sigma_{x,x^*}\otimes\mathbb{1}_{x^{**}}} x^* \otimes x \otimes x^{**} \xrightarrow{\varepsilon_x\otimes\mathbb{1}_{x^{**}}} x^{**}. \tag{27.43}$$

Property 27.4.3. Let $\mathbf{C}$ be a braided monoidal category. Consider the canonical natural isomorphism $u_x : x \cong x^{**}$ defined above. Any natural isomorphism $\psi_x : x \cong x^{**}$ can be written as $u_x\theta_x$ where $\theta \in \mathrm{Aut}(\mathbb{1}_{\mathbf{C}})$. It is not hard to see that this natural isomorphism is monoidal (hence pivotal) exactly when θ is a twist. If $\mathbf{C}$ is a fusion category then the pivotal structure is spherical if and only if θ gives a ribbon structure.

Definition 27.4.4 (Premodular category). A ribbon fusion category. Equivalently, a spherical braided fusion category.

Definition 27.4.5 (S-matrix). Given a premodular category $\mathbf{M}$ (with braiding σ) one defines the S-matrix as follows:

$$S_{x,y} := \mathrm{tr}(\sigma_{y,x} \circ \sigma_{x,y}), \tag{27.44}$$

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 27.4.6 (Modular category[7]). A premodular category for which the S-matrix is invertible.

Property 27.4.7. Let $\mathbf{M}$ be a modular category with S-matrix S and define the following matrix:

$$E_{x,y} := \begin{cases} 1 & x = y^* \\ 0 & \text{otherwise.} \end{cases} \tag{27.45}$$

The following relation with the categorical dimension of $\mathbf{M}$ is obtained:

$$S^2 = \dim(\mathbf{M})E. \tag{27.46}$$

Formula 27.4.8 (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 dimension associated to the spherical structure on $\mathbf{M}$. Using the formula

$$S_{x,y}S_{x,z} = \dim(x) \sum_{w \in \mathcal{O}(\mathbf{M})} N^w_{y,z} S_{x,w} \tag{27.47}$$

for all $x, y, z \in \mathcal{O}(\mathbf{M})$, one obtains the following important relation:

$$\sum_{w \in \mathcal{O}(\mathbf{M})} \frac{S_{w,y}S_{w,z}S_{w,x^*}}{\dim(w)} = \dim(\mathbf{M})N^x_{y,z}. \tag{27.48}$$

This property implies that the S-matrix of a modular category determines the fusion coefficients of the underlying fusion category.

27.5 Module categories

By categorifying the definition of a module over a ring 3.6.29, one obtains the notion of a module category:

Definition 27.5.1 (Module category). Let $\mathbf{M}$ be a monoidal category. A left $\mathbf{M}$-module (category) is a linear 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 27.5.2. Similar to how a G-set can be defined as a functor $\mathbf{B}G \to \mathbf{Set}$ (Property 4.10.3), one can define a module category as a 2-functor $\mathbf{BM} \to \mathbf{Cat}$.

Analogous to Definition 4.6.19 one can also define internal homs for module categories:

[7] "Modular tensor category" is often abbreviated as **MTC**.

Definition 27.5.3 (Internal hom). Consider a left $\mathbf{M}$-module $\mathbf{C}$. Given two objects $x, y \in \mathrm{ob}(\mathbf{C})$ one defines their internal hom (if it exists) as the object $\underline{\mathrm{Hom}}(x, y) \in \mathrm{ob}(\mathbf{M})$ satisfying the following condition

$$\mathbf{C}(m \triangleright x, y) \cong \mathbf{M}(m, \underline{\mathrm{Hom}}(x, y)) \tag{27.49}$$

for all $m \in \mathrm{ob}(\mathbf{M})$.

Property 27.5.4. 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 4.6.19 as a particular case.

27.6 Higher vector spaces

27.6.1 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 ground field K. Coincidentally this ground 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 27.6.1 (2-vector space). To define 2-vector spaces, *Kapranov* and *Voevodsky* lifted these observations to categories by replacing the ground 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**: **2-matrices**, i.e. collections $\|A_{ij}\|$ of finite-dimensional K-vector spaces.
- **2-morphisms**: Collections (f_{ij}) of linear maps between finite-dimensional K-vector spaces.

The multiplication (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 27.6.2. A 2-vector space is a lax module category over the ring category **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.

27.6.2 Baez-Crans 2-vector spaces

Definition 27.6.3 (2-vector space). A category internal to **Vect**. The morphism are **linear functor**, i.e. functors internal to **Vect**.

Remark 27.6.4. The above definition should not be confused with that of categories and functors enriched over **Vect**.

Example 27.6.5 (Ground field). The ground 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 27.6.6 (Chain complexes). There exists an equivalence between the (2-)categories of 2-vector spaces and 2-term chain complexes.

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 27.6.7. The equivalence (on the level of ordinary categories) is an instance of the Dold-Kan correspondence 12.1.17.

Definition 27.6.8 (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:

$$\vec{f} := f - e(s(f)) \tag{27.50}$$

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:

$$g \circ f = \big(s(f), \vec{f} + \vec{g}\big). \tag{27.51}$$

Definition 27.6.9 (Antisymmetric natural morphism). A natural morphism between n-linear functors in **2Vect** is said to be **completely antisymmetric** if its arrow part is completely antisymmetric.

27.7 Higher Lie theory

27.7.1 Lie superalgebras

Definition 27.7.1 (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 27.7.2 (Lie superalgebra). When using the braiding

$$\sigma(x \otimes y) = (-1)^{\deg(x)\deg(y)} y \otimes x \tag{27.52}$$

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

$$\sigma(x \otimes y) = (-1)^{(\deg(x)-k)(\deg(y)-k)} y \otimes x. \tag{27.53}$$

It is simply a Lie bracket on the k-suspension $\Pi^k V$.

Example 27.7.3 (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 30.2.3:

Definition 27.7.4 (Gerstenhaber algebra). A graded-commutative algebra equipped with a degree-1 Lie bracket that acts as a graded derivation:

$$[x, yz] = [x, y]z + (-1)^{\deg(x)(\deg(y)-1)}y[x, z]. \tag{27.54}$$

Definition 27.7.5 (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

  $$J_{x,y,z} : [[x, y], z] \to [x, [y, z]] + [[x, z], y], \tag{27.55}$$

  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 27.7.6 (Lie 2-algebra). Consider a 2-term chain complex in the category **FinVect**:

$$0 \longrightarrow L_1 \longrightarrow L_0 \longrightarrow 0. \tag{27.56}$$

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

  $$J : [\cdot, [\cdot,\cdot]] \Rightarrow [[\cdot,\cdot],\cdot] + [\cdot, [\cdot,\cdot]] \circ (\sigma \otimes \mathbb{1}), \tag{27.57}$$

  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)!

27.7.2 Lie n-algebras

Definition 27.7.7 (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 27.7.8 (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

$$l_n(v_{\sigma(1)}\dots v_{\sigma(n)}) = \chi(\sigma; v_1,\dots,v_n)l_n(v_1\dots v_n) \tag{27.58}$$

and

$$\sum_{\substack{i+j=n+1\\ \sigma\in\mathrm{Unshuff}(i,j-1)}} (-1)^{i(j-1)}\chi(\sigma;v_1,\dots,v_n)l_i\big(l_j(v_{\sigma(1)}\cdots v_{\sigma(j)})v_{\sigma(j+1)}\cdots v_{\sigma(n)}\big) = 0, \tag{27.59}$$

where Unshuff denotes the collection of unshuffles 3.2.33.

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 27.7.9. The definition can be rephrased in terms of graded maps $\hat{l}_n : \mathrm{Alt}^\bullet V \to V$.

Remark 27.7.10 (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

$$l_1\circ l_1 = l_2(l_0,-). \tag{27.60}$$

This terminology stems from the situation where l_1 is identified with the exterior covariant derivative on an associated vector bundle (see Formula 33.4.18).

Example 27.7.11 (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 27.7.12. 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 (Section 30.6).

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 27.7.13. 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

$$H^3(\mathfrak{g};\mathbb{R}) \cong \mathbb{R}. \tag{27.61}$$

The different classes can be represented by scalar multiples of the Killing cocycle (see Example 30.6.11). For every such scalar $\lambda\in\mathbb{R}$, one denotes the resulting Lie 2-algebra by $\mathfrak{g}_\lambda$.

[8] Also called a **strong(ly) homotopy Lie algebra** (abbreviated to **sh Lie algebra**).

Lie algebras and L_∞-algebras can also be dually characterized in terms of their Chevalley-Eilenberg algebra (see Definition 30.6.1):

Alternative Definition 27.7.14 (Lie algebra). Consider a finite-dimensional Lie algebra $\mathfrak{g}$. The transpose/dual of the Lie bracket $[\cdot,\cdot]:\mathfrak{g}\wedge\mathfrak{g}\rightarrow\mathfrak{g}$ is a morphism $\delta:\mathfrak{g}^*\rightarrow\mathfrak{g}^*\wedge\mathfrak{g}^*$:

$$\delta\omega(g,h) := \omega([g,h]). \tag{27.62}$$

In fact, it is not hard to see that this is exactly the Chevalley-Eilenberg differential of $\mathrm{CE}(\mathfrak{g})$. Conversely, given a semifree dgca $(\mathrm{Alt}^\bullet V^*, d)$, 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 $d^2=0$ 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 27.7.15 (L_∞-algebra). The (graded) Leibniz rule implies that the differential δ is completely defined by its restriction to the generators $V^*\leq\mathrm{Alt}^\bullet V^*$. The differential can be decomposed as follows:

$$\delta t^a := -\sum_{k=1}^{\infty}\frac{1}{k!}[t_{a_1},\ldots,t_{a_k}]_k^a\, t^{a_1}\wedge\cdots\wedge t^{a_k}, \tag{27.63}$$

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 $[\cdots]_k:\mathrm{Alt}^k V\rightarrow V$ of degree $n-1$ (some sources rephrase these brackets as morphism from the symmetric algebra $\mathrm{Sym}^\bullet V$, in which case their degree is just -1, cf. décalage 27.1.9).

The nilpotency condition $\delta^2=0$ implies a list of (quadratic) relations on the brackets $[\cdots]_k$ (with $d:=[\cdot]_1$):

$$\begin{aligned}
d^2 &= 0\\
d[\cdot,\cdot]_2 &= [d\cdot,\cdot]_2+[\cdot,d\cdot]_2\\
[[v_1,v_2],v_3]_2+\text{cyc. perm.} &= d[v_1,v_2,v_3]_3-[dv_1,v_2,v_3]_3-[v_1,dv_2,v_3]_3-[v_1,v_2,dv_3]_3\\
&\ \ \vdots
\end{aligned}$$

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 27.7.16 (Maurer-Cartan element). An element a of an L_∞-algebra V that satisfies the equation

$$\sum_{k=0}^{\infty}\frac{1}{k!}[a,\ldots,a]_k = 0. \tag{27.64}$$

For dg-Lie algebras this reduces to the ordinary Maurer-Cartan equation (see 33.3.23):

$$da+\frac{1}{2}[a,a]=0. \tag{27.65}$$

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.

27.8 Monoidal n-categories

Definition 27.8.1 (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 4.9.8. For the explicit definitions of monoidal bi- and tricategories, see the papers [91] and [90] respectively.

If one would put multiple compatible monoidal products on an n-category, by a version of the Eckmann-Hilton argument 4.5.1 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 27.8.2 (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 $\boldsymbol{k}\mathbf{Mon}\boldsymbol{n}\mathbf{Cat}$.

Example 27.8.3. 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
  - $k \geq 3$: symmetric monoidal category.

The stabilization occurring for higher values of k is the content of the following hypothesis[10] by *Baez & Dolan*:

Theorem 27.8.4 (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* $\boldsymbol{k}\mathbf{Mon}\boldsymbol{n}\mathbf{Cat} \hookrightarrow \boldsymbol{(n+2)}\mathbf{Mon}\boldsymbol{n}\mathbf{Cat}$ *becomes an equivalence.*

27.8.1 Relation with group cohomology

See Definition 3.4.1 or Section 5.4.2 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

$$g * h := e^{i\omega(g,h)}gh. \tag{27.66}$$

[9] As in category with a specified element not as in category with a zero object 4.4.14.

[10] For certain definitions of higher categories this has been proven in full generality.

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 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 $H^4(G; \mathrm{U}(1))$.

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 27.8.5. This section is strongly related to the twisting procedure in n-dimensional Dijkgraaf-Witten theories.

[11] These can be parametrized by the *Stasheff polytopes/associahedra*.

Part V

Differential Geometry

Chapter 28

Classic Differential Geomtry

28.1 Curves

Definition 28.1.1 (Regular curve). Let $c : I \to \mathbb{R}^n$ be a curve on the interval I. It is said to be regular at $t \in I$ if $\frac{dc}{dt} \neq 0$. (This is generalized in 29.3.7.)

Definition 28.1.2 (Geometric property). A property that is invariant under:

1. parameter transformations, and
2. orientation-preserving basis transformations.

Property 28.1.3. Let $c, d : I \to \mathbb{R}^n$ be two curves with the same image. The following relation holds:

$$c(t) \text{ is regular } \iff d(t) \text{ is regular.} \tag{28.1}$$

28.1.1 Arc length

Definition 28.1.4 (Natural parameter). Let $c : I \to \mathbb{R}^n$ be a curve. The parameter t is said to be a natural parameter if

$$\left\| \frac{dc}{dt} \right\| = 1 \tag{28.2}$$

for all values of $t \in I$.

Formula 28.1.5 (Arc length). The following function is a bijective map and gives a natural parameter for the curve c:

$$\phi(t) := \int_{t_0}^{t} \|\dot{c}(t)\| dt. \tag{28.3}$$

This natural parameter is often denoted by the letter s.

Property 28.1.6. 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

$$u = \pm s + \alpha,$$

where s denotes the arc length.

Remark 28.1.7. This property implies that there does not exist a unique natural parameter for any curve.

28.1.2 Frenet-Serret frame

Definition 28.1.8 (Tangent vector). Let $c(s)$ be a curve parametrized by arc length. The tangent vector field $t(s)$ is defined as follows:

$$t(s) := c'(s). \tag{28.4}$$

A different notation for the derivative, e.g. a dot, is used whenever the differentiation is performed with respect to a natural parameter.

Property 28.1.9. From the definition of natural parameters 28.1.4 and the previous definition it follows that the tangent vector is automatically a unit vector.

Definition 28.1.10 (Principal normal vector). Let $c(s)$ be a curve parametrized by arc length. The principal normal vector field is defined as follows:

$$n(s) := \frac{t'(s)}{\|t'(s)\|}. \tag{28.5}$$

Property 28.1.11. From Property 28.1.9 and the definition of the principal normal vector it follows that the tangent vector and principal normal vector are always orthogonal.

Definition 28.1.12 (Binormal vector). Let $c(s)$ be a curve parametrized by arc length. The binormal vector field is defined as follows:

$$b(s) := t(s) \times n(s). \tag{28.6}$$

Definition 28.1.13 (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 $\big\{t(s), n(s), b(s)\big\}$ is called the Frenet-Serret frame.

Remark. This basis does not have to be the same for every value of the parameter s.

Definition 28.1.14 (Curvature). Let $c(s)$ be a curve parametrized by arc length. The curvature of $c(s)$ is defined as follows:

$$\frac{1}{\rho(s)} := \|t'(s)\|. \tag{28.7}$$

Definition 28.1.15 (Torsion). Let $c(s)$ be a curve parametrized by arc length. The torsion of $c(s)$ is defined as follows:

$$\tau(s) := \rho(s)^2\,(t\ t'\ t''), \tag{28.8}$$

where $(a\ b\ c)$ denotes the **triple product** $a \cdot (b \times c)$.

Formula 28.1.16 (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:

$$\begin{cases} t'(s) &= & & \frac{1}{\rho(s)}n(s) & \\ n'(s) &= & -\frac{1}{\rho(s)}t(s) & + & \tau(s)b(s). \\ b'(s) &= & & -\tau(s)n(s) & \end{cases} \tag{28.9}$$

Theorem 28.1.17 (Fundamental theorem for curves). *Let $k, w : I \rightarrow \mathbb{R}$ be two C^1-functions with k nonnegative. There exists an interval $]-\varepsilon, \varepsilon[\ \subset I$ and a curve $c :]-\varepsilon, \varepsilon[\ \rightarrow \mathbb{R}^3$ with natural parameter s such that c has k as its curvature and w as its torsion.*

28.2 Surfaces

Notation 28.2.1. 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:

$$\sigma_i := \frac{\partial \sigma}{\partial q^i}. \tag{28.10}$$

Definition 28.2.2 (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:

$$T_p\Sigma := \left\{v \in \mathbb{R}^3 \,\middle|\, \left[v - \sigma(q_0^1, q_0^2)\right] \cdot \left[\sigma_1(q_0^1, q_0^2) \times \sigma_2(q_0^1, q_0^2)\right] = 0\right\}. \tag{28.11}$$

Definition 28.2.3 (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

$$N := \frac{\sigma_1 \times \sigma_2}{\|\sigma_1 \times \sigma_2\|}. \tag{28.12}$$

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.

28.2.1 First fundamental form

Definition 28.2.4 (Metric coefficients). Let σ be the parametrization of a surface. The metric coefficients g_{ij} are defined as follows:

$$g_{ij} := \sigma_i \cdot \sigma_j. \tag{28.13}$$

Definition 28.2.5 (Scale factor). The following factors are often used in vector calculus:

$$g_{ii} =: h_i^2. \tag{28.14}$$

Definition 28.2.6 (First fundamental form). Let σ be the parametrization of a surface. One can define a bilinear form $\mathrm{I}_p : T_p\Sigma \times T_p\Sigma \to \mathbb{R}$ that restricts the inner product on $\mathbb{R}^n$ to $T_p\Sigma$:

$$\mathrm{I}_p(v, w) := v \cdot w. \tag{28.15}$$

This bilinear form is called the first fundamental form or **metric**.

Corollary 28.2.7. 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 (28.13):

$$\mathrm{I}_p(v, w) = v^i\sigma_i \cdot w^j\sigma_j = g_{ij}v^iw^j. \tag{28.16}$$

Notation 28.2.8. The (arc) length 28.1.5 of a curve $c : I \to \mathbb{R}^n$ can be written as follows:

$$s = \int \|\dot{c}(t)\|dt \equiv \int \sqrt{ds^2}, \tag{28.17}$$

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):

$$ds^2 \equiv g_{ij}dq^idq^j. \tag{28.18}$$

Formula 28.2.9 (Inverse metric). Let (g_{ij}) denote the metric tensor. The matrix (g^{ij}) is defined as its inverse:

$$(g^{ij}) := \frac{1}{\det(g_{ij})}\begin{pmatrix} g_{22} & -g_{12} \\ -g_{12} & g_{11} \end{pmatrix}. \tag{28.19}$$

[1] The symbol σ denotes the embedding of the surface as a vector field while Σ denotes the geometric image of σ.

28.2.2 Isometries

Definition 28.2.10 (Isometry). An isometry is a 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 28.2.11. A diffeomorphism Φ is an isometry if and only if the metric coefficients of σ and σ' are the same.

Definition 28.2.12 (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 28.2.13. A diffeomorphism Φ is conformal if and only if the metric coefficients of σ and σ' are proportional.

Definition 28.2.14 (Area-preserving map). A diffeomorphism $\Phi : \Sigma \to \Sigma'$ is said to be area-preserving if it maps a subset of Σ to a subset of Σ' with the same area.

Property 28.2.15. A diffeomorphism Φ is area-preserving if and only if the metric coefficients of σ and σ' satisfy

$$g_{11}g_{22} - g_{12}^2 = g'_{11}g'_{22} - (g'_{12})^2 \tag{28.20}$$

for all points (q^1, q^2), i.e. if it preserves the determinant of the metric.

Corollary 28.2.16. A map that is area-preserving and conformal is also isometric.

28.2.3 Second fundamental form

Definition 28.2.17 (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:

$$\mathrm{II}_p(v, w) := L_{ij}(p)v^i w^j, \tag{28.21}$$

where $L_{ij} := N \cdot \sigma_{ij}$.

Definition 28.2.18 (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

$$\frac{1}{\rho_n(s)} := c''(s) \cdot N(s). \tag{28.22}$$

From the definition of the second fundamental form it follows that the normal curvature can be written as follows:

$$\frac{1}{\rho_n(s)} = \mathrm{II}(t, t) = \frac{\mathrm{II}(\dot{c}(\lambda), \dot{c}(\lambda))}{\mathrm{I}(\dot{c}(\lambda), \dot{c}(\lambda))}, \tag{28.23}$$

where the last equality holds for any given parameter λ.

Theorem 28.2.19 (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 28.2.20. 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 28.2.21 (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:

$$\frac{1}{\rho_g(s)} := \big(N(s)\ t(s)\ t'(s)\big). \tag{28.24}$$

Formula 28.2.22. Let c be a curve defined on a surface σ. From the definitions of the normal and geodesic curvature it follows that

$$\frac{1}{\rho^2} = \frac{1}{\rho_n^2} + \frac{1}{\rho_g^2}. \tag{28.25}$$

28.2.4 Curvature of a surface

Definition 28.2.23 (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:

$$L_p(\sigma_1) := -N_1 \qquad \text{and} \qquad L_p(\sigma_2) := -N_2. \tag{28.26}$$

Formula 28.2.24. Let $v, w \in T_p\Sigma$. The following equalities relate the second fundamental form and the Weingarten map:

$$L_p(v) \cdot w = L_p(w) \cdot v = \mathrm{II}_p(v, w). \tag{28.27}$$

Formula 28.2.25. Let $\big(g^{ij}\big)$ be the inverse of the metric. The matrix elements of L_p can be expressed as follows:

$$L_j^k = g^{ki} L_{ij}. \tag{28.28}$$

Formula 28.2.26 (Weingarten formulas).

$$N_j = -L_j^k \sigma_k \tag{28.29}$$

Definition 28.2.27 (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.

[2]Sometimes called the **shape operator**.

Property 28.2.28. 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 28.2.29 (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 28.2.30 (Rodrigues's formula). A curve is a line of curvature if and only if it is a solution of the following differential equation:

$$\frac{dN}{dt}(t) = -\frac{1}{R(t)}\frac{dc}{dt}(t). \tag{28.30}$$

If the curve satisfies this formula, the scalar function $1/R(t)$ coincides with the principal curvature along the curve.

Formula 28.2.31 (Differential equation for curvature lines).

$$\begin{vmatrix} (\dot{q}^2)^2 & -\dot{q}^1\dot{q}^2 & (\dot{q}^1)^2 \\ g_{11} & g_{12} & g_{22} \\ L_{11} & L_{12} & L_{22} \end{vmatrix} = 0 \tag{28.31}$$

Property 28.2.32. From Definition 28.2.27 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 28.2.33 (Gaussian curvature). The Gaussian curvature of a surface is defined as the determinant of the Weingarten map:

$$K := \frac{1}{R_1R_2}. \tag{28.32}$$

Definition 28.2.34 (Mean curvature). The mean curvature of a surface is defined as the trace of the Weingarten map:

$$H := \frac{1}{2}\left(\frac{1}{R_1} + \frac{1}{R_2}\right). \tag{28.33}$$

Property 28.2.35. The principal curvatures are the solutions of the following equation:

$$x^2 - 2Hx + K = 0. \tag{28.34}$$

This is the characteristic equation (20.60) of the Weingarten map.

Definition 28.2.36. 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 28.2.37. A surface Σ containing only umbilics is either part of a sphere or part of a plane.

Formula 28.2.38. 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

$$x_3 = \frac{1}{2}\left(\frac{x_1^2}{R_1} + \frac{x_2^2}{R_2}\right) \tag{28.35}$$

up to order $O(x^2)$.

Formula 28.2.39 (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

$$\frac{1}{\rho_n} = \frac{\cos^2\theta}{R_1} + \frac{\sin^2\theta}{R_2}. \tag{28.36}$$

Definition 28.2.40 (Asymptotic curve). A curve that is at every point tangent to a direction with zero normal curvature.

Formula 28.2.41 (Differential equation for asymptotic curves).

$$L_{11}\left(\dot{q}^1(t)\right)^2 + 2L_{12}\dot{q}^1(t)\dot{q}^2(t) + L_{22}\left(\dot{q}^2(t)\right)^2 = 0 \tag{28.37}$$

Property 28.2.42. 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.

28.2.5 Christoffel symbols and geodesics

Formula 28.2.43 (Gauss's formulas). The second derivatives of the surface σ are given by

$$\sigma_{ij} = L_{ij}N + \Gamma^k{}_{ij}\sigma_k, \tag{28.38}$$

where the **Christoffel symbols** $\Gamma^k{}_{ij}$ are defined as

$$\Gamma^k{}_{ij} := g^{kl}\sigma_l \cdot \sigma_{ij}. \tag{28.39}$$

Corollary 28.2.44. From the expression of the Christoffel symbols one can derive an alternative expression only in terms of the metric g_{ij}:

$$\Gamma^k{}_{ij} = \frac{1}{2}g^{kl}\left(\frac{\partial g_{il}}{\partial q^j} - \frac{\partial g_{ij}}{\partial q^l} + \frac{\partial g_{jl}}{\partial q^i}\right). \tag{28.40}$$

Definition 28.2.45 (Geodesic). A curve with zero geodesic curvature.

Property 28.2.46. 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 28.2.47 (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:

$$q''^k + \Gamma^k{}_{ij}q'^i q'^j = 0. \tag{28.41}$$

28.2.6 Theorema Egregium

Formula 28.2.48 (Codazzi-Mainardi equations).

$$\frac{\partial L_{ij}}{\partial q^k} - \frac{\partial L_{ik}}{\partial q^j} = {\Gamma^l}_{ik} L_{lj} - {\Gamma^l}_{ij} L_{lk} \tag{28.42}$$

Definition 28.2.49 (Riemann curvature tensor).

$${R^l}_{ijk} := \frac{\partial {\Gamma^l}_{ik}}{\partial q^j} - \frac{\partial {\Gamma^l}_{ij}}{\partial q^k} + {\Gamma^s}_{ik} {\Gamma^l}_{sj} - {\Gamma^s}_{ij} {\Gamma^l}_{ks} \tag{28.43}$$

Formula 28.2.50 (Gauss's equations).

$${R^l}_{ijk} = L_{ik} L^l_j - L_{ij} L^l_k \tag{28.44}$$

Theorem 28.2.51 (Theorema Egregium). *The Gaussian curvature is completely determined by the metric tensor g_{ij} and its derivatives:*

$$K = \frac{{R^l}_{121} g_{l2}}{g_{11} g_{22} - g_{12}^2}. \tag{28.45}$$

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 28.2.52. From the condition of isometries 28.2.11 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 28.2.53. 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 29

Manifolds

References for this chapter (and Part V in general) are [9, 10, 37, 59, 64, 114].

29.1 Charts

Definition 29.1.1 (Chart). Consider a topological space M and consider an open subset of $U \subseteq M$ such that there exists a homeomorphism $\varphi : U \to O$ where O is an open subset of $\mathbb{R}^n$. The pair (U, φ) is called a chart on M.

Definition 29.1.2 (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 29.1.3 (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.

Definition 29.1.4 (Maximal atlas). Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be two atlases on the same topological space. If $\mathcal{A}_1 \cup \mathcal{A}_2 = \mathcal{A}$ 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 29.1.5 (Manifold). A topological space equipped with a maximal C^0-atlas is called a **topological** manifold. An alternative definition (often used in topology) is that of a locally Euclidean Hausdorff space. 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 (among others) the property of paracompactness 7.5.15 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 32.7).

(For an alternative definition of manifolds in the context of *smooth spaces* see Section 42.2.)

If all transition maps are C^k-diffeomorphisms, the manifold is called a C^k-manifold. The limiting case, a C^∞-manifold, is also called a **smooth manifold**. If the transition maps are not only smooth, but even analytic 14.6.7, 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**.

Definition 29.1.6 (Structure sheaf ♣). Let M be a C^k-manifold. The structure sheaf $\mathcal{O}_M$ is defined as the sheaf 9.2.1 that assigns to every open set $U \subseteq M$ the set of C^k-functions $f : U \to \mathbb{R}$.

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 9.5.2 that is locally isomorphic to $\mathbb{R}^n$ equipped with its standard space of differentiable functions. (This is an extension of the algebro-geometric constructions from Sections 11.2 and 11.3.)

Property 29.1.7. 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. The manifolds can, up to isomorphism, be completely reconstructed from this algebraic data. (This can be seen as an analogue of the Gel'fand-Naimark theorem 24.1.41. However, no compactness is required here.)

Theorem 29.1.8 (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 automatically smooth.*

Theorem 29.1.9 (Radó-Moise). *In dimensions $1, 2$ and 3 there exists for every topological manifold a unique smooth structure.*

Theorem 29.1.10. *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 29.1.11. In dimension 4 there only exist partial results. For noncompact manifolds there uncountably many distinct smooth structures exist, while for compact manifolds no complete characterization has been found.

Definition 29.1.12 (Smooth function). Let $f : M \to N$ be a function between two smooth manifolds. It is said to be smooth if there exist charts (U, φ) and (V, ψ) for M and N with $f(U) \subseteq V$ such that the function

$$f_{\varphi\psi} = \psi \circ f \circ \varphi^{-1} \tag{29.1}$$

is smooth on $\mathbb{R}^n$. This function is called a **local representation** of f.

Definition 29.1.13 (Diffeomorphism). A homeomorphism f such that both f and f^{-1} are smooth.

Notation 29.1.14. The set of all C^∞-functions on a manifold M, defined on a neighbourhood of $p \in M$, is denoted by $C^\infty_p(M)$. This set forms a commutative ring when equipped with the usual sum and product (composition) of functions.

Remark 29.1.15. 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 29.1.16 (Differentiably good cover). A good cover 8.1.8 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 29.1.17. Every paracompact smooth manifold admits a (differentiably) good cover. Furthermore, if the manifold is compact, it admits a finite good cover.

29.2 Tangent vectors

Definition 29.2.1 (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 germs of smooth functions at p, i.e. a map $v_p : C_p^\infty(M) \to \mathbb{R}$ satisfying the properties

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**[1].

Property 29.2.2 (Constant functions). Constant functions $c : p \mapsto c$ lie in the kernel of all tangent vectors:

$$v_p(c) = 0. \tag{29.2}$$

Definition 29.2.3 (Tangent space). The set of all tangent vectors at a point $p \in M$ admits the structure of a vector space T_pM. A canonical choice of basis vectors is given by

$$\left.\frac{\partial}{\partial x^i}\right|_p : C_p^\infty(M) \to \mathbb{R} : f \mapsto \frac{\partial}{\partial x^i}\big(f \circ \varphi^{-1}\big)(\varphi(p)), \tag{29.3}$$

where (U, φ) is a coordinate chart such that $p \in U$ with local coordinates $(x^1, \ldots, x^n)$. The above basis vector are also often denoted by ∂_i.

Due to the explicit dependence of the tangent vectors on the point $p \in M$, it is clear that for curved manifolds the tangent spaces belonging to different points are not the same. However, they are related through the following property:

Property 29.2.4. For a smooth connected manifold, the tangent spaces satisfy

$$\dim(T_pM) = \dim(M) \tag{29.4}$$

for all $p \in M$. Theorem 20.2.14 then implies that the tangent spaces over two distinct points $p, q \in M$ are isomorphic. (A way to relate distinct tangent spaces will be presented in Sections 32.6 and 33.4.)

Alternative Definition 29.2.5 (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:

$$\frac{d(\varphi \circ \gamma_1)}{dt}(0) = \frac{d(\varphi \circ \gamma_2)}{dt}(0). \tag{29.5}$$

This defines an equivalence relation[2] on the set of smooth curves through p. The tangent space at p is then defined as the set of equivalence classes of tangent curves through p. These equivalence classes can be explicitly constructed as follows. The tangent vector to the curve $c(t)$ through p is defined by the following formula:

$$v_p(f) := \left.\frac{d(f \circ c)}{dt}\right|_{t=0}. \tag{29.6}$$

[1] More generally, every operation that satisfies the Leibniz property is called a derivation.

[2] The relation is well-defined because the transition functions (and their Jacobian matrices) are invertible and thus nonsingular.

Applying the chain rule gives

$$v_p(f) = \frac{\partial(f \circ \varphi^{-1})}{\partial x^i}(\varphi(p))\frac{dx^i}{dt}(0), \tag{29.7}$$

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 equation (29.5) define the same tangent vector.

Proof of equivalence. Let (U, φ) be a chart around the point $p \in M$. Using the first definition of a tangent vector 29.2.3, i.e.

$$\left.\frac{\partial}{\partial q^i}\right|_p : C_p^\infty(M, \mathbb{R}) \to \mathbb{R} : f \mapsto \frac{\partial}{\partial q^i}\big(f \circ \varphi^{-1}\big)(\varphi(p)),$$

one can rewrite Equation (29.7)

$$v_p(f) = \frac{\partial(f \circ \varphi^{-1})}{\partial q^i}(\varphi(p))\frac{dq^i}{dt}(0)$$

as follows:

$$v_p(f) = \left.\frac{\partial f}{\partial q^i}\right|_p \frac{dq^i}{dt}(0).$$

Because the partial derivatives as defined in 29.2.3 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 $v = (v^1, \ldots, v^n)$ 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 Equation

(29.7) one can find

$$\begin{aligned} v_{p,\chi}(f) &= \frac{\partial(f\circ\varphi^{-1})}{\partial q^i}(\varphi(p))\frac{d(\varphi^i\circ\chi)}{dt}(0) \\ &= \frac{\partial(f\circ\varphi^{-1})}{\partial q^i}(\varphi(p))\frac{d}{dt}\big(\varphi^i\circ\gamma+\varphi^i\circ\delta-\varphi^i(p)\big) \\ &= \frac{\partial(f\circ\varphi^{-1})}{\partial q^i}(\varphi(p))\left(\frac{d(\varphi^i\circ\gamma)}{dt}+\frac{d(\varphi^i\circ\delta)}{dt}\right) \\ &= v_{p,\gamma}(f)+v_{p,\delta}(f). \end{aligned}$$

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 $\lambda\in K$ can be proven similarly by defining the curve

$$\chi(t)=\varphi^{-1}\circ\Big[\lambda\Big(\varphi\circ\gamma(t)\Big)\Big].$$

□

[a] More precisely, representatives of equivalence classes of vectors tangent to curves.

29.3 Submanifolds

29.3.1 Immersions and submersions

In this section the tangent map induced by a smooth function $f:M\to N$ is denoted by $T_pf:T_pM\to T_{f(p)}N$. A formal definition is given in Equation (32.1). For now this will be the map that is locally represented by the Jacobian of f.

Definition 29.3.1 (Immersion). A differentiable function $f:M\to N$ between smooth manifoldsfor which the derivative is everywhere injective or, equivalently, such that its derivative has maximal rank everywhere:

$$\mathrm{rk}(T_pf)=\dim(M)\qquad\forall p\in M. \tag{29.8}$$

Definition 29.3.2 (Critical point). A point $p\in\mathrm{dom}(f)$ is said to be critical if the rank of the Jacobian T_pf is not maximal. The image of a critical point is called a **critical value**.

At a critical point $p\in M$ the Hessian of f gives a well-defined quadratic form. A critical point is said to be **nondegenerate** if the Hessian is nonsingular there.

Property 29.3.3 (Criticality). A point $p\in\mathrm{dom}(f)$ is critical if and only if there exists a chart (U,φ) containing p for which $\partial_i f(p)=0$.

Theorem 29.3.4 (Sard). *Consider a differentiable function $\psi:M\to N$, where $\dim(M)=m$ and $\dim(N)=n$ and let $k_0=\max\{1,m-n+1\}$. If ψ is of class C^k, with $k\geq k_0$, the set of critical values of ψ has Lebesgue measure 0.*

Definition 29.3.5 (Regular point). A regular point of f is a point $p\in M$ such that T_pf is surjective.

Definition 29.3.6 (Regular value). Let $f:M\to N$ be a differentiable function between smooth manifolds. A point $y\in N$ is called a **regular value** if every point in the preimage $f^{-1}(y)$ is a regular point or, equivalently, if it is not a critical value.

Corollary 29.3.7. It follows from Property 29.3.3 that a point $p\in\mathrm{dom}(f)$ is regular if and only if $\partial_i f(p)\neq 0$ for all charts (U,φ) containing p.

Definition 29.3.8 (Submersion). A differentiable function $f : M \to N$ between smooth manifolds such that all $p \in M$ are regular or, equivalently, such that

$$\mathrm{rk}(T_pf) = \dim(N) \qquad \forall p \in M. \tag{29.9}$$

29.3.2 Submanifolds

Definition 29.3.9 (Embedding). A differentiable function between smooth manifolds that is both an immersion and an embedding in the topological sense 7.2.11. This implies that the submanifold topology coincides with the subspace topology 7.1.4.

Definition 29.3.10 (Embedded submanifold). Let M be a manifold. A smooth manifold N is called an embedded or **regular submanifold** (of M) if there exists an embedding $f : M \hookrightarrow N$.

Definition 29.3.11 (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:

$$\mathbb{R}^m \cong \mathbb{R}^m \times \{0, \dots, 0\} \hookrightarrow \mathbb{R}^m \times \mathbb{R}^{n-m} \cong \mathbb{R}^n. \tag{29.10}$$

Subspaces obtained in this way, i.e. by setting a number of coordinates equal to 0 (or any other constant), are called slices.

Alternative Definition 29.3.12 (Embedded submanifold). A subset N of M for which there exists a positive integer k and such that for every point $p \in N$ there exists a chart (U, φ) that satisfies

$$\varphi(U \cap N) = \varphi(U) \cap (\mathbb{R}^k \times \underbrace{\{0, \dots, 0\}}_{\dim(M)-k}). \tag{29.11}$$

The set $U \cap N$ is called a **slice** of (U, φ) in analogy with the previous definition of a (standard) slice.

Definition 29.3.13 (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 29.3.14 (Submersion theorem[3]). *Consider a smooth map $f : M_1 \to M_2$ between smooth manifolds and let $y \in M_2$ be a regular value. Then $N = f^{-1}(y)$ is a submanifold of M_1 with codimension* $\dim(M_2)$.

Definition 29.3.15 (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 in fact an embedded submanifold. It is called a closed embedded manifold.

Example 29.3.16 (Stiefel manifold). Let V be an inner product space 20.3.1 over a field K. The set of orthonormal k-frames can be embedded in $K^{n\times k}$. It is a compact embedded submanifold, called the Stiefel manifold of k-frames over V.

Definition 29.3.17 (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 the following relation holds:

$$T_pX + T_pY = T_pM. \tag{29.12}$$

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).

[3] Also called the **regular value theorem**.

Property 29.3.18 (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 29.3.19 (Intersection number). By the above property two closed submanifolds $X, Y \subset M$ with complementary dimension that intersect transversally, have a finite number of intersection points. Given an orientation on M, 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

$$I(H(X,0), Y) = I(H(X,1), Y). \tag{29.13}$$

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.

Property 29.3.20 (Euler characteristic). Consider a closed manifold M. The Euler characteristic 8.2.18 is given by

$$\chi(M) = I(\Delta_M, \Delta_M), \tag{29.14}$$

where $\Delta_M \in M \times M$ is the diagonal of M.

29.4 Manifolds with boundary

Definition 29.4.1 (Manifold with boundary). Let $\mathbb{H}^n$ denote the upper half space:

$$\mathbb{H}^n := \mathbb{R}^{n-1} \times \mathbb{R}^+ = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid x_n \geq 0\}. \tag{29.15}$$

An n-dimensional manifold with boundary is defined as a topological space M equipped with a maximal atlas consisting of (regular) charts (U, φ) such that U is diffeomorphic to $\mathbb{R}^n$ (these points are called **interior points**) and **boundary charts** (V, ϕ) such that V is diffeomorphic to $\mathbb{H}^n$ (these points are called **boundary points**).

Remark 29.4.2 (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.

Property 29.4.3. Let M be an n-dimensional manifold with boundary and let (U, φ) be a chart for $p \in \partial M$.

$$\varphi(p) \in \partial\mathbb{H}^n = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid x_n = 0\} \tag{29.16}$$

Definition 29.4.4 (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

$$\varphi : U \to \mathbb{R}^k \times (\mathbb{R}^+)^l.$$

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:

$$\varphi \circ \psi^{-1}(V \cap \{0\} \times \mathbb{R}^k) \subset \{0\} \times \mathbb{R}^k.$$

Remark 29.4.5. 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. There exists no such diffeomorphism and accordingly one has to make a distinction between the type of corners.

29.4.1 Cobordisms ♣

Definition 29.4.6 (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 one can also define a cobordism, but there the manifolds X, Y should respect the orientation of ∂M.

Definition 29.4.7 (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 commutative graded ring $\Omega = \bigoplus_{n=0}^\infty \Omega_n$.

?? COMPLETE ??

29.5 Morse theory

29.5.1 Morse functions

Definition 29.5.1 (Morse function). Let M be a smooth manifold. A smooth function is called a Morse function if it has no degenerate critical points 29.3.2.

Property 29.5.2 (Density). The set of Morse functions is open and dense in the C^2-topology (see Section 38.3 on jet spaces).

Definition 29.5.3 (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 or every proper function satisfies this condition.

Corollary 29.5.4. 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.

Corollary 29.5.5. Let γ be a flow line on M such that $f(\gamma)$ is bounded. The flow line is complete and its limits are critical points of f. Moreover, the convergence at $t \longrightarrow \pm\infty$ is exponential.

Definition 29.5.6 (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 $\lambda_p(f)$.

To any Morse function one can associate a series called the **Morse counting-series**:

$$M_t(f) := \sum_{p\in\mathrm{crit}(f)} t^{\lambda_p(f)}. \tag{29.17}$$

If M is compact, the nondegeneracy condition implies that the above sum only has a finite number of terms.

Property 29.5.7 (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 $(U, x_1, \ldots, x_n)$ around p such that $x_i(p) = 0$ and

$$f|_U(x) = f(p) - x_1^2 - \cdots + x_k^2 + \cdots, \tag{29.18}$$

where k is the Morse index of f.

Corollary 29.5.8. The critical points of a Morse function are isolated.

Remark 29.5.9 (Morse-Palais lemma). The Morse lemma can be generalized to open subsets of Banach spaces (and thus to infinite-dimensional manifolds).

Definition 29.5.10 (Self-indexing function). A Morse function is said to be self-indexing if at every critical points its value is equal to its index.

?? COMPLETE ??

29.5.2 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 (it is assumed that M comes equipped with a covariant derivative):

Definition 29.5.11 (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.

29.5.3 Morse homology

Definition 29.5.12 (Gradient-like vector field). Consider a Morse function $f \in C^\infty(M)$. A vector field X is said to be gradient-like 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

$$X = -2\sum_{i=1}^{\lambda_p(f)} x^i\partial_i + 2\sum_{i=\lambda_p(f)+1}^{\dim(M)} x^i\partial_i. \tag{29.19}$$

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 29.5.13. Let $f \in C^\infty(M)$ be a Morse function on a compact manifold and consider a gradient-like vector field X (with respect to f). For every $p \in M$ the limits of the flow line of $-X$, passing through p, are critical points of f.

Definition 29.5.14 (Stable and unstable manifold). Let $f \in C^\infty(M)$ be a Morse function and consider a gradient-like vector field X (with respect to f). For every critical point p of f, one defines the stable and unstable manifold of X as follows:

$$W_p^\pm(X) := \{x \in M \mid \lim_{t\to\pm\infty}\Phi_t(x) = p\}, \tag{29.20}$$

where Φ_t denotes the flow of $-X$. These sets carry the structure of a smooth manifold, locally diffeomorphic to $\mathbb{R}^{\dim(M)-\lambda_p(f)}$ and $\mathbb{R}^{\lambda_p(f)}$, respectively.

Definition 29.5.15 (Morse-Smale pair). Let $f \in C^\infty(M)$ be a Morse function and consider a gradient-like vector field X (with respect to f). If for all critical points $p, q \in \mathrm{Crit}(f)$ one has that

$$W_p^+(X) \pitchfork W_q^-(X), \tag{29.21}$$

the pair (f, X) is called a Morse-Smale pair.

Property 29.5.16. If M is compact, there exists a self-indexing Morse-Smale pair.

Property 29.5.17. 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. By $\mathcal{M}(p,q)$ one denotes the set of integral curves of $-\nabla f$ that start at p and end at q, i.e. the integral curves γ that satisfy $\gamma([0,1]) \subset W_p^-(\nabla f) \cap W_q^+(\nabla f)$. 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 29.5.18 (Morse homology). The chain groups are defined as follows

$$CM_k(M,f) := \bigoplus_{\substack{p\in\mathrm{Crit}(f)\\ \lambda_p(f)=k}} \mathbb{Z}\langle p\rangle. \tag{29.22}$$

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:

$$\partial p := \sum_{\substack{q\in\mathrm{Crit}(f)\\ \lambda_q(f)=\lambda_p(f)-1}} \left|\overline{\mathcal{M}}(p,q)\right|\langle q\rangle. \tag{29.23}$$

One can show that $\partial^2 = 0$ and Morse homology is defined as the homology of this complex:

$$HM_\bullet(M,f) := \frac{\ker(\partial)}{\mathrm{im}(\partial)}. \tag{29.24}$$

29.6 Surgery theory ♣

Definition 29.6.1 (Dehn twist). Consider an orientable surface M together with a simple closed curve c. A tubular neighbourhood[4] T 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 ??

[4]See Definition 32.1.13 for a formal definition.

Chapter 30

Lie groups and Lie algebras

References for this chapter are [50, 62]. The main reference for the section on *coadjoint orbits* is [104]. For some concepts such as vector fields and pushforwards the reader is referred to Chapter 32.

30.1 Lie groups

Definition 30.1.1 (Lie group). A group that is also a differentiable manifold such that both the multiplication and inversion maps are smooth functions.[1]

Definition 30.1.2 (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 30.1.3 (Closed subgroup theorem[2]). *A closed subgroup of a Lie group is a regular Lie subgroup.*

Property 30.1.4 (Generating neighbourhoods). Let G be a connected Lie group. Every neighbourhood U_e of the identity e generates G, i.e. every element $g \in G$ can be written as a word in U_e.

Definition 30.1.5 (Isogeny). Let G, H be two Lie groups. G and H are said to be isogenous if one is a covering space 7.2.13 of the other. The covering map is then called an isogeny between G and H.

30.1.1 Left invariant vector fields

Definition 30.1.6 (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 $g \in G$:

$$L_{g,*}X(h) = X(gh), \tag{30.1}$$

where L denotes the regular (left) action on G. The term "left-invariant vector field" is often abbreviated as **LIVF**.

Property 30.1.7. The set $\mathfrak{X}^L(G)$ of LIVFs on a (real) Lie group G is a vector space over $\mathbb{R}$.

Property 30.1.8 (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 of G. Furthermore, for every $v \in T_eG$ there exists a LIVF $X \in \mathfrak{X}^L(G)$ such that $X(e) = v$ and this mapping is an isomorphism from T_eG to $\mathfrak{X}^L(G)$.

[1] For complex Lie groups one requires the definition of a complex manifold (see Chapter 37).

[2] Sometimes called **Cartan's theorem**.

30.1.2 One-parameter subgroups

Definition 30.1.9 (One-parameter subgroup). A Lie group morphism $\Phi : \mathbb{R} \to G$ from the additive group of real numbers to G.

Property 30.1.10. Let $\Phi : \mathbb{R} \to G$ be a one-parameter subgroup of G 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 30.1.11. 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 30.1.12 (Left-invariant vector fields). All LIVFs X are complete[3], 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[4] $D(X)$ 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.

30.2 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.

30.2.1 Definitions

Definition 30.2.1 (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**: $[\lambda u + v, w] = \lambda[u, w] + [v, w]$ for all $\lambda \in K$,
2. **Alternativity**: $[v, v] = 0$, and
3. **Jacobi identity**: $[u, [v, w]] + [v, [w, u]] + [w, [u, v]] = 0$.

Remark 30.2.2. Note that often the alternativity condition is replaced by an antisymmetry condition. However, this is only equivalent over fields of characteristic $\neq 2$. For $\mathrm{char}(K) = 2$ one only has that alternativity implies antisymmetry. Since the fields will almost exclusively be $\mathbb{R}$ or $\mathbb{C}$, this will not pose any problems and, hence, the antisymmetry condition will always be assumed.

Definition 30.2.3 (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 29.2.1 with respect to the operation $\star$, i.e.

$$\{x, y \star z\} = \{x, y\} \star z + y \star \{x, z\}.$$

[3]See Definition 32.3.13 for a general statement.
[4]See Definition 32.3.11.

Definition 30.2.4 (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 $\{e_k\}_{k\in I}$:

$$[e_i, e_j] = \sum_{k\in I} {c_{ij}}^k e_k. \tag{30.2}$$

The coefficients ${c_{ij}}^k$ are called the structure constants of the Lie algebra. (Note that these constants are basis-dependent.)

Property 30.2.5 (Isomorphism). Two Lie algebras $\mathfrak{g}, \mathfrak{h}$ are isomorphic if one can find bases $\mathcal{B}$ for $\mathfrak{g}$ and $\mathcal{C}$ for $\mathfrak{h}$ such that the associated structure constants are equal.

Example 30.2.6 (Lie algebra of LIVFs). Consider the vector space $\mathfrak{X}^L(G)$ of LIVFs on a Lie group G. Using Equation (32.27) one can show that the commutator (Lie bracket) 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 30.1.8 is used to relate the above Lie algebra of left-invariant vector fields and the tangent space to the identity:

Alternative Definition 30.2.7 (Lie algebra of Lie group). Let G be a Lie group. The tangent space $\mathfrak{g} := T_eG$ has the structure of a Lie algebra where the Lie bracket is induced by the commutator of vector fields (32.27) in the following way:

$$[x, y]_{\mathfrak{g}} := L_{g^{-1},*}[L_{g,*}x, L_{g,*}y], \tag{30.3}$$

where $x, y \in T_eG$ 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 30.2.8. Lie algebras are generally denoted by fraktur symbols. For example, the Lie algebra associated with the Lie group G is often denoted by $\mathfrak{g}$.

Theorem 30.2.9 (Ado). *Every finite-dimensional Lie algebra can be embedded as a subalgebra of* $\mathfrak{gl}_n \cong M_n$.

Theorem 30.2.10 (Lie's third theorem). *Every finite-dimensional Lie algebra* $\mathfrak{g}$ *is the Lie algebra of a unique simply-connected Lie group* G.

Definition 30.2.11 (Lie algebra morphism). A map $\Phi : \mathfrak{g} \to \mathfrak{h}$ is called a Lie algebra morphism if it satisfies the equation

$$\Phi([x, y]) = [\Phi(x), \Phi(y)] \tag{30.4}$$

for all $x, y \in \mathfrak{g}$.

Property 30.2.12 (Homomorphism theorem[5]). Let G, H be two Lie groups with G 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 $\Phi = \phi_*$. Conversely, every Lie group morphism induces a Lie algebra morphism through its differential (see Formula 30.3.8).

Definition 30.2.13 (Derivation). Given a Lie algebra $\mathfrak{g}$, space of derivations $\mathfrak{Der}(\mathfrak{g})$ is defined as the space of linear maps d such that

$$d([x, y]) = [dx, y] + [x, dy] \tag{30.5}$$

for all $x, y \in \mathfrak{g}$. This vector space becomes a Lie algebra when equipped with the commutator of linear maps.

[5] Also called **Lie's second theorem**.

30.2.2 Exponential map

Formula 30.2.14 (Exponential map). Let X be a LIVF on G. The exponential map $\exp : \mathfrak{g} \to G$ is defined as

$$\exp(X) := \gamma_X(1), \tag{30.6}$$

where γ_X is the associated one-parameter subgroup from Property 30.1.12.

Property 30.2.15 (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 30.2.16. Because the identity morphism $\mathbb{1}_{\mathfrak{g}} = \exp_*$ is an isomorphism, the inverse function theorem (see Theorem 32.1.10) implies that the image of exp will contain a neighbourhood of the identity $e \in G$. If G is connected, Property 30.1.4 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 30.2.17 (Matrix Lie groups). For matrix Lie groups one can define the classic matrix exponential:

$$e^{tA} := \sum_{k=0}^{\infty} \frac{(tA)^k}{k!}. \tag{30.7}$$

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 $A \in M_{m,n}$ and that it is invertible with inverse given by $\exp(-A)$. Using Ado's theorem 30.2.9 one can then use this matrix exponential to represent the exponential map for any (finite-dimensional) Lie algebra.

Remark 30.2.18. If G is a 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 30.2.19 (Baker-Campbell-Hausdorff formula). Consider the equation

$$z = \log(\exp(x)\exp(y)), \tag{30.8}$$

where $x, y \in \mathfrak{g}$. The solution is given by the following formula:

$$e^x e^y = \exp\left(x + y + \frac{1}{2}[x,y] + \frac{1}{12}[x,[x,y]] - \frac{1}{12}[y,[x,y]] + \cdots\right). \tag{30.9}$$

One should note that this formula will only converge if x, y are sufficiently small. For matrix Lie algebras this means that $\|x\| + \|y\| < \frac{\ln(2)}{2}$ in the Hilbert-Schmidt norm 20.4.7. 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 30.2.20 (Lie product formula[6]). Let $\mathfrak{g}$ be a Lie algebra. The following formula applies to any $x, y \in \mathfrak{g}$:

$$e^{x+y} = \lim_{n\to\infty}\left(e^{\frac{x}{n}} e^{\frac{y}{n}}\right)^n. \tag{30.10}$$

[6] Also called the **Lie-Trotter formula**. Later, extensions where given by *Kato* and *Suzuki* for certain unbounded operators.

30.2.3 Examples

Example 30.2.21 (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 30.2.22 (General linear group). An interesting example is the Lie algebra associated to the Lie group of invertible complex[7] matrices $GL(n, \mathbb{C})$. 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 $A \in GL(n, \mathbb{C})$ and every $B \in \mathfrak{gl}(n, \mathbb{C})$ the following equality holds:

$$L_{A,*}(B) = L_A(B). \tag{30.11}$$

Corollary 30.2.23. By noting that the endomorphism ring $\mathrm{End}(V)$ of an n-dimensional vector space V is given by the matrix ring $M_n(K)$, it can be seen that $\mathrm{End}(V)$ also forms a Lie algebra when equipped with the commutator of linear maps.

Example 30.2.24 (Isometries). The Lie algebra associated $\mathfrak{isom}(V)$ with the group of isometries $\mathrm{Isom}(V)$ of a nondegenerate Hermitian form satisfies the condition

$$\langle Xv, w\rangle = -\langle v, Xw\rangle \tag{30.12}$$

for all Lie algebra elements $X \in \mathfrak{isom}(V)$. It follows that the Lie algebra consists of all skew-Hermitian operators.

Two explicit examples are:

Example 30.2.25 (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 21.4.8, i.e. $c_{ijk} = \varepsilon_{ijk}$.

Example 30.2.26 (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 restricting $GL(2, \mathbb{C})$ to the subset of matrices with unit determinant:

Example 30.2.27 (Lie algebra of $SL(2, \mathbb{C})$). 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:

$$l_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right),*}\left(\left.\frac{\partial}{\partial x^i}\right|_e\right) = \begin{pmatrix} a & 0 & b \\ -b & a & 0 \\ c & 0 & \frac{1+bc}{a} \end{pmatrix}^m_{\ i} \left.\frac{\partial}{\partial x^m}\right|_{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)}, \tag{30.13}$$

where the coordinate chart (U, ϕ) defined by

$$U = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{C}) \,\middle|\, a \neq 0 \right\}$$

and

$$\phi : U \to \mathbb{C}^3 : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto (a, b, c)$$

was used. One can then use this formula to work out the Lie bracket of the basis vectors ∂_i to obtain the following structure constants:

$$\begin{aligned} [\partial_1, \partial_2] &= 2\partial_2 \\ [\partial_1, \partial_3] &= -2\partial_3 \\ [\partial_2, \partial_3] &= \partial_1. \end{aligned} \tag{30.14}$$

[7]This result is also valid for real matrices.

30.2.4 Solvable Lie algebras

Definition 30.2.28 (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 $[x, S] \subseteq S$.

Definition 30.2.29 (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 30.2.30 (Derived algebra). Let $\mathfrak{g}$ be a Lie algebra. The derived Lie algebra is defined as follows:

$$[\mathfrak{g}, \mathfrak{g}] := \big\{[x, y] \,\big|\, x, y \in \mathfrak{g}\big\}. \tag{30.15}$$

Definition 30.2.31 (Solvable Lie algebra). Consider the sequence of derived Lie algebras

$$\mathfrak{g} \geq [\mathfrak{g}, \mathfrak{g}] \geq [[\mathfrak{g}, \mathfrak{g}], [\mathfrak{g}, \mathfrak{g}]] \geq \cdots \tag{30.16}$$

If this sequence ends in the zero space, the Lie algebra $\mathfrak{g}$ is said to be solvable.

Remark 30.2.32. In general one can define the derived series for any ideal and, accordingly, define solvability for ideals.

Definition 30.2.33 (Radical). The largest solvable ideal of a Lie algebra.

30.2.5 Simple Lie algebras

Definition 30.2.34 (Direct sum). The direct sum $\mathfrak{g} \oplus \mathfrak{h}$ of two Lie algebras $\mathfrak{g}, \mathfrak{h}$ is defined as the direct sum 20.1.15 in the sense of vector spaces where the Lie bracket is extended using the relation

$$[x, y] = 0 \tag{30.17}$$

for all $x \in \mathfrak{g}$ and $y \in \mathfrak{h}$.

Definition 30.2.35 (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 $\rho : \mathfrak{g} \to \mathrm{Der}(\mathfrak{h})$ is defined as the direct sum in the sense of vector spaces where the Lie bracket is extended using the relation

$$[g, h] := \rho(g)(h) \tag{30.18}$$

for all $g \in \mathfrak{g}, h \in \mathfrak{h}$. This also turns $\mathfrak{h}$ into an ideal.

Definition 30.2.36 (Simple Lie algebra). A Lie algebra is said to be simple if it is non-Abelian and if it has no nontrivial ideals.

Definition 30.2.37 (Semisimple Lie algebra). A Lie algebra is said to be semisimple if it is the direct sum of simple Lie algebras.

Theorem 30.2.38 (Levi decomposition). *Let $\mathfrak{g}$ be a finite-dimensional Lie algebra. It can be decomposed as follows:*

$$\mathfrak{g} = \mathfrak{R} \ltimes (\mathfrak{L}_1 \oplus \cdots \oplus \mathfrak{L}_n), \tag{30.19}$$

where $\mathfrak{R}$ is the radical of $\mathfrak{g}$ and the algebras $\mathfrak{L}_i$ are simple subalgebras.

Definition 30.2.39. 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 $\mathfrak{g}$.

30.2.6 Central extensions

Definition 30.2.40 (Central extension). A central extension of a Lie algebra $\mathfrak{g}$ by an Abelian Lie algebra $\mathfrak{a}$ is an exact sequence of Lie algebras of the form

$$0 \longrightarrow \mathfrak{a} \longrightarrow \mathfrak{h} \longrightarrow \mathfrak{g} \longrightarrow 0, \tag{30.20}$$

where the image of $\mathfrak{a}$ lies in the center of $\mathfrak{h}$.

Construction 30.2.41 (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. $\Theta([u, v], w) + \Theta([v, w], u) + \Theta([w, u], v) = 0$.

Such a morphism is called a **Lie algebra 2-cocycle** (see Section 30.6 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:
>
> $$[v \oplus \lambda, w \oplus \mu] := [v, w]_{\mathfrak{g}} \oplus \Theta(v, w). \tag{30.21}$$

30.3 Representation theory

30.3.1 Lie groups

Definition 30.3.1 (Representation of Lie groups). Let G be a Lie group and let V be a vector space. A representation of G on V is a Lie group morphism $\rho : G \to \mathrm{GL}(V)$.

Example 30.3.2 (Adjoint representation of Lie groups). Let G be a Lie group and consider the adjoint action $\mathrm{Ad}_g : h \mapsto ghg^{-1}$. The adjoint representation of G on $\mathfrak{g}$ is defined as the differential $\mathrm{Ad}_{g,*}$ at the identity element. For matrix Lie groups this becomes

$$\mathrm{Ad}_g : T_eG \to T_eG : x \mapsto gxg^{-1}. \tag{30.22}$$

30.3.2 Lie algebras

Definition 30.3.3 (Representation of Lie algebras). Let $\mathfrak{g}$ be a Lie algebra and let V be a vector space. A representation of $\mathfrak{g}$ on V is a Lie algebra morphism $\rho : \mathfrak{g} \to \mathrm{End}(V)$.

Example 30.3.4 (Adjoint representation of Lie algebras). Using the fact that the adjoint representation of Lie groups is smooth, one can define the adjoint representation of Lie algebras as

$$\mathrm{ad}_x := \mathrm{Ad}_{g,*}, \tag{30.23}$$

where $g = e^{tx}$. More explicitly, the adjoint map is given by

$$\mathrm{ad}_x(y) = [x, y], \tag{30.24}$$

for all $x, y \in \mathfrak{g}$.

Property 30.3.5 (Faithful). The adjoint representation ad_x is faithful for all $x \in \mathfrak{g}$.

Property 30.3.6 (Jacobi identity). Given the antisymmetry of the Lie bracket, the Jacobi identity is equivalent to $\mathrm{ad} : \mathfrak{g} \rightarrow \mathrm{End}(\mathfrak{g})$ being a Lie algebra morphism, i.e. $\mathrm{ad}_{[x,y]} = [\mathrm{ad}_x, \mathrm{ad}_y]$.

Formula 30.3.7 (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:

$$(\mathrm{ad}_{e_i})^j_k = c^j_{ik}. \tag{30.25}$$

Formula 30.3.8 (Induced morphism). Let $\phi : G \rightarrow H$ be a Lie group morphism with G simply-connected. This morphism induces a Lie algebra morphism $\Phi : \mathfrak{g} \rightarrow \mathfrak{h}$ given by

$$\Phi(x) := \left.\frac{d}{dt}\phi\big(e^{tx}\big)\right|_{t=0} \tag{30.26}$$

or, equivalently,

$$\phi\big(e^{tx}\big) = e^{t\Phi(x)}. \tag{30.27}$$

The morphism induced by $\mathrm{Ad} : G \rightarrow H$ is precisely $\mathrm{ad} : \mathfrak{g} \rightarrow \mathfrak{h}$. Informally one can thus say that the infinitesimal version of the similarity transformation is given by the commutator:

Corollary 30.3.9 (Commutator). For the algebra of the general linear group GL_n the Lie bracket is given by the commutator:

$$[A, B] := AB - BA. \tag{30.28}$$

30.3.3 Coadjoint orbits

Definition 30.3.10 (Coadjoint representation). The representation of a Lie group G on the dual space $\mathfrak{g}^*$ defined by

$$\langle \mathrm{Ad}^*_g(\omega), v\rangle := \langle \omega, \mathrm{Ad}^{-1}_g(v)\rangle. \tag{30.29}$$

Infinitesimally this induces a representation of $\mathfrak{g}$ on its linear dual. It is given by

$$\mathrm{ad}^*_x : \omega \mapsto -\omega \circ \mathrm{ad}_x. \tag{30.30}$$

Definition 30.3.11 (Coadjoint orbit). Given an element $\omega \in \mathfrak{g}^*$, the coadjoint orbit Ω_ω is defined as as the orbit of ω under the action of G. This orbit can also be defined as the homogeneous space G/G_ω.

The following important construction shows that every coadjoint orbit is in fact canonically a symplectic manifold (Chapter 35):

Definition 30.3.12 (Kirillov-Kostant-Souriau form). Consider a coadjoint orbit Ω_α. Because α 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 $v = \mathrm{ad}^*_x(\alpha)$ for some $x \in \mathfrak{g}$. Now, a symplectic form on Ω_α is defined as follows:

$$\omega_\alpha(\mathrm{ad}^*_x(\alpha), \mathrm{ad}^*_y(\alpha)) = \langle \alpha, [x, y]\rangle. \tag{30.31}$$

?? COMPLETE (perhaps move to chapter on symplectic geometry) ??

30.4 Structure

30.4.1 Killing form

Definition 30.4.1 (Killing form). Let $\mathfrak{g}$ be a finite-dimensional Lie algebra over the field k. The Killing form (sometimes **Cartan-Killing** form) on $\mathfrak{g}$ is defined as the following symmetric bilinear form:

$$K : \mathfrak{g} \otimes \mathfrak{g} \to k : (x, y) \mapsto \mathrm{tr}(\mathrm{ad}_x \circ \mathrm{ad}_y). \tag{30.32}$$

The trace can be calculated by choosing a (finite-dimensional) representation of the Lie algebra using Ado's theorem 30.2.9.

From Formula 30.3.7 one can derive the value of the Killing form on the basis $\{e_i\}_{i\leq n}$:

$$K_{ij} = c^l_{ik} c^k_{jl}, \tag{30.33}$$

where c^k_{ij} are the structure constants of the Lie algebra.

Theorem 30.4.2 (Cartan's criterion). *A Lie algebra is semisimple if and only if its Killing form is nondegenerate.*

Property 30.4.3. If a Lie group G is compact, the Killing form of its associated Lie algebra $\mathfrak{g}$ is negative-definite.

Corollary 30.4.4. Let G be a compact Lie group. If its Lie algebra is semisimple, the Killing form K induces a metric

$$g : (x, y) \mapsto -\mathrm{tr}(\mathrm{ad}_x \circ \mathrm{ad}_y) = -K(x, y), \tag{30.34}$$

which turns the corresponding Lie group G into a *Riemannian manifold* (see Chapter 34).

Property 30.4.5 (Killing form is invariant). The Killing-form is Ad-invariant:

$$K(\mathrm{Ad}_g(x), \mathrm{Ad}_g(y)) = K(x, y) \tag{30.35}$$

for all $g \in G$. More generally, the Killing form is invariant under all automorphisms of $\mathfrak{g}$.

Corollary 30.4.6. The adjoint map ad_z is antisymmetric with respect to the Killing form:

$$K(\mathrm{ad}_z x, y) = -K(x, \mathrm{ad}_z y). \tag{30.36}$$

Property 30.4.7 (Invariant forms). For a simple Lie algebra, every (ad-)invariant symmetric bilinear form is a scalar multiple of the Killing form.

Example 30.4.8. 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

$$\mathrm{tr}(AB) = 2nK(A, B). \tag{30.37}$$

Property 30.4.9 (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:

$$c_{ijk} = c_{jki}. \tag{30.38}$$

A corollary of this property is also that the structure constants become totally antisymmetric.

Construction 30.4.10 (Induced Killing form). Let $\mathfrak{g}$ be a Lie algebra and let V be a vector space equipped with a Lie algebra representation $\rho : \mathfrak{g} \to \mathrm{End}(V)$. One can define a Killing form associated with ρ in the following way:

$$K_\rho(x, y) := \mathrm{tr}\left(\rho(x) \circ \rho(y)\right). \tag{30.39}$$

This is a generalization of Definition 30.4.1 which reduces to the Killing form K when choosing V to be $\mathfrak{g}$ in the adjoint representation.

30.4.2 Weights, roots and Dynkin diagrams

From here on the base field is assumed to be algebraically closed ($\mathbb{C}$ for simplicity).

Definition 30.4.11 (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:

$$\exists n \in \mathbb{N} : \underbrace{[\mathfrak{h}, [\mathfrak{h}, [\mathfrak{h}, \ldots]]]}_{n \text{ times}} = 0. \tag{30.40}$$

2. **Self-normalizing**:

$$\forall x \in \mathfrak{h} : [x, y] \in \mathfrak{h} \implies y \in \mathfrak{h}. \tag{30.41}$$

From here on it will also be assumed that all Lie algebras are finite-dimensional. This assumption is motivated by the following property:

Property 30.4.12. Every finite-dimensional Lie algebra contains a Cartan subalgebra.

Property 30.4.13. If $\mathfrak{g}$ is semisimple, its Cartan subalgebra is Abelian.

Construction 30.4.14. 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 vectors $\{h_i\}_{i\leq k} \subset \mathfrak{g}$ such that $\forall i, j \leq k : [h_i, h_j] = 0$. (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 $i \in I$, then the algebra $\mathfrak{h} = \mathrm{span}\{h_i\}_{i\leq k}$ is a Cartan subalgebra.

Definition 30.4.15 (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:

$$V_\lambda := \{v \in V \mid \forall h \in \mathfrak{h} : h \cdot v = \lambda(H)v\}. \tag{30.42}$$

The nonzero elements of a weight space are called **weight vectors**. If the representation V can be decomposed as a direct sum of weigth spaces, it is called a **weight module**:

$$V = \bigoplus_{\lambda\in\mathfrak{h}^*} V_\lambda. \tag{30.43}$$

In the case where V is the adjoint representation, the (nonzero) weights are called **roots**:

Definition 30.4.16 (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 $h \in \mathfrak{h}$:

$$[h, g_j] = \alpha_j(h)g_j, \tag{30.44}$$

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 $\mathfrak{g}$:

$$\mathfrak{g} = \mathfrak{h} \oplus \bigoplus_{\lambda\in\Phi} \mathfrak{g}_\lambda, \tag{30.45}$$

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 30.4.17. 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 30.4.18 (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

$$\lambda = \pm \sum_i^n a_i\lambda_i, \tag{30.46}$$

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 certain root λ to be either all positive or all negative.

More generally one can define the following equivalence relation on a root system:

Definition 30.4.19 (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 30.4.20 (Triangular decomposition). Given a choice of positive roots Φ^+, one can decompose the Lie algebra $\mathfrak{g}$ as follows:

$$\mathfrak{g} = \mathfrak{n}_- \oplus \mathfrak{h} \oplus \mathfrak{n}_+, \tag{30.47}$$

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 $\mathfrak{g}$.

Property 30.4.21 (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 30.4.22 (Weyl group). For every simple root λ one can construct a Householder transformation 20.3.16 as follows:

$$\sigma_\lambda : \mathrm{span}_\mathbb{R}(\Delta) \to \mathrm{span}_\mathbb{R}(\Delta) : \mu \mapsto \mu - 2\frac{\langle\mu|\lambda\rangle}{\langle\lambda|\lambda\rangle}\lambda, \tag{30.48}$$

where the inner product $\langle\cdot|\cdot\rangle$ is the dual Killing form. The Weyl group W is defined as the group generated by all these transformations.

Property 30.4.23 (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 30.4.24 (Coroot). Consider the *sharp* map (34.2), 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

$$K^*(\cdot,\cdot) = K(\cdot^\sharp, \cdot^\sharp). \tag{30.49}$$

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):

$$\alpha^\vee := 2\frac{\langle\alpha|\cdot\rangle}{\langle\alpha|\alpha\rangle} \equiv 2\frac{\alpha}{\langle\alpha|\alpha\rangle}. \tag{30.50}$$

With this definition the Weyl transformations can be rewritten as follows:

$$\sigma_\lambda : \mathrm{span}_{\mathbb{R}}(\Delta) \to \mathrm{span}_{\mathbb{R}}(\Delta) : \mu \mapsto \mu - \mu(\lambda^\vee)\lambda. \tag{30.51}$$

Notation 30.4.25 (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 30.4.26 (Weyl chamber). Given a choice of positive roots Φ^+, the closed (fundamental) Weyl chamber associated to this ordering is defined as the subset $\mathcal{W} \subset \mathfrak{h}_0$ that contains the elements w satisfying the following equation for all $\gamma \in \Phi^+$:

$$w(H^\gamma) \geq 0. \tag{30.52}$$

Elements of this Weyl chamber are called **dominant weights**.

Property 30.4.27 (Weyl group). The Weyl group acts transitively on the set of Weyl chambers and, accordingly, on the orderings of the root system.

Property 30.4.28. 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 $\mathrm{span}\{E^\alpha, F^\alpha, [E^\alpha, F^\alpha]\}$ 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 30.4.29 (Cartan matrix). Let $\lambda_i, \lambda_j \in \Delta$ be simple roots. Because the Weyl group is the symmetry group of the root system,

$$\sigma_{\lambda_i}(\lambda_j) = \lambda_j - 2\frac{\langle\lambda_i|\lambda_j\rangle}{\langle\lambda_i|\lambda_i\rangle}\lambda_i$$

is a root. From the properties above it then follows that the quantity

$$C_{ij} := 2\frac{\langle\lambda_i|\lambda_j\rangle}{\langle\lambda_i|\lambda_i\rangle} = \lambda_j(H^{\lambda_i}) \tag{30.53}$$

is an integer. The matrix formed by these numbers is called the Cartan matrix.

Property 30.4.30 (Properties of Cartan matrix). The Cartan matrix C_{ij} 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 30.4.31 (Bond number). For all indices $i \neq j$ the bond number n_{ij} is defined as follows:

$$n_{ij} := C_{ij}C_{ji}. \tag{30.54}$$

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 from the definition).

Remark 30.4.32. 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 $\langle\lambda_i|\lambda_i\rangle < \langle\lambda_j|\lambda_j\rangle$ and $\langle\lambda_i|\lambda_i\rangle > \langle\lambda_j|\lambda_j\rangle$.

Construction 30.4.33 (Dynkin diagram). For a semisimple Lie algebra $\mathfrak{g}$ with simple roots Δ one can draw a so-called Dynkin diagram using the following rules:

1. For every simple root $\lambda \in \Delta$ draw a circle $\bigcirc$.
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).

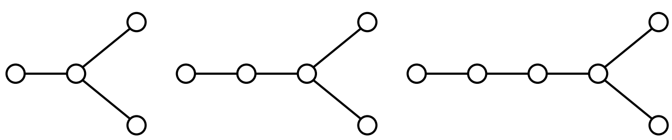
Property 30.4.34 (Classification). The Dynkin diagrams can be classified as follows (for every type the first three examples are given):

- 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 30.4.35 (Special linear group). By looking at the Lie brackets in Equation (30.14) 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 30.4.36 (Cartan & Killing). *Every finite-dimensional simple Lie algebra over $\mathbb{C}$ can be reconstructed from its set of simple roots Δ.*

Construction 30.4.37 (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 30.4.33. 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 30.4.38. 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 30.4.39 ($\mathfrak{sl}_2$). For every Cartan matrix A of rank 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.

30.4.3 Highest weight theory

Definition 30.4.40 (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 30.4.41 (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 $i, j \leq |\Delta|$:

$$\omega_i(H^{\alpha_j}) = \delta_{ij}. \tag{30.55}$$

This implies that an element $\lambda \in \mathfrak{h}_0$ is algebraically integral if it is an integral combination of fundamental weights.

Definition 30.4.42 (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:

$$\lambda \geq \mu \iff \lambda - \mu \in \mathrm{span}_{\mathbb{N}}\big(\Phi^+\big). \tag{30.56}$$

Definition 30.4.43 (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 30.4.44 (Highest weight theorem). *Let $\mathfrak{g}$ be a finite-dimensional Lie algebra. The following statements hold:*

- *Every finite-dimensional irreducible representation of $\mathfrak{g}$ 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.*

30.4.4 Coxeter groups

The properties about root systems in Section 30.4.2 can be abstracted into the following definition:

Definition 30.4.45 (Root system). Let $(E, \langle\cdot|\cdot\rangle)$ 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**: $E = \mathrm{span}_{\mathbb{R}}(\Phi)$.
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 ans simple roots, also carry over from the case of Lie algebras.

Remark 30.4.46 (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 30.4.47 (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 30.4.22.

The Weyl group induced by a root system is a specific instance of the following notion:

Definition 30.4.48 (Coxeter group). A group admitting a presentation of the form

$$\langle r_1, \ldots, r_n \mid (r_i r_j)^{m_{ij}} = e\rangle, \tag{30.57}$$

where $m_{ii} = 1$ and $m_{ij} \geq 2$ for $i \neq j$. If $m_{ij} = \infty$, no relation is imposed.

30.4.5 Kac-Moody algebras ♣

Construction 30.4.49 (Kac-Moody algebra). Consider the Cartan matrix A associated to a finite-dimensional (semi)simple Lie algebra. This matrix has the properties listed in 30.4.30. If one drops the positivity condition, the definition of a **generalized Cartan matrix** is obtained. For such a matrix one can construct a (possibly infinite-dimensional) Lie algebra using an analogue of the Chevalley-Serre relations.

Since a generalized Cartan matrix might have a vanishing determinant, one cannot use the Chevalley-Serre presentation as given in Construction 30.4.37, because the constructed roots might be linearly dependent. However, 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.[8]

Although an analogue of the Serre relations is not included, it can be shown that these relations still hold (see for example [115]). In fact when A is symmetrizable, the ideal $\mathfrak{m}$ is exactly generated by the Serre relations.

Remark 30.4.50 (Classification). There exist three distinct classes of Kac-Moody algebras (based on the definiteness of A):

- 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 fact, it can be shown for generalized Cartan matrices A 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 30.4.51 (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:

$$[x\otimes t^k,y\otimes t^l]:=[x,y]_{\mathfrak{g}}\otimes t^{k+l}. \tag{30.58}$$

Because this strongly resembles the definition of the ring of Laurent polynomials $K[t,t^{-1}]$ over a field K, the loop algebra is sometimes denoted by $\mathfrak{g}[t,t^{-1}]$.

[8] *Kac* proved that for symmetrizable matrices this construction is equivalent to the Chevalley-Serre presentation with the generators as given here.

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 8.1.12, then LG has the natural structure of a (infinite-dimensional) Lie group and its Lie algebra is precisely $L\mathfrak{g}$.

Definition 30.4.52 (Affine Lie algebra). Given a simple Lie algebra $\mathfrak{g}$, one can define the affine Lie algebra $\widehat{\mathfrak{g}}$ as the central extension of the loop algebra $L\mathfrak{g}$ by $\mathbb{C}$ associated to the cocycle

$$\Theta : (x \otimes t^k, y \otimes t^l) \mapsto kK(x,y)\delta_{k+l,0},$$

where $K(\cdot,\cdot)$ 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 $\widehat{\mathfrak{g}}$. This cocycle can also be defined using the residue 15.5.17 of a Laurent polynomial:

$$\Theta(f,g) := \mathrm{Res}[K(f,g)]_t, \tag{30.59}$$

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 will be denoted by c):

$$[d, x \otimes P(t)] = x \otimes t\frac{d}{dt}P(t) \tag{30.60}$$

$$[d, c] = 0, \tag{30.61}$$

where $P(t) \in \mathbb{C}[t, t^{-1}]$.

Definition 30.4.53 (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

$$[E_0, F_i] = 0 \tag{30.62}$$

$$[F_0, E_i] = 0 \tag{30.63}$$

for all $1 \leq i \leq n$. This also implies that $[E_0, F_0] =: H_0$ 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:

$$[H_0, E_0] = 2E_0 \tag{30.64}$$

$$[H_0, F_0] = -2F_0. \tag{30.65}$$

The extended Cartan matrix $\widehat{A}$ is defined by adjoining a new row and column to A. These new entries are defined by

$$[H_0, E_i] =: a_{0i}E_i \tag{30.66}$$

$$[H_i, E_0] =: a_{i0}E_0. \tag{30.67}$$

Definition 30.4.54 (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 30.4.49, is isomorphic to the affine Kac-Moody algebra as constructed above, starting from the simple Lie algebra $\mathfrak{g}(B)$, if and only if $A = \widehat{B}$.

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 30.4.55. Let $\mathfrak{g}$ be a simple Lie algebra with Cartan matrix A. The affine Lie algebra $\widehat{\mathfrak{g}}$ is isomorphic to the derived subalgebra $[\mathfrak{g}(\widehat{A}), \mathfrak{g}(\widehat{A})]$ of the Kac-Moody algebra $\mathfrak{g}(\widehat{A})$.

30.4.6 Universal enveloping algebra

Definition 30.4.56 (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 30.4.57. If the Chevalley-Serre presentation from Construction 30.4.37 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 30.4.58 (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 30.4.59 (Casimir invariant[9]). Let $\mathfrak{g}$ be a Lie algebra. A Casimir invariant is an element in the center of $U(\mathfrak{g})$.

Formula 30.4.60 (Quadratic Casimir invariant). Consider a Lie algebra representation $\rho:\mathfrak{g}\to\mathrm{End}(V)$ 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

$$\Omega_\rho := \sum_{i=1}^{\dim(\mathfrak{g})} \rho(X_i)\circ\rho(\xi_i), \tag{30.68}$$

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 30.4.10.

Property 30.4.61 (Casimir invariants of irreducible representations). When the representation $\rho:\mathfrak{g}\to\mathrm{End}(V)$ is irreducible, Schur's lemma 22.2.3 says that

$$\Omega_\rho = c_\rho\mathbb{1}_V. \tag{30.69}$$

By taking the trace of this formula and using 30.4.10, it can be seen that $c_\rho=\frac{\dim(\mathfrak{g})}{\dim(V)}$.

Definition 30.4.62 (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[10]:

$$V(\lambda) := U(\mathfrak{g})\otimes_{U(\mathfrak{b})}\mathbb{C}_\lambda, \tag{30.70}$$

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 PBW theorem 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 30.4.63. 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 $U(\mathfrak{g})/I_\lambda$.

The importance of Verma modules is given by the following property:

[9] Also known as a **Casimir operator** or **Casimir element**.

[10] This can be seen as an "extension of scalars"-procedure where a $U(\mathfrak{b})$-module is turned into a $U(\mathfrak{g})$-module.

Property 30.4.64 (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 30.4.65 (Basis of Verma module). A basis for $V(\lambda)$ is given by the monomials

$$F_{\alpha_1}^{m_1} F_{\alpha_2}^{m_2} \cdots F_{\alpha_N}^{m_N} v_\lambda, \tag{30.71}$$

where v_λ is the highest weight vector, α_i are negative roots, $m_i \in \mathbb{N}$ and $F_{\alpha_i} \in \mathfrak{g}_{\alpha_i}$.

30.5 Group contractions

Definition 30.5.1 (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:

$$U = u + \varepsilon w.$$

The group contraction is obtained by taking the limit $\varepsilon \to 0$. 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 30.5.2 (Galilei group). The *Galilei group* in d 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.

30.6 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 is only valid for finite-dimensional algebras):

Definition 30.6.1 (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 let $\{t^a\}_{a\leq\dim(\mathfrak{g})}$ be its linear dual. The Chevalley-Eilenberg algebra $\mathrm{CE}(\mathfrak{g})$ is defined as the Grassmann algebra $\Lambda^\bullet\mathfrak{g}^*$ with a dg-algebra structure induced by the differential[11]

$$dt^a := -\frac{1}{2}c^a_{bc} t^b \wedge t^c, \tag{30.72}$$

where c^a_{bc} are the structure constants of $\mathfrak{g}$.

By analogy with the case of group (co)homology as in Section 5.4.2, 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

[11] In Section 27.7.2 it was explained how this differential can be obtained as the dual of the Lie bracket.

the construction are defined with respect to the trivial $U(\mathfrak{g})$-module K (the underlying field of the Lie algebra). This gives:

$$H^i_{\text{Lie}}(\mathfrak{g};M) := \text{Ext}^i_{U(\mathfrak{g})}(K,M) \tag{30.73}$$

$$H_i^{\text{Lie}}(\mathfrak{g};M) := \text{Tor}_i^{U(\mathfrak{g})}(K,M), \tag{30.74}$$

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 K to determine the Ext-functor in terms of hom-sets $\text{Hom}(\cdot,M)$. It can be shown that such a resolution is given by the tensor product $U(\mathfrak{g})\otimes\Lambda^\bullet\mathfrak{g}$:

$$\text{Ext}^i_{U(\mathfrak{g})}(K,M) = H^i(\text{Hom}_{\mathfrak{g}}(U(\mathfrak{g})\otimes_K\Lambda^\bullet\mathfrak{g},M)) \cong H^i(\text{Hom}_K(\Lambda^\bullet\mathfrak{g},M)), \tag{30.75}$$

where the differential of the (middle) complex is given by

$$\begin{aligned} d(u\otimes g_1\wedge\cdots\wedge g_n) := &\sum_i(-1)^{i+1}ug_i\otimes(g_1\wedge\cdots\wedge\hat{g}_i\wedge\cdots\wedge g_n) \\ &+\sum_{i<j}(-1)^{i+j}u\otimes[g_i,g_j]\wedge\cdots\wedge\hat{g}_i\wedge\cdots\wedge\hat{g}_j\wedge\cdots\wedge g_n, \end{aligned} \tag{30.76}$$

where as usual the caret $\hat{}$ indicates the omission of a factor. In the case $M=K$, the hom-complex can easily be seen to be the Chevalley-Eilenberg algebra $\text{CE}(\mathfrak{g})\cong\Lambda^\bullet\mathfrak{g}^*$ (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 $(-1)^n dm\otimes\omega$ with $dm(g) := g\cdot m$.

Property 30.6.2 (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:

$$H^0(\mathfrak{g};M) = \{m\in M \mid \mathfrak{g}\cdot m = m\}. \tag{30.77}$$

The first cohomology group with coefficients in K is isomorphic to the quotient of $\mathfrak{g}$ by its first derived ideal:

$$H^1(\mathfrak{g}) \cong \mathfrak{g}/[\mathfrak{g},\mathfrak{g}]. \tag{30.78}$$

Property 30.6.3 (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 30.6.4 (Classification of central extensions). The 2-cocycles (with values in K) from Construction 30.2.41 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 k are classified by $H^2(\mathfrak{g})$.

Corollary 30.6.5. Semisimple Lie algebras do not admit nontrivial central extensions.

Definition 30.6.6 (Weil algebra). Consider a Lie algebra $\mathfrak{g}$. Its Weil algebra is defined as the dg-algebra $\Lambda^\bullet(\mathfrak{g}^*\oplus\mathfrak{g}^*[1])$ with differential $d_W := d_{\text{CE}}+\Pi$, where d_{CE} is the differential on the Chevalley-Eilenberg subalgebra $\text{CE}(\mathfrak{g})\subset W(\mathfrak{g})$ and Π shifts the degree by 1. The action of d_{CE} on shifted generators is defined through the relation $[\Pi,d_{\text{CE}}]=0$.

Definition 30.6.7 (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

$$0 \to \ker(p) \to W(\mathfrak{g}) \overset{p}{\twoheadrightarrow} \mathrm{CE}(\mathfrak{g}) \to 0, \tag{30.79}$$

where p is the obvious projection map. An important subpace of $\ker(p)$ is given by the algebra of invariant polynomials $\mathrm{inv}(\mathfrak{g})$:

Definition 30.6.8 (Invariant polynomial). A horizontal element ω for which $d\omega$ is also horizontal. (Sometimes the horizontality condition is replaced by $d\omega = 0$.)

It should be noted that although this definition might seem complicated, it is (for ordinary Lie algebras[12]) equivalent to the definition in terms of Ad-invariant polynomials:

Alternative Definition 30.6.9 (Invariant polynomial). Let G be a Lie group with Lie algebra $\mathfrak{g}$. A polynomial $P \in K[\mathfrak{g}]$, where K is the base field, is said to be invariant (or Ad-invariant) if

$$P(x) = P(gxg^{-1}) \tag{30.80}$$

for all $x \in \mathfrak{g}$ and $g \in G$. This subalgebra of $K[\mathfrak{g}]$ is denoted by $K[\mathfrak{g}]^G$.

A concept that will be important later on in the study of characteristic classes on fibre bundles is the transgression map:

Definition 30.6.10 (Transgression). The exact sequence (30.79) induces a long exact sequence in cohomology. An invariant polynomial is said to be in transgression with a cocycle in $\mathrm{CE}(\mathfrak{g})$ if their cohomology classes are related by the connecting morphism. More explicitly, by definition of invariant polynomials, one has $d\omega = 0$ and, since $W(\mathfrak{g})$ has vanishing cohomology, there exists an element c_ω such that $\omega = dc_\omega$. By restricting c_ω to $\mathrm{CE}(\mathfrak{g})$, one obtains a $\mathfrak{g}$-cocycle, since $d_{\mathrm{CE}}c_\omega = 0$.

Example 30.6.11 (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 $\langle\cdot,[\cdot,\cdot]\rangle$.

[12]The above definition leads to a straightforward generalization in the context of L_∞-algebras.

Chapter 31

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 he wants to work, e.g. topological space, topological group and continuous map in the case of **Top**.

31.1 Bundles

Definition 31.1.1 (Bundle). A triple (E, B, π) where E and B are spaces and π is a morphism. Sometimes the map π is also required to be surjective. However, under this additional restriction one cannot make the association $\mathbf{Bundle}(B) \cong \mathbf{C}/B$ of categories anymore.

An explicit example in the category **Diff** is the following:

Example 31.1.2 (Fibred manifold). A surjective submersion 29.3.8

$$\pi : E \to B,$$

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 31.1.3 (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 following **Čech cocycle condition**:

$$g_{ij} = g_{ik} \circ g_{kj}. \tag{31.1}$$

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

$$\lambda_{i,r} g_{ij} \lambda_{j,s}^{-1} = h_{rs} \tag{31.2}$$

whenever this is well-defined. The resulting quotient set is denoted by $\check{H}^1(B; G)$.[1]

[1] The notation stems from the fact that this is the first Čech cohomology group with values in G (Section 9.3.2).

Property 31.1.4 (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 31.1.5 (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:

$$\varphi_j \circ \varphi_i^{-1}(b, x) = (b, g_{ji}(b) \cdot x). \tag{31.3}$$

The compatibility conditions satisfied by the functions g_{ji}, obtained by considering triple intersections, are exactly the cocycle conditions (31.1). Moreover, it can be shown that this action of G on every fibre is faithful 3.3.11.

Remark 31.1.6. One should pay attention to the fact that the bundle charts are not coordinate charts in the sense of manifolds 29.1.1 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 31.1.7. A fibre bundle (E, B, π, F, G) is often denoted by $F \hookrightarrow E \xrightarrow{\pi} B$ or even $\pi : E \to B$ if the fibre is not important. A drawback of such notations is that the structure group of the bundle is not shown.

Definition 31.1.8 (Numerable fibre bundle). A fibre bundle that admits a local trivialization over a numerable open cover 7.5.19.

Definition 31.1.9 (Compatible bundle charts). A bundle chart (V, ψ) is said to be **admissable** 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

$$\psi \circ \varphi_i^{-1}(b, x) = (b, h_i(b)x) \tag{31.4}$$

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.

[2]This terminology follows from the fact that the bundle is locally isomorphic to a (trivial) product space: $E \cong U \times F$.

Definition 31.1.10 (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 31.1 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 31.1: Bundle map between fibre bundles.

Definition 31.1.11 (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

$$\lambda_i := \varphi'_i \circ \varphi_i^{-1}. \tag{31.5}$$

Property 31.1.12 (Isomorphism). Two fibre bundles over the same base space are equivalent if and only if they are isomorphic.

Definition 31.1.13 (Trivial bundle). A fibre bundle (E, B, π, F) is said to be trivial if there exists an equivalence $E \cong B \times F$.

31.1.1 Constructions

Construction 31.1.14 (Fibre bundle 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 31.1.3 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 $T := \bigsqcup_{i\in I} U_i \times F$ and equip it with the disjoint union topology 7.3.2.
3. From this disjoint union construct a quotient space, equipped with the quotient space topology 7.3.3, induced by the following equivalence relations for all $i, j \in I$:

$$(b, f) \sim (b, g_{ji}(b) \cdot f), \tag{31.6}$$

   where $b \in U_i \cap U_j$ and $f \in F$ (note that the disjoint union indices are suppresed in this notation).

The fibre bundle E is equal to this quotient space, where the projection π is the quotient space projection $\pi : E \to B : [(b, f)] \mapsto b$, where $[A]$ denotes the equivalence class of A in E. Local trivializations are given by the maps $\varphi_i : \pi^{-1}(U_i) \to U_i \times F$ that satisfy

$$\varphi_i^{-1} : (b, f) \mapsto [(b, f)]. \tag{31.7}$$

[3]Sometimes also called the **clutching theorem**, see below for an explanation.

Remark 31.1.15 (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 31.1.16 (Homotopy invariance). Homotopic transition functions give rise to equivalent bundles.

Remark 31.1.17 (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**.

Definition 31.1.18 (Pullback bundle). Let $\pi : E \rightarrow B$ be a fibre bundle and let $f : B' \rightarrow B$ be a morphism of spaces. The pullback 4.4.55 of the bundle projection and f gives the total space of the pullback bundle f^*E:

$$f^*E := \big\{(b', e) \in B' \times E \,\big|\, f(b') = \pi(e)\big\}. \tag{31.8}$$

The topology on f^*E is induced by the subspace topology of the product $B' \times E$. The projection onto the second factor gives a map of total spaces $f^*E \rightarrow E$.

Definition 31.1.19 (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:

$$E_1 \times_B E_2 := \big\{(p, q) \in E_1 \times E_2 \,\big|\, \pi_1(p) = \pi_2(q)\big\}. \tag{31.9}$$

It is the pullback of one bundle along the projection of another bundle.

31.1.2 Sections

Definition 31.1.20 (Section). A (**global**) section of a fibre bundle $\pi : E \rightarrow B$ is a morphism $s : B \rightarrow E$ such that $\pi \circ s = \mathbb{1}_B$, i.e. it is a section of π in the sense of Definition 4.4.1. For any open subset $U \subset B$, a **local** section is defined as a morphism $s_U : U \rightarrow E$ such that $\pi \circ s_U = \mathbb{1}_U$.

Notation 31.1.21. 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 $\Gamma(E) \equiv \Gamma(B, E)$.

Property 31.1.22 (Pullback of sections). Consider a fibre bundle $\pi : E \rightarrow B$ together with a morphism of spaces $f : B' \rightarrow B$. The sections of E pullback to the pullback bundle f^*E by defining $f^*s := s \circ f$.

Chapter 32

Vector Bundles

The main reference for this chapter is [34].

32.1 Tangent bundle

The tangent space, as introduced in Section 29.2, can also be introduced in a more abstract/general way. Because it is the most important example of a vector bundle, tangent bundles are introduced first.

Construction 32.1.1 (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 an associated smooth function on TU, called the **differential** or **derivative** of f, by

$$Tf : U \times \mathbb{R}^n \to f(U) \times \mathbb{R}^n : (p, v) \mapsto (f(p), Df(p)v), \tag{32.1}$$

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

$$\widetilde{\psi}_{ji} := T\psi_{ji} : TU_i \to TU_j$$

given by

$$\widetilde{\psi}_{ji}(\varphi_i(p), v) := \big(\varphi_j(p), D(\varphi_j \circ \varphi_i^{-1})(\varphi_i(p))v\big). \tag{32.2}$$

Because the transition functions are diffeomorphisms, the associated Jacobians are invertible. This implies that the maps $\widetilde{\psi}_{ji}$ are elements of $\mathrm{GL}(\mathbb{R}^n)$. The tangent bundle is then obtained by applying the fibre bundle construction theorem 31.1.14 to the triple $(M, \mathbb{R}^n, \mathrm{GL}(\mathbb{R}^n))$ together with the cover $\{U_i\}_{i\leq n}$ and the cocycle $\{\widetilde{\psi}_{ji}\}_{i,j\leq n}$.

Definition 32.1.2 (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 32.1.3 (Tangent space). Consider a point $p \in M$. The definition of the tangent space in the above setting is given by the fibre

$$T_pM := \pi_{TM}^{-1}(p). \tag{32.3}$$

If one uses the natural charts to map T_pM to 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 32.1.4 (Smooth structure). An atlas on TM is given by the charts (TU_i, θ) with

$$\theta : TM \to \mathbb{R}^{2n} : (p, X) \mapsto (\varphi_i \circ \pi(p), X^1, \ldots, X^n), \tag{32.4}$$

where (U_i, φ_i) is a chart around $p \in M$ such that X can be expressed as $X^i \partial_i \in T_pM$.

Property 32.1.5 (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

$$\dim(TM) = 2\dim(M). \tag{32.5}$$

Remark 32.1.6 (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 32.1.7 (Differential). The map T defined in Equation (32.1) can be generalized to arbitrary smooth manifolds as the map $Tf : TM \to TN$, where the Jacobian now acts on the (linear) fibres. 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 32.1.8. 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 on the category of smooth manifolds. One can view T as a "functorial derivative".

Definition 32.1.9 (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 20.2.11.

Theorem 32.1.10 (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 32.1.11 (Parallelizable manifold). A manifold with a trivial tangent bundle.

Definition 32.1.12 (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 32.1.13 (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 32.1.14 (Tubular neighbourhood theorem). *Every embedded submanifold admits a tubular neighbourhood, namely its normal bundle. Furthermore, all tubular neighbourhoods are diffeomorphic.*

Corollary 32.1.15 (Submanifolds and NDR pairs). Consider a smooth manifold M with a submanifold S. The pair (M, S) is an NDR pair 8.2.33. 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.

32.2 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 32.2.1 (Vector bundle). Consider a manifold B with atlas $\{(U_i, \varphi_i)\}_{i\in I}$ together with a cocycle $\{g_{ji} : U_i \cap U_j \to G\}_{i,j\in I}$ with values in a group G and a representation $\rho : G \to \mathrm{GL}(V)$ on a (finite-dimensional) vector space V. This data can be used to construct a fibre bundle through Construction 31.1.14. The dimension of the typical fibre V is called the **rank** of the vector bundle.

Remark 32.2.2. 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 32.2.3 (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 57.3 on geometric quantization for more information.)

Property 32.2.4 (Vector bundles over a sphere). The clutching theorem 31.1.17 and the homotopy invariance imply that vector bundles over the sphere are determined by homotopy classes of functions $S^{n-1} \to \mathrm{GL}_p(K)$, i.e. they are classified by the homotopy group $\pi_{n-1}(\mathrm{GL}_p(K))$.

32.2.1 Sections

Definition 32.2.5 (Frame). A frame of a vector bundle E is a tuple $(s_1, \ldots, s_n)$ of sections such that $(s_1(b), \ldots, s_n(b))$ is a basis for the fibre $\pi^{-1}(b)$ for all $b \in B$.

Property 32.2.6 (Trivial bundles). A vector bundle is trivial if and only if there exists a global frame.

Theorem 32.2.7 (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 X is a finitely-generated projective $C(X)$-module.*

Property 32.2.8 (Zero section). The zero section s_0 of a vector bundle $E \to B$ is the map that assigns to every point $b \in B$ the zero vector of the associated vector space 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.

32.2.2 Sums and product

Definition 32.2.9 (Whitney sum). Consider two vector bundles E, E' with typical fibres W, W' over the same base manifold. One can construct a new vector bundle $E \oplus E'$ by taking the typical fibre to be the direct sum $W \oplus W'$, i.e. the fibre over p is given by $W_p \oplus W'_p$. This operation is called the Whitney sum or **direct sum** of vector bundles.

The existence property for complements 20.1.20 from linear algebra can be generalized in the following way:

Property 32.2.10. 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 32.2.11. 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 32.2.12 (Stable isomorphism). Two vector bundles E, E' over a base manifold B are said to be stably isomorphic if there exist integers $m, n \in \mathbb{N}$ such that

$$E \oplus (B \times \mathbb{R}^m) \cong E' \oplus (B \times \mathbb{R}^n). \tag{32.6}$$

Construction 32.2.13 (Tensor product). The tensor product $E \otimes E' \to M$ of two vector bundles $\pi : E \to M$ and $\pi' : E' \to M$ is given by the fibrewise tensor product.

Construction 32.2.14 (Exterior tensor product). The exterior tensor product $E \boxtimes E' \to M \times N$ of two vector bundles $\pi : E \to M$ and $\pi' : E' \to N$ is given by the tensor product of fibres over the Cartesian product $M \times N$.

32.2.3 Associated vector bundles

Construction 32.2.15 (Associated vector bundle). Consider a representation

$$\rho : \mathrm{GL}(\mathbb{R}^n) \to \mathrm{GL}(\mathbb{R}^l)$$

together with the tangent bundle cocycle $\{t_{ji} := D(\psi_{ji}) \circ \varphi_i\}_{i,j \leq n}$. The composite

$$\rho \circ t_{ji} : U_i \cap U_j \overset{t_{ji}}{\to} \mathrm{GL}(\mathbb{R}^n) \overset{\rho}{\to} \mathrm{GL}(\mathbb{R}^l)$$

is again a cocycle and can thus be used to define a new vector bundle on M through the fibre bundle construction theorem. The vector bundle $E \equiv \rho(TM)$ is called the associated (vector) bundle of the tangent bundle induced by ρ.

Example 32.2.16 (Contravariant vectors). By noting that the k^{th} 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 32.2.17 (Cotangent bundle). Another useful construction is given by the contragredient representation $A \mapsto (\rho^T)^{-1} = (\rho^{-1})^T$. 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 32.2.18. 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 32.2.19 (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 32.2.20 (Determinant line bundle). Consider a rank-n vector bundle $\pi : E \to B$. The determinant map induces an associated line bundle $\det(\pi) : \det(E) := \bigwedge^n E \to B$ 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 are called **densitized bundles**.

Example 32.2.21 (Canonical bundle). Consider a smooth n-dimensional manifold M. The canonical (line) bundle of M is given by $\det(T^*M) \equiv \bigwedge^n T^*M$, i.e. the determinant line bundle of the cotangent bundle of M.

32.2.4 Grassmann bundle

Looking at Property 20.8.2 and noting that $\mathrm{GL}_n(\mathbb{R})$ is a Lie group, it is clear that one can endow the Grassmannian $\mathrm{Gr}(k, \mathbb{R}^n)$ from Definition 20.8.1 with a differentiable structure, thereby turning it into a smooth manifold. This allows the construction of a new bundle by applying the fibre bundle construction theorem. Because the Grassmannian is not a vector space, the resulting bundle will be a general fibre bundle and not a vector bundle.

Construction 32.2.22 (Grassmann bundle). Define a new set of transition functions

$$\psi_{ji} : (\varphi_i(p), V) \mapsto (\varphi_j(p), t_{ji}(p) \cdot V), \tag{32.7}$$

where $\{t_{ji}\}_{i,j\leq n}$ is the tangent bundle cocycle. These transition functions can be used to create a new fibre bundle with typical fibre $\mathrm{Gr}(k, \mathbb{R}^n)$. The fibre over a point $p \in M$ is the Grassmannian $\mathrm{Gr}(k, T_pM)$ 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 32.2.23. The Grassmann k-plane bundle of a vector bundle E is denoted by $\mathrm{Gr}(k, E)$.

Definition 32.2.24 (Tautological bundle). Consider the Grassmannian $\mathrm{Gr}(n, V)$ 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 $W \in \mathrm{Gr}(n, V)$ and $w \in W$. Local trivializations are constructed as follows:

$$\varphi_Z : \pi^{-1}(U) \to \mathrm{Gr}(n, V) \times Z : (W, w) \mapsto (W, \mathrm{pr}_Z(w)), \tag{32.8}$$

where pr_Z is the orthogonal projection onto the subspace $Z \in \mathrm{Gr}(n, V)$. This bundle inherits a natural vector bundle structure from V.

Definition 32.2.25. Consider the tautological line bundle J over a projective space $\mathbb{CP}^n \cong \mathrm{Gr}(1, \mathbb{C}^n)$. The dual line bundle $\mathrm{Hom}(J, \mathbb{C})$ is often denoted by $\mathscr{O}_{\mathbb{CP}^n}(1)$ or $\mathscr{O}(1)$ and is sometimes called **Serre's twisting sheaf**. Tensor powers of this bundle are accordingly denoted by $\mathscr{O}_{\mathbb{CP}^n}(k)$. To also allow for factors of J one can extend the notation to negative indices: $\mathscr{O}_{\mathbb{CP}^n}(-k)$, e.g. the tautological bundle is denoted by $\mathscr{O}_{\mathbb{CP}^n}(-1)$.

Property 32.2.26 (Euler sequence). The twisting sheaf $\mathscr{O}_{\mathbb{CP}^n}(1)$ fits in a short exact sequence

$$0 \longrightarrow \underline{\mathbb{C}} \longrightarrow \mathscr{O}_{\mathbb{CP}^n}(1)^{\oplus(n+1)} \longrightarrow T\mathbb{CP}^n \longrightarrow 0, \tag{32.9}$$

where $\underline{\mathbb{C}}$ denotes the trivial line bundle.

Sketch of proof. Note that vector fields on $\mathbb{CP}^n$ can be obtained by pairing (coordinate-induced) basis vectors ∂_i on $\mathbb{C}^{n+1}$ with linear functions on $\mathbb{C}^{n+1}$, i.e. with sections of $\mathscr{O}_{\mathbb{CP}^n}(1)$. The kernel of this map is given by the Euler vector field $\mathbb{E} := x^i\partial_i$ and its scalar multiples (Definition 14.6.12).

32.3 Vector fields

From here on the theory will be specialized to the smooth setting, i.e. all manifolds will be assumed to be smooth and all morphism will be required to be smooth (unless stated otherwise).

Definition 32.3.1 (Vector field). A section $s \in \Gamma(TM)$ of the tangent bundle. By the Serre-Swan theorem 32.2.7 the set of vector fields forms a $C^\infty(M)$-module.

Notation 32.3.2. The set of all vector fields on a manifold M is often denoted by $\mathfrak{X}(M)$.

Definition 32.3.3 (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 $\mathrm{ind}_X(p)$ of X at p is defined as the degree 8.2.38 of the function

$$f : S^{n-1} \to S^{n-1} : m \mapsto \frac{X(m)}{\|X(m)\|}.$$

Property 32.3.4 (Winding number). The winding number of a vector field X along a curve γ (where it is assumed that X does not vanish on γ) is equal to the sum of indices of zeroes of X lying inside γ.

Theorem 32.3.5 (Poincaré-Hopf). *Let M be a compact manifold and consider a vector field X having only isolated zeroes. The Euler characteristic 8.2.18 is given by*

$$\chi(M) = \sum_{p \in X^{-1}(0)} \mathrm{ind}_X(p). \tag{32.10}$$

Remark 32.3.6. This is essentially a restatement of Property 29.3.20, where one submanifold is given by the zero section and the other is given by the graph of X.

An immediate consequence of the Poincaré-Hopf theorem is the following well-known result:

Theorem 32.3.7 (Hairy ball theorem). *There exists no nowhere-vanishing vector field on an even-dimensional sphere S^{2n}.*

Definition 32.3.8 (Pullback). Let X be vector field on N and let $\varphi : M \to N$ be a diffeomorphism. The pullback of X along φ is defined as

$$(\varphi^* X)_p := T\varphi^{-1}(X_{\varphi(p)}). \tag{32.11}$$

Definition 32.3.9 (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:

$$(\varphi_* X)_{\varphi(p)} := T\varphi(X_p). \tag{32.12}$$

This can be rewritten using the pullback as follows:

$$\varphi_* X = (\varphi^{-1})^* X. \tag{32.13}$$

Equivalently, one can define a vector field on N by

$$(\varphi_* X)_q(f) := X_{\varphi^{-1}(q)}(f \circ \varphi) \tag{32.14}$$

for all smooth functions $f : N \to \mathbb{R}$ and points $q \in N$.

32.3.1 Integral curves

Definition 32.3.10 (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

$$\gamma'(t) = X(\gamma(t)) \tag{32.15}$$

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 18.2.1, 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 32.3.11 (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

$$\sigma_t(p) := \gamma_p(t), \tag{32.16}$$

is called the flow of X at time t. The **flow domain** is defined as the set

$$D(X) := \big\{(t,p) \in \mathbb{R} \times M \,\big|\, p \in M, t \in \,]a_p, b_p[\,\big\}, \tag{32.17}$$

where $]a_p, b_p[$ is the maximal interval on which γ_p is defined.

Property 32.3.12. 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 smooth, bijective 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 32.3.13 (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 32.3.14. If the manifold M is compact, every vector field $X \in \mathfrak{X}(M)$ is complete.

Property 32.3.15 (Winding number). The winding number of a vector field along a closed integral curve is 1.

32.3.2 Lie derivative

Formula 32.3.16 (Lie derivative for smooth functions). Let $X \in \mathfrak{X}(M)$ and let $f \in C^\infty(M)$. The Lie derivative of f with respect to X at $p \in M$ is defined as

$$\mathcal{L}_X f(p) := \lim_{t\rightarrow 0} \frac{f(\gamma_p(t)) - f(p)}{t}. \tag{32.18}$$

The definition of the Lie derivative closely resembles the definition of the ordinary derivative on Euclidean space. This is not a coincidence:

Formula 32.3.17. Working out the definition of the Lie derivative and rewriting it as an operator equality gives

$$\mathcal{L}_X = \sum_k X_k \frac{\partial}{\partial x^k}. \tag{32.19}$$

It is clear that this is just the vector field X expanded in the basis 29.2.3. This way, one also recovers the behaviour of a tangent vector as a derivation. For smooth functions $f : M \rightarrow \mathbb{R}$ this gives

$$\mathcal{L}_X f(p) = X_p(f). \tag{32.20}$$

Explanation. In this derivation Landau's little-o notation is used:

$$\lim_{t\to 0}\frac{o(t)}{t} = 0. \tag{32.21}$$

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 $p \in U$:

$$\begin{aligned}
\mathcal{L}_X f(p) &= \lim_{t\to 0}\left[\frac{f(\beta_p(0) + t\beta_p'(0) + o(t)) - f(p)}{t}\right]\\
&= \lim_{t\to 0}\left[\frac{f(p + tX(p) + o(t)) - f(p)}{t}\right]\\
&= \lim_{t\to 0}\left[\frac{f(p) + tDf(p)\cdot X(p) + o(t) - f(p)}{t}\right]\\
&= \sum_k \frac{\partial f}{\partial x^k}(p)X_k(p) + \lim_{t\to 0}\frac{o(t)}{t}\\
&= \sum_k \frac{\partial f}{\partial x^k}(p)X_k(p),
\end{aligned} \tag{32.22}$$

where the defining condition 32.3.10 for integral curves was used on the second line. If this equation is rewritten as an operator equality, one obtains

$$\mathcal{L}_X = \sum_k X_k \frac{\partial}{\partial x^k}. \tag{32.23}$$

[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.

Formula 32.3.18 (Lie derivative for vector fields). Let $X, Y \in \mathfrak{X}(M)$.

$$\mathcal{L}_X Y(p) := \frac{d}{dt}(\sigma_t^* X)(\gamma_p(t))\bigg|_{t=0}. \tag{32.24}$$

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 32.3.11:

$$\mathcal{L}_X Y = \lim_{t\to 0}\frac{(T\sigma_t)^{-1}X(\gamma_p(t)) - X(p)}{t}, \tag{32.25}$$

where $T\sigma_t$ is the differential 32.1.7 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 32.3.11 was used. One can also rewrite the second term in the numerator of (32.25) using the flow:

$$X(p) = X(\sigma_0(p)) = T\sigma_0(X).$$

Using the definition of the pushforward of vector fields (32.12), the Lie derivative can be rewritten as follows:

$$\begin{aligned}\mathcal{L}_X Y &= \lim_{t\to 0} \frac{\sigma_{-t*}X(\gamma_p(t)) - \sigma_{0*}X(\gamma_p(0))}{t} \\ &= \frac{d}{dt}(\sigma_{-t*}X)(\gamma_p(t))\bigg|_{t=0}.\end{aligned}$$

Finally, by using the relation between pushforward and pullback (32.13) this becomes

$$\mathcal{L}_X Y = \frac{d}{dt}(\sigma_t^* X)(\gamma_p(t))\bigg|_{t=0}. \tag{32.26}$$

Property 32.3.19. Let $X, Y \in \mathfrak{X}(M)$ be vector fields of class C^k. The Lie derivative has the following properties:

- $\mathcal{L}_X Y$ is a vector field.
- **Lie bracket**: The Lie derivative of vector fields coincides with the commutator:

$$\mathcal{L}_X Y = [X, Y]. \tag{32.27}$$

  The fact that this is indeed a derivation on $C^{k-1}(M,\mathbb{R})$ follows from Schwarz's theorem 14.6.10. This result shows that the Lie derivative on vector fields turns the space $\mathfrak{X}(M)$ into a (real) Lie algebra.
- The previous point also implies that the Lie derivative is antisymmetric:

$$\mathcal{L}_X Y = -\mathcal{L}_Y X. \tag{32.28}$$

Definition 32.3.20 (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:

$$\mathcal{L}_{e_i} e_j = 0. \tag{32.29}$$

Equivalently, a basis is holonomic if the associated structure coefficients of the Lie algebra $\mathfrak{X}(M)$ vanish on U.

Property 32.3.21. For every holonomic basis there exists a coordinate system on M such that the basis coincides with the coordinate-induced basis.

32.4 Differential k-forms

Definition 32.4.1 (Differential form). A differential k-form is a map

$$\omega : T^k M \to \mathbb{R} \tag{32.30}$$

such that the restriction of ω to each fibre of the bundle $T^k M$ is multilinear and antisymmetric. 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. The space $\Omega^0(M)$ is defined as the space of smooth functions $C^\infty(M)$.

Differential forms can also be constructed as sections of an associated vector bundle:

Alternative Definition 32.4.2. Consider the representation

$$\rho_k : \mathrm{GL}(\mathbb{R}^{m*}) \to \mathrm{GL}(\Lambda^k\mathbb{R}^{m*}) : A \mapsto A \wedge \cdots \wedge A.$$

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 $\rho_k(\pi_{T^*M})$:

$$\Omega^k(M) := \Gamma(\rho_k(\pi_{T^*M})). \tag{32.31}$$

Construction 32.4.3 (Exterior algebra). One can construct a Grassmann algebra 21.4.20 by equipping the graded vector space

$$\Omega^\bullet(M) := \bigoplus_{k\geq 0}\Omega^k(M) \tag{32.32}$$

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 32.4.4 (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

$$f^*(\omega) := \omega \circ f_*. \tag{32.33}$$

This defines a map $f^* : \Omega^\bullet(N) \to \Omega^\bullet(M)$.

Definition 32.4.5 (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

$$f_*(\omega) := \omega \circ (f^{-1})_*. \tag{32.34}$$

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 32.4.6 (Dual basis). Consider the coordinate basis from Definition 29.2.3 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:

$$\left\langle \frac{\partial}{\partial x^i}, dx^j \right\rangle = \delta_i^j. \tag{32.35}$$

It should be noted that dx^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.

32.4.1 Exterior derivative

Definition 32.4.7 (Exterior derivative). The exterior derivative d_k is a morphism constructed on the graded algebra of differential k-forms:

$$d_k : \Omega^k(M) \to \Omega^{k+1}(M). \tag{32.36}$$

For $k = 0$ it is defined by

$$df := \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i \tag{32.37}$$

The object $df \in \Omega^1(M)$ is often called the **differential** of f. This formula can be generalized to higher degree forms as follows:

$$d(f dx_{i_1} \wedge \cdots \wedge dx_{i_k}) := df \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_k}. \tag{32.38}$$

Remark 32.4.8. Equation (32.37) 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 dx^i are not infinitesimal quantities but vectors of unit norm.

Property 32.4.9. The exterior derivatives satisfy the following properties for all $k \geq 0$:

- **Nilpotency**: For all $\omega \in \Omega^k(M)$:

$$d_k \circ d_{k+1} = 0. \tag{32.39}$$

- **Linearity**: d_k is an $\mathbb{R}$-linear map.

These two items say that $(\Omega^\bullet(M), d)$ is not just a graded algebra, but in fact a dg-algebra 5.1.9.

- **Graded Leibniz rule**: (hence d is a graded derivation):

$$d(\omega_1 \wedge \omega_2) = d\omega_1 \wedge \omega_2 + (-1)^j \omega_1 \wedge d\omega_2, \tag{32.40}$$

  where $\omega_1 \in \Omega^j(M)$ and $\omega_2 \in \Omega^k(M)$.

- **Naturality**: If $f \in C^\infty(M)$, then $f^*(d\omega) = d(f^*\omega)$.

Example 32.4.10. Let $f \in C^\infty(M, \mathbb{R})$ and let γ be a curve on M. From Definition 32.4.6 of the basis $\{dx_k\}_{k\leq n}$ one obtains the following result:

$$\langle df(x), \gamma'(t)\rangle = \sum_k \frac{\partial f}{\partial x_k}(x)\gamma_k'(t) = (f \circ \gamma')(t). \tag{32.41}$$

Example 32.4.11. An explicit formula for the exterior derivative of a k-form Φ is

$$\begin{aligned} d\Phi(X_1, \ldots, X_{k+1}) = &\sum_{i=1}^{k+1} (-1)^{i+1} X_i(\Phi(X_1, \ldots, \hat{X}_i, \ldots, X_{k+1})) \\ &+ \sum_{i<j} (-1)^{i+j} \Phi([X_i, X_j], X_1, \ldots, \hat{X}_i, \ldots, \hat{X}_j, \ldots, X_{k+1}), \end{aligned} \tag{32.42}$$

where $\hat{X}$ indicates that this argument is omitted.

32.4.2 Lie derivative

Formula 32.4.12 (Lie derivative of smooth functions). Using the definition of the exterior derivative of smooth functions (32.37) and the definition of the dual basis 32.4.6, one can rewrite the Lie derivative 32.3.16 as

$$Xf(p) = df_p(X(p)). \tag{32.43}$$

Formula 32.4.13 (Lie derivative of differential forms).

$$\mathcal{L}_X\omega(p) := \lim_{t\to 0}\frac{\sigma_t^*\omega - \omega}{t}(p) \tag{32.44}$$

Property 32.4.14. The Lie derivative has the following Leibniz-type property with respect to differential forms:

$$\mathcal{L}_X(\omega(Y)) = (\mathcal{L}_X\omega)(Y) + \omega(\mathcal{L}_XY), \tag{32.45}$$

where X, Y are two vector fields and ω is a one-form.

32.4.3 Interior product

Definition 32.4.15 (Interior product). Aside from the exterior derivative one can also define another operation on the algebra of differential forms:

$$\iota_X : (\iota_X\omega)(v_1, \ldots, v_{k-1}) \mapsto \omega(X, v_1, \ldots, v_{k-1}). \tag{32.46}$$

This antiderivation of degree -1 is called the **interior product** or **interior derivative**. It can be seen as a generalization of the contraction map 21.3.15.

Notation 32.4.16. In certain situations the above notation might become cumbersome. For this reason the notation $X \lrcorner\, \omega$ is frequently used.

Formula 32.4.17 (Cartan's magic formula[1]). 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:

$$\mathcal{L}_X\omega = \iota_X(d\omega) + d(\iota_X\omega). \tag{32.47}$$

32.4.4 Lie derivative of tensor fields

Formula 32.4.18 (Lie derivative of tensor fields). By comparing the definitions of the Lie derivatives of vector fields 32.3.18 and differential forms 32.4.13, 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 $\mathcal{T} \in \Gamma(T^{(k,l)}M)$:

$$\mathcal{L}_X\mathcal{T}(p) := \left.\frac{d}{dt}\sigma_t^*\mathcal{T}(\gamma_p(t))\right|_{t=0}. \tag{32.48}$$

Alternative Definition 32.4.19 (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. $\mathcal{L}_X$ 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 32.4.20 (Derivations). Every derivation D of the tensor algebra can be decomposed as

$$D = \mathcal{L}_X + S \tag{32.49}$$

for some vector field X and some endomorphism S.

[1] Sometimes called **Cartan's (infinitesimal) homotopy formula**.

32.4.5 Vector-valued differential forms

Definition 32.4.21 (Vector-valued form). Consider a vector space V and let $E \to M$ be a vector bundle with typical fibre V. A vector-valued differential form on M can be defined in two ways. A vector-valued k-form can be defined as a map $\omega : \Gamma(T^kM) \to V$ or, more generally, as a section of the following associated bundle:

$$\Omega^k(M; E) := \Gamma(E \otimes \Lambda^k T^*M). \tag{32.50}$$

The latter construction is also often called a **vector bundle-valued differential form**.

Construction 32.4.22 (Wedge product). Let $\omega \in \Omega^p(M; E_1)$ and $\nu \in \Omega^q(M; E_2)$. The wedge product of these differential forms is defined as follows:

$$\omega \wedge \nu(v_1, \ldots, v_{p+q}) := \frac{1}{p!q!} \sum_{\sigma \in S_{p+q}} \mathrm{sgn}(\sigma)\omega(v_{\sigma(1)}, \ldots, v_{\sigma(p)}) \otimes \nu(v_{\sigma(p+1)}, \ldots, v_{\sigma(q)}). \tag{32.51}$$

This is a direct generalization of the formula for the wedge product of ordinary differential forms, where the scalar product (product in the algebra $\mathbb{R}$) is replaced by the tensor product (product in 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 $E_1 \otimes E_2$.

Remark 32.4.23 (Differential vs. pushforward). At this point the reason why the pushforward 32.12 is also sometimes called the differential (as in Definition 32.1.1) can be given.

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

$$df := df^i \otimes \partial_i, \tag{32.52}$$

where $(f^1, \ldots, f^{\dim(N)})$ is a local expression for f and $\{\partial_1, \ldots, \partial_{\dim(N)}\}$ 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 32.4.24 (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 $de^i := 0$ one can again construct an exterior derivative through the Leibniz formula.

Remark 32.4.25. 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 32.4.26 (Lie algebra-valued form). A vector-valued differential form where the vector space V is equipped with a Lie algebra structure.

Formula 32.4.27 (Wedge product). Let $\omega \in \Omega^p(M; \mathfrak{g})$ and $\nu \in \Omega^q(M; \mathfrak{g})$ where $\mathfrak{g}$ is a Lie algebra. The wedge product of these differential forms is defined as follows:

$$[\omega \wedge \nu](v_1, \ldots, v_{p+q}) := \frac{1}{p!q!} \sum_{\sigma \in S_{p+q}} \mathrm{sgn}(\sigma)[\omega(v_{\sigma(1)}, \ldots, v_{\sigma(p)}), \nu(v_{\sigma(p+1)}, \ldots, v_{\sigma(q)})], \tag{32.53}$$

where $[\cdot, \cdot]$ denotes the Lie bracket on $\mathfrak{g}$.

Formula 32.4.28. Let $\{e_a\}_{a\leq\dim(\mathfrak{g})}$ be a basis for the Lie algebra $\mathfrak{g}$. One can write any Lie algebra-valued differential forms as $\phi = \phi^a \otimes e_a$ and $\psi = \psi^b \otimes eb$, where ϕ^a and ψ^b are ordinary differential forms. The above formula for the wedge product can now be rewritten more elegantly as

$$[\phi \wedge \psi] = (\phi^a \wedge \psi^b) \otimes [e_a, e_b], \tag{32.54}$$

where $\wedge$ is the wedge product on $\Omega^\bullet(M)$.

Corollary 32.4.29 (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:

$$[\phi \wedge \psi] = (-1)^{pq+1}[\psi \wedge \phi], \tag{32.55}$$

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:

$$(-1)^{pr}[\phi \wedge [\psi \wedge \theta]] + (-1)^{pq}[\psi \wedge [\theta \wedge \phi]] + (-1)^{qr}[\theta \wedge [\phi \wedge \psi]] = 0, \tag{32.56}$$

where $\theta \in \Omega^r(M;\mathfrak{g})$.

32.5 Frobenius's theorem

Definition 32.5.1 (Distribution). A section of the Grassmann k-plane bundle 32.2.22.

Definition 32.5.2 (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 $\forall p \in N : W(p) = T_pN$. W is said to be integrable if there exists such a submanifold N.

Property 32.5.3. If a distribution on M is integrable, M can be written as the (disjoint) union of maximal connected, integrable manifolds. These submanifolds are also called the **leaves** of the distribution (the decomposition in leaves defines a *foliation*).

Definition 32.5.4 (Frobenius's integrability condition). A distribution W on a manifold M is said to satisfy the Frobenius integrability condition on an open set $U \subseteq M$ if 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 32.5.5 (Frobenius's integrability theorem). *Let W be a distribution over a manifold M. W is integrable if and only if it satisfies the Frobenius integrability condition.*

This theorem also admits a formulation in terms of differential forms.

Property 32.5.6 (Differential formulation). Consider a rank-r distribution D. The annihilator $I(D)$ is the ideal containing all differential forms satisfying

$$\omega(X_1, \ldots, X_k) = 0 \tag{32.57}$$

whenever all $X_i \in D$. The Frobenius theorem says that the following conditions are equivalent:

- D is integrable.

- $I(D)$ is a differential ideal.

Property 32.5.7 (Pfaffian system). The fact that D is a distribution implies that locally there exists a set 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.

Given a Pfaffian system $\{\theta^\alpha\}_{\alpha \leq \dim(M)-r}$, a third equivalent condition for (**complete**) integrability is that

$$d\theta^1 \wedge \theta^2 \wedge \cdots = 0. \tag{32.58}$$

Property 32.5.8 (Integral manifold). An integral manifold of a rank-r Pfaffian system $\{\theta^\alpha\}_{\alpha \leq \dim(M)-r}$ is a smooth function $f : N \to M$ such that

$$f^*\theta^\alpha = 0 \tag{32.59}$$

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 $\{dy^\alpha\}_{\alpha \leq \dim(M)-r}$ generates the same differential ideal as the Pfaffian system. The integral submanifolds of the distribution are given by the systems

$$\begin{cases} y^1 = c^1, \\ y^2 = c^2, \\ \quad \vdots \end{cases} \tag{32.60}$$

for constants c^α.

Definition 32.5.9 (Characteristic system). Consider a system P of exterior differential equations $\{\omega^\alpha = 0\}$ where the degree of all forms is nonzero. As for a smooth distribution, one can locally find a set of one-forms that generate the same ideal. This is again called the **Pfaffian system**. The characteristic system of P is defined as the Pfaffian system of the differential closure of P.

If P also contained functions, i.e. degree-zero forms, the characteristic system is that of $\{\omega^\alpha, df^\beta\}$ extended by the functions f^β themselves.

Definition 32.5.10 (Integral invariant). Consider a Pfaffian system $P \equiv \{\theta^\alpha\}$. A differential form ω is called an invariant of P if

$$\mathcal{L}_X\omega = 0 \tag{32.61}$$

whenever $\theta(X) = 0$ for al $\theta \in P$. 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 $d\varphi$ is an (absolute integral) invariant of P, then φ is called a relative integral invariant.

32.6 Linear connections

32.6.1 Koszul connections

Definition 32.6.1 (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

$$\nabla(f\sigma) = f\nabla\sigma + df \otimes \sigma \tag{32.62}$$

for all $f \in C^\infty(M)$. When evaluated on a vector field X, ∇_X is often called a **covariant derivative**.

Property 32.6.2. Because $\nabla\sigma$ takes a vector field as input, which is a $C^\infty(M)$-linear operation, the connection satisfies the following linearity property:

$$\nabla_{fX+Y}\sigma = f\nabla_X\sigma + \nabla_Y\sigma. \tag{32.63}$$

Formula 32.6.3. Let E, E' be two vector bundles over the same manifold M. Koszul connections on E and E' induce a connection on the tensor product bundle $E \otimes E'$ as follows:

$$\nabla(X \otimes Y) := \nabla X \otimes Y + X \otimes \nabla Y \tag{32.64}$$

for $X \in \Gamma(E), Y \in \Gamma(E')$.

Example 32.6.4 (Affine connection). Let M be a manifold. An affine connection $\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ is a Koszul connection on the tangent bundle.

Property 32.6.5 (Local behaviour). Consider a vector $v \in T_pM$. If two vector fields $X, Y \in \Gamma(TM)$ coincide on some neighbourhood U of p, then $\nabla_v X = \nabla_v Y$ at p. Furthermore, given a curve $c : [0,1] \to M$ and two vector fields $X, Y \in \Gamma(TM)$ such that $X \circ c = Y \circ c$, one finds that $\nabla_{\dot c} X = \nabla_{\dot c} Y$. This implies that an affine connection only depends on the local behaviour of the given section.

Remark 32.6.6. The above property shows the major difference between the Lie derivative and the covariant derivative when acting on sections of the tangent bundle σ. 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 32.6.7 (Parallel tensor fields). 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 32.6.8. 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 34).

Property 32.6.9 (Affinity). Consider two affine connections $\nabla, \overline{\nabla}$ on a smooth manifold M. The operator $\nabla - \overline{\nabla}$ is an endomorphism of E, i.e. $\nabla - \overline{\nabla} \in \Omega^1(M; \mathrm{End}(E))$. It follows that the set of affine connections forms an affine space (hence the name).

Definition 32.6.10 (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

$$\nabla e_i = \Gamma^k{}_{ji} e_k \otimes f^j. \tag{32.65}$$

The quantities $\Gamma^k{}_{ji}$ are called the connection coefficients or **Christoffel symbols** of ∇. For a general vector field $\sigma = \sigma^i e_i$ one then obtains (if $\{e_i, f^i\}_{1\leq i\leq k}$ are coordinate-induced):

$$\begin{aligned}\nabla\sigma &= (\nabla\sigma^i) \otimes e_i + \sigma^i(\nabla e_i)\\ &= (\partial_j\sigma^k)e_k \otimes f^j + \sigma^i(\Gamma^k{}_{ji} e_k \otimes f^j)\\ &= (\partial_j\sigma^k + \Gamma^k{}_{ji}\sigma^i)e_k \otimes f^j.\end{aligned} \tag{32.66}$$

Definition 32.6.11 (Curvature). Let ∇ be a Koszul connection on a vector bundle $E \to M$. The associated curvature 2-form $F_\nabla \in \Omega^2(M; \mathrm{End}(E))$ is defined as follows:

$$F_\nabla(X, Y)\sigma := [\nabla_X, \nabla_Y]\sigma - \nabla_{[X,Y]}\sigma. \tag{32.67}$$

Definition 32.6.12 (Torsion). Let ∇ be an affine connection on smooth manifold M. The associated torsion 2-form $T \in \Omega^2(M;TM)$ is defined as follows:

$$T(X,Y) := \nabla_X Y - \nabla_Y X - [X,Y]. \tag{32.68}$$

Property 32.6.13 (Bianchi identities). Let ∇ be an affine connection on smooth manifold M. The associated curvature and torsion forms are related as follows:

$$R(X,Y)Z - T(T(X,Y),Z) + \text{cyclic} = \nabla_X T(Y,Z) \tag{32.69}$$
$$\nabla_X R(Y,Z) + R(T(X,Y),Z) + \text{cyclic} = 0, \tag{32.70}$$

where cyclic indicates that all terms obtained by cyclic permutations of the vector fields have to be included.

32.6.2 Induced connections

Formula 32.6.14 (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:

$$\begin{aligned}\nabla_Y T(\omega^1,\ldots,\omega^k,X_1,\ldots,X_l) := Y\Big(T(\omega^1,\ldots,\omega^k,X_1,\ldots,X_l)\Big)&\\ -\sum_{i=1}^k T(\omega^1,\ldots,\nabla_Y\omega^i,\ldots,\omega^k,X_1,\ldots,X_l)&\\ -\sum_{i=1}^l T(\omega^1,\ldots,\omega^l,X_1,\ldots,\nabla_Y X_i,\ldots,X_l),&\end{aligned} \tag{32.71}$$

where $Y,X_1,\ldots,X_l \in \mathfrak{X}(M)$ and $\omega^1,\ldots,\omega^k \in \Omega^1(M)$.

Corollary 32.6.15 (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:

$$(\nabla_X\nabla)_Y Z = \nabla_X(\nabla_Y Z) - \nabla_{\nabla_X Y} Z - \nabla_Y(\nabla_X Z). \tag{32.72}$$

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):

$$\nabla^2_{X,Y} S := \nabla_X(\nabla_Y S) - \nabla_{\nabla_X Y} S. \tag{32.73}$$

As an example of the second covariant derivative the definition of the Hessian on arbitrary smooth manifolds is given:

Definition 32.6.16 (Hessian). Consider a manifold with connection ∇. The Hessian of a function $f \in C^\infty(M)$ is defined as the iterated covariant derivative:

$$\mathrm{Hess}(f) := \nabla^2 f, \tag{32.74}$$

where one should note that by the above definition the first covariant derivative also acts on the second one, i.e

$$\nabla^2 f(X,Y) = \nabla_X(\nabla_Y f) - \nabla_{\nabla_X Y} f. \tag{32.75}$$

For a scalar function one knows that $\nabla f = df$ and for covector fields one knows that (in local coordinates)

$$\nabla_i \sigma_j = \partial_i \sigma_j - \Gamma^k_{ij}\sigma_k,$$

where Γ^k_{ij} are the connection coefficients. Combining these facts one obtains the following local expression for the Hessian of f:

$$\mathrm{Hess}(f) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} - \Gamma^k_{ij}\frac{\partial f}{\partial x_k}\right) dx^i \otimes dx^j. \tag{32.76}$$

Definition 32.6.17 (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 31.1.18 there exists a unique Koszul connection ∇' satisfying

$$\nabla'(f^*\chi) = f^*(\nabla\chi) \tag{32.77}$$

for any section χ of E.

Definition 32.6.18 (Invariant connection). Let G be a Lie group acting on a vector bundle $E \to M$. A Koszul connection ∇ on E is said to be invariant with respect to the G-action if it satisfies:

$$g^*\nabla = \nabla \tag{32.78}$$

for all $g \in G$.

32.7 Integration Theory

For the theory of measure spaces and Lebesgue integration, see Chapter 16.

32.7.1 Orientation and densities

One can define an orientation of manifolds by generalizing the situation for vector spaces 21.4.24:

Definition 32.7.1 (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 $\mathrm{Vol} \in \Omega^{\dim(M)}(M)$. 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. The existence of such a volume form turns a differentiable manifold into an **orientable manifold**.

Alternatively, an (orientable) manifold with volume form Vol is said to be **positively oriented** if it comes equipped with a smooth choice of bases $\{v_1, \ldots, v_n\}$ for T_pM such that

$$\mathrm{Vol}_p(v_1, \ldots, v_n) > 0. \tag{32.79}$$

Example 32.7.2. Let $M = \mathbb{R}^n$. The canonical Euclidean volume form is given by the determinant map

$$\det : (u_1, \ldots, u_n) \mapsto \det(u_1, \ldots, u_n), \tag{32.80}$$

where the u_n's are expressed in the canonical basis $(e_1, \ldots, e_n)$. 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 $\{u_1, \ldots, u_n\}$.

Property 32.7.3. 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 32.7.4 (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 32.7.5 (Orientable manifold). A manifold is orientable if its orientation bundle is trivial.

Remark 32.7.6. By definition of the orientation bundle, the transition functions are those that have a positive determinant. This gives the equivalence with Definition 32.7.1. In the next chapter on principal bundles yet another (equivalent) definition of orientability in terms of G-structures will be given (see Example 33.5.5).

Further below, integration theory will be generalized from orientable manifolds to non-orientable manifolds. To achieve this goal the notion of differential forms needs to be generalized. A good introduction for this is [122].

Definition 32.7.7 (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

$$\mathbf{1}_{\text{sgn}} : g \mapsto \det(\phi(g)). \tag{32.81}$$

The notation $\mathbf{1}_{\text{sgn}}$ refers to the fact that this is a generalization of the *alternating* (or *sign*) *representation* of the permutation groups S_n. Any Riemannian manifold admits a canonical pseudoscalar bundle Ψ associated to its (orthogonal) frame bundle.

Sections of a vector bundle with transition functions defined by $\mathbf{1}_{\text{sgn}}$ 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**.

Definition 32.7.8 (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

$$\rho : A \mapsto \det(A)^{-s}. \tag{32.82}$$

The number s 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 32.7.9 (Pseudovectors). The representation

$$\rho : A \mapsto \operatorname{sgn}\det(A)A \tag{32.83}$$

gives rise to a bundle similar to the tangent bundle, where the sign of the cocycles t_{ji} 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 Ψ (hence its name).

Remark 32.7.10 (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. These are the objects one can integrate over any manifold, even the non-orientable 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 combining Definition 32.7.8 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 32.7.11 (Orientability). A smooth manifold is orientable if and only if its canonical line bundle 32.2.21 is trivial. Furthermore, for orientable manifolds there exists an isomorphism $\Gamma(\det(T^*M)) \cong \Gamma(|\Omega|(M))$.

32.7.2 Orientation in homology

In this section a characterization of orientability in terms of the homology of a manifold is given. Smoothness is not required here. See Sections 8.2 and 8.3 for an introduction to homology.

Begin with the canonical example $\mathbb{R}^n$. Intuitively one would expect an orientation on 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 $H_n(S^n) \cong H_n(\mathbb{R}^n, \mathbb{R}^n\backslash\{*\})$. 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 $H_n(\mathbb{R}^n, \mathbb{R}^n\backslash\{x\})$.

For a given manifold M one then defines a global orientation (if it exists) as 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.

Property 32.7.12 (Orientability). If a closed connected manifold is ($\mathbb{Z}$-)orientable, there exists an isomorphism

$$H_n(M) \cong H_n(M, M\backslash\{p\}) \tag{32.84}$$

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**.

In the case where 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 8.5.1).

The above definition and property can be generalized to arbitrary unital rings R:

Definition 32.7.13 (R-orientability). A manifold is M-orientable if a consistent choice of local R-orientation exists or, equivalently, if $H_n(M;R) \cong R$.

Property 32.7.14 (Non-orientable 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 32.7.15 (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.

32.7.3 Integration of top-dimensional forms

Definition 32.7.16 (Measure zero). A subset $U \subset M$ 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 32.7.17 (Integrable form). A differential form for which its components with respect to any basis of $\Omega^k(M)$ are Lebesgue integrable on $\mathbb{R}^n$.

Formula 32.7.18 (Integration with compact support). Consider a top-dimensional form $\omega \in \Omega^{\dim(M)}$ on M with compact support on a coordinate patch $U \subset M$.

$$\int_M \omega = \int_U \omega := \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \omega_{12\ldots n}(x)dx^1dx^2\cdots dx^n. \tag{32.85}$$

This integral is well-defined because under an orientation-preserving change of coordinates 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 then 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:

$$\int_M \omega := \sum_{i\in I}\int_{U_i} \rho_i\omega. \tag{32.86}$$

Remark 32.7.19. 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 32.7.20 (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 32.7.21 (Invariance under pullbacks). Consider an orientation-preserving diffeomorphism $f : M \to N$.

$$\int_M f^*\omega = \int_N \omega \tag{32.87}$$

Notation 32.7.22. Because the integral of differential forms is linear in the integrand and additive over disjoint unions, it can be interpreted as a linear pairing. This motivates the following notation:

$$\langle M, \omega\rangle := \int_M \omega. \tag{32.88}$$

32.7.4 Stokes's theorem

Theorem 32.7.23 (Stokes's theorem). *Let M be an orientable manifold with boundary ∂M and let ω be a differential k-form on M.*

$$\int_{\partial M}\omega = \int_M d\omega. \tag{32.89}$$

Corollary 32.7.24. The Kelvin-Stokes theorem 21.2.5, the divergence theorem 21.2.6 and Green's identity 21.2.7 are immediate results of this (generalized) Stokes's theorem.

Definition 32.7.25 (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**: $d\omega = 0$.
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 32.7.26. Calibrated submanifolds minimize the volume within their homology class.

32.7.5 Distributions ♣

For more information on the theory of distributions on Euclidean space, see Chapter 17.

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 [9]) 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 (see Chapter 34). This data allows to turn the space of smooth sections of any tensor bundle over M into a Fréchet space 23.3.10 using a generalization of the seminorms (17.2), 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 $K\subset M$ 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 32.7.27 (Tensor distribution). The space of tensor distributions of order p is defined as 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 32.7.28 (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):

$$\langle\nabla T,\sigma\rangle := -\langle T, g\cdot\nabla\sigma\rangle, \tag{32.90}$$

where $g \cdot \nabla\sigma$ denotes the **internal contraction** (generalizing the divergence of a vector field) which, in local coordinates, is given by

$$(g \cdot \nabla\sigma)^{i_1 \dots i_p} = \nabla_j \sigma^{j i_1 \dots i_p}. \tag{32.91}$$

Definition 17.3.20 can easily be generalized to smooth manifolds:

Definition 32.7.29 (Wave front set). The wave front set of a distribution $\phi \in D'(M)$ is defined as follows:

$$\mathrm{WF}(\phi) := \{(x, v) \in T^*M_0 \mid v \in \Sigma_x(\phi)\}, \tag{32.92}$$

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).

?? COMPLETE? ??

32.8 Cohomology

32.8.1 de Rham complex

Definition 32.8.1 (Exact form). If $\omega \in \Omega^k(M)$ can be written as $\omega = d\chi$ for some $\chi \in \Omega^{k-1}(M)$, it is said to be exact.

Definition 32.8.2 (Closed form). Let $\omega \in \Omega^k(M)$. If $d\omega = 0$, it is said to be closed.

Definition 32.8.3 (de Rham complex). The sequence

$$0 \longrightarrow \Omega^0(M) \longrightarrow \Omega^1(M) \longrightarrow \cdots \longrightarrow \Omega^{\dim(M)}(M) \longrightarrow 0 \tag{32.93}$$

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 32.8.8 below for more information).

Definition 32.8.4 (de Rham cohomology). Following Definition 5.1.3 the k^{th} de Rham cohomology group on M is defined as the k^{th} (co)homology group of the de Rham complex:

$$H^k_{\mathrm{dR}}(M) := \frac{\ker(d_k)}{\mathrm{im}(d_{k-1})}. \tag{32.94}$$

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 32.8.5 (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 k-cycle with integral coefficients gives an integer.

Formula 32.8.6 (Cup product). Let $[\nu] \in H^k_{\mathrm{dR}}$ and $[\omega] \in H^l_{\mathrm{dR}}$. The cup product on de Rham cohomology is given by

$$[\nu] \smile [\omega] := [\nu \wedge \omega]. \tag{32.95}$$

The following theorem allows to write $H^\bullet(M)$ for the de Rham cohomology on M:

Theorem 32.8.7 (de Rham). *The de Rham cohomology over a smooth manifold is isomorphic to its singular cohomology (Section 8.3).*

Theorem 32.8.8 (Poincaré lemma[2]). *For every point $p \in M$ there exists a neighbourhood on which the de Rham cohomology is trivial:*

$$\forall p \in M : \exists U \subseteq M : H^k(U) = 0. \tag{32.96}$$

This implies that every closed form is locally exact, i.e. if $d\omega = 0$ at the point $p \in M$, there exist a neighbourhood $U \subseteq M$ of p and a differential form λ such that

$$\omega = d\lambda \tag{32.97}$$

at all points $p' \in U$. More generally, this lemma says that the following isomorphism exists for every smooth manifold M:

$$H^\bullet(M \times \mathbb{R}^n) \cong H^\bullet(M). \tag{32.98}$$

In fact, this can even be further generalized due to the homotopy axiom of de Rham cohomology:

$$H^\bullet(E) \cong H^\bullet(M) \tag{32.99}$$

for every vector bundle E over M.

Definition 32.8.9 (Relative cohomology). Consider the submanifold inclusion $\iota : S \hookrightarrow M$. The relative de Rham complex is defined as follows:

$$\Omega^n(M, S) := \Omega^n(M) \oplus \Omega^{n-1}(S), \tag{32.100}$$

where the coboundary operator d is defined by

$$d(\omega, \lambda) := (d\omega, \iota^*\omega - d\lambda). \tag{32.101}$$

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 in fact 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 is exists:

$$\cdots \longrightarrow H^k(f) \longrightarrow H^k(M) \longrightarrow H^k(N) \longrightarrow H^{k+1}(f) \longrightarrow \cdots. \tag{32.102}$$

Definition 32.8.10 (Twisted de Rham complex). Consider the usual (graded) de Rham ring $\Omega^\bullet(M)$ on a smooth manifold M. For every degree-3 class $\alpha \in H^3(M)$, one can define the twisted de Rham differential

$$d_\alpha := d + \alpha\wedge. \tag{32.103}$$

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 8.6.1:

[2] The original theorem states that on a contractible space 8.1.7 every closed form is exact.

Property 32.8.11 (Equivariant de Rham theorem ♣). Consider a smooth manifold M with a smooth G-action and let $W(\mathfrak{g})$ be the Weil algebra 30.6.6 of G. Construct the dgca $\Omega^\bullet(M) \otimes W(\mathfrak{g})$ with differential $d_{\mathrm{dR}} + d_W$. 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 $\iota_\xi, \mathcal{L}_\xi$ for all $\xi \in \mathfrak{g}$. This subcomplex, with the induced differential $d_{\mathrm{dR}} + d_W$, 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 8.6.1.

Property 32.8.12 (Cohomological models). The intersection of the basic subcomplex with $\Omega^\bullet(M)$ can be identified with the complex of tensorial (basic) differential forms (see Definition 33.3.14), 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 (the Weil algebra itself gives a model for the total spacee EG).

32.8.2 Integration

At this point a little side note can be given about why the de Rham cohomology groups 32.8.4 really constitute a cohomology theory. Some concepts from homology are needed that can be found in Section 8.2.

Let M be a compact manifold and let $\{\lambda_i : \Delta^k \to M\}$ be the set of singular k-simplexes on M. Suppose that one wants to integrate a form over a singular k-chain $C = \sum_{i=0}^k a_i\lambda_i$ 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:

$$\langle\cdot, C\rangle : \Omega^k(M) \to \mathbb{R} : \omega \mapsto \int_C \omega = \sum_{i=0}^k a_i \int_{\Delta_k} \lambda_i^*\omega, \tag{32.104}$$

where λ_i^* pulls ω back to Δ^k, which is a subset of $\mathbb{R}^k$ as required. Stokes's theorem 32.7.23 then says that

$$\int_C d\omega = \int_{\partial C} \omega. \tag{32.105}$$

Using the pairing $\langle\cdot,\cdot\rangle$ this can be rewritten more explicitly as

$$\langle d\omega, C\rangle = \langle\omega, \partial C\rangle. \tag{32.106}$$

The operators d and ∂ can thus be interpreted as formal adjoints. After confirming (again using Stokes' theorem) that 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 $\langle\omega, C\rangle$, 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 32.8.4.

32.8.3 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 states otherwise).

The following definition characterizes cohomology with compact support directly through its relation to compact sets:

Definition 32.8.13 (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:

$$H_c^\bullet(M) := \varinjlim_{\text{compact } K} H^\bullet(M, M\backslash K). \tag{32.107}$$

Property 32.8.14 (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 $H^\bullet(X, X\backslash K \cup V) \cong H^\bullet(U, U\backslash K)$ 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

$$H_c^\bullet(M) \cong H^\bullet(\widehat{M}, *) \cong \widetilde{H}^\bullet(\widehat{M}), \tag{32.108}$$

where the fact that both $*$ and $\widehat{M}$ are compact (this is true for Hausdorff spaces) is used.

Theorem 32.8.15 (Poincaré duality). *Let M be an m-dimensional manifold. The pairing $\int : H^k(M) \otimes H_c^{m-k}(M) \rightarrow \mathbb{R}$ induces an isomorphism on cohomology:*

$$H^k(M) \cong \left(H_c^{m-k}(M)\right)^*. \tag{32.109}$$

If M is of finite type, the converse also holds:

$$H_c^k(M) \cong \left(H^{m-k}(M)\right)^*. \tag{32.110}$$

Corollary 32.8.16 (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:

$$H_c^\bullet(E) \cong H_c^{\bullet-n}(M). \tag{32.111}$$

Definition 32.8.17 (Poincaré dual). Let M be an m-dimensional manifold and let $i : S \rightarrow M$ be a closed k-dimensional submanifold. The Poincaré dual of S in M is the unique cohomology class $[\eta_S] \in H^{m-k}(M)$ such that

$$\int_S i^*\omega = \int_M \omega \wedge \eta_S \tag{32.112}$$

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

$$\int_S i^*\omega = \int_M \omega \wedge \eta'_S \tag{32.113}$$

  for all $\omega \in H^k(M)$.

Remark 32.8.18. 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 32.8.19 (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 32.8.20 (Transversal intersections). The Poincaré dual of a transversal intersection is equal to the wedge product of the individual Poincaré duals:

$$\eta_{S\pitchfork T} = \eta_S \wedge \eta_T. \tag{32.114}$$

Definition 32.8.21 (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 $\Omega^\bullet_{cv}(E)$ if $\mathrm{supp}(\omega) \cap \pi^{-1}(K)$ is compact for every compact subset $K \subset M$. The cohomology of this complex is called the **de Rham cohomology with compact support in the vertical direction**.

Remark. The above definition implies that $\omega \in \Omega^\bullet_{cv}(E)$ is compactly supported on each fibre $\pi^{-1}(p), p \in M$. This observation explains the name of the cohomology theory.

Definition 32.8.22 (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 $f(x,u)\pi^*\phi \wedge (du_{i_1} \wedge \cdots \wedge du_{i_k})$, where ϕ is a form on the base manifold M, f has compact support and $k < n$.
- **Type 2**: those locally of the form $f(x,u)\pi^*\phi \wedge (du_1 \wedge \cdots \wedge du_n)$ where ϕ is a form on the base manifold M and f has compact support.

The fibre integration map $\pi_* : \Omega^\bullet_{cv}(E) \to \Omega^{\bullet-n}(M)$ is defined as follows. If ω is of type 1, then $\pi_*\omega := 0$. If ω is of type 2, then

$$\pi_*\omega := \left(\int\!\!\cdots\!\!\int f(x,u)du_1\cdots du_n\right)\phi. \tag{32.115}$$

Formula 32.8.23 (Projection formula). Consider a rank-n vector bundle $\pi : E \to M$. For every pair of forms $\phi \in \Omega^\bullet(M)$ and $\omega \in \Omega^\bullet_{cv}(E)$, the following formula holds:

$$\pi_*(\pi^*\phi \wedge \omega) = \phi \wedge \pi_*\omega. \tag{32.116}$$

Furthermore, if $\phi \in \Omega^k_c(M)$ and $\omega \in \Omega^{m+n-k}_{cv}(E)$,

$$\int_E \pi^*\phi \wedge \omega = \int_M \phi \wedge \pi_*\omega. \tag{32.117}$$

Theorem 32.8.24 (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:*

$$\pi_* : H^\bullet_{cv}(E) \cong H^{\bullet-n}(M) : \mathcal{T}. \tag{32.118}$$

Corollary 32.8.25 (Poincaré lemma for vertically compact cohomology).

$$H^\bullet_{cv}(M \times \mathbb{R}^n) \cong H^{\bullet-n}(M). \tag{32.119}$$

Formula 32.8.26 (Thom isomorphism). Denote the **Thom class** of M by $\Phi := \mathcal{T}(1) \in H^0(M)$. Because $\mathcal{T}$ and π_* are mutual inverses and, hence, $\pi_*\Phi = 1$, the projection formula implies that

$$\mathcal{T}(\omega) = \pi^*\omega \wedge \Phi. \tag{32.120}$$

Property 32.8.27 (Orientation class). The Thom class Φ restricts to a generator of the cohomology of the typical fibre:

$$H^n_c(V) \cong \widetilde{H}^n(S^n) \cong H^n(S^n).$$

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 32.7.12.

Property 32.8.28 (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.

Definition 32.8.29 (Thom space). Let $E \to M$ be a vector bundle. For every fibre in E one can construct its one-point compactification 7.5.27 and by gluing these together the **sphere bundle** $\mathrm{Sph}(E)$ is obtained. The quotient space $\mathrm{Sph}(E)/B$, where all the adjoined points are identified, is called the Thom space $\mathrm{Th}(E)$.

By equipping E with a metric (see Chapter 34), 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 $\mathrm{Th}(E)$ can be shown to be isomorphic to the quotient space $D(E)/S(E)$.

Property 32.8.30. If the base manifold M is compact, the Thom space is obtained as the one-point compactification of the total space E.

Property 32.8.31 (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 32.2.8. Homotopy invariance implies that

$$H^\bullet(D(E), S(E)) \cong H^\bullet(E, E_0). \tag{32.121}$$

Then, using Result 32.1.15 together with the dual of Property 8.2.33, one can show that the reduced cohomology of the Thom space $\mathrm{Th}(E)$ is isomorphic to the relative cohomology of the pair (E, E_0):

$$\widetilde{H}^\bullet(\mathrm{Th}(E)) \cong H^\bullet(E, E_0). \tag{32.122}$$

To relate this to vertically compact cohomology, Property 32.8.14 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:

$$\begin{tikzcd} & \widetilde{H}^\bullet(\mathrm{Th}(E)) \arrow[dl, "\cong"'] \arrow[dr, "\cong"] & \\ H^\bullet(E, E_0) \arrow[rr, "\cong"'] & & H^\bullet_{cv}(E) \end{tikzcd} \tag{32.123}$$

Definition 32.8.32 (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:

$$(\Sigma^\infty \mathrm{Th}(E))_n \cong \mathrm{Th}(\mathbb{R}^n \oplus E), \tag{32.124}$$

where $\mathrm{Th}(\mathbb{R}^n \oplus E) \cong \Sigma^n \mathrm{Th}(E)$ was used.

Now, consider the sequence $(\xi_n)_{n\in\mathbb{N}}$ of *universal vector bundles*. For every ξ_n, define the n^{th} component space as follows:

$$MO_n := \mathrm{Th}(\xi_n). \tag{32.125}$$

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 n^{th} structure map of "the" Thom spectrum MO.[3]

Definition 32.8.33 (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 32.8.34. 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 32.8.35 (Normalization). If the vector bundle admits a nowhere-vanishing section, its Euler class vanishes.

32.8.4 Čech-de Rham complex

Theorem 32.8.36 (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:*

$$0 \longrightarrow H^\bullet(M) \overset{\iota_U \oplus \iota_V}{\longrightarrow} H^\bullet(U) \oplus H^\bullet(V) \overset{\pi_2 - \pi_1}{\longrightarrow} H^\bullet(U \cap V) \longrightarrow 0. \tag{32.126}$$

Definition 32.8.37 (Čech-de Rham complex). The Čech complex 9.3.10 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 for the Čech-de Rham complex:

Property 32.8.38 (Mayer-Vietoris sequence). The horizontal complex

$$0 \longrightarrow \Omega^\bullet(M) \longrightarrow \prod_{i_0} \Omega^\bullet(U_{i_0}) \longrightarrow \prod_{i_0, i_1} \Omega^\bullet(U_{i_0 i_1}) \longrightarrow \cdots \tag{32.127}$$

is acyclic, i.e. the δ-cohomology of the Čech-de Rham complex vanishes.

[3] Note that the Thom spectrum as defined here is not an Ω-spectrum 8.5.6, it is merely a sequential spectrum (prespectrum).

An important corollary is that one can compute the (de Rham) cohomology of M using the above double complex:

Property 32.8.39. 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 $U_{i_0\ldots i_p}$. 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 32.8.40 (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:*

$$H^\bullet(M) \cong \check{H}^\bullet(\mathcal{U};\flat\mathbb{R}). \tag{32.128}$$

By noting that good covers are *cofinal* in the set of open covers, one can pass to the full Čech cohomology:

$$H^\bullet(M) \cong \check{H}^\bullet(M;\flat\mathbb{R}). \tag{32.129}$$

Corollary 32.8.41. All compact manifolds admit a finite good cover and hence have finite-dimensional de Rham cohomology.

Property 32.8.42 (Exponential sequence). Consider the following exact sequence of topological groups:

$$0 \longrightarrow \mathbb{Z} \xrightarrow{2\pi} \mathbb{C} \xrightarrow{\exp} \mathrm{U}(1) \longrightarrow 0. \tag{32.130}$$

Let $(M,\mathcal{O}_M)$ be a complex manifold (or a smooth manifold and restrict the above sequence to the real numbers). The exact sequence induces an exact sequence of structure sheaves:

$$0 \longrightarrow \flat\mathbb{Z} \longrightarrow \mathcal{O}_M \longrightarrow \mathcal{O}_M^\times \longrightarrow 0. \tag{32.131}$$

This in turn induces a long exact sequence in cohomology and by Example 9.3.6 (if M is paracompact) the connecting homomorphism leads to an isomorphism

$$H^{\bullet+1}(M;\mathbb{Z}) \cong \check{H}^\bullet(M;\mathrm{U}(1)). \tag{32.132}$$

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.

32.8.5 Non-orientable manifolds

This section gives a differential-geometric incarnation of Section 8.4 on local coefficients.

Definition 32.8.43 (Twisted cohomology). Let $E \to M$ be a flat vector bundle over M. By Construction 32.4.24 the algebra $\Omega^\bullet(M)\otimes E$ can be given the structure of a differential graded algebra 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 32.8.44. According to the remark following Construction 32.4.24 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 $a\in\mathrm{GL}(n,\mathbb{R})$.

If one takes $E = o(M)$ to be the orientation line bundle over M, the (honest) densities of Remark 32.7.10 are obtained. The cohomology of this complex is simply called the **twisted de Rham cohomology**.

Property 32.8.45 (Isomorphism). The twisted cohomologies defined by two trivializations induced from atlases on M are isomorphic.[4]

Property 32.8.46 (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 32.8.15 can be generalized almost verbatim to the twisted case:

Theorem 32.8.47 (Poincaré duality). *Integration of densities induces the following isomorphism:*

$$H^k(M) \cong \Big(H_c^{m-k}(M; o(M))\Big)^*. \tag{32.133}$$

If M is of finite type, the converse also holds:

$$H_c^k(M) \cong \Big(H^{m-k}(M; o(M))\Big)^*. \tag{32.134}$$

The Thom isomorphism also holds for non-orientable bundles:

Theorem 32.8.48 (Thom isomorphism). *Let $E \to M$ be a rank-n vector bundle. Fibre integration gives the following isomorphism:*

$$H_{cv}^{\bullet+n}(E) \cong H^\bullet(M; o(E)). \tag{32.135}$$

32.8.6 Generalized cohomology ♣

In this section some statements from singular/de Rham cohomology on vector bundles are generalized to statements about generalized (Eilenberg-Steenrod) cohomology theories (Section 8.5). In the remainder of this section $E^\bullet$ will denote a multiplicative generalized cohomology theory.[5]

Definition 32.8.49 (Orientation). Let $\pi : E \to M$ be a rank-n vector bundle. An E-orientation or E-**Thom class** is a cohomology class $u \in \widetilde{E}^n(\mathrm{Th}(V))$ that restricts to a generator on every fibre of V.

32.9 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 19.

Definition 32.9.1 (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

$$\sigma_D : T^*M \to \mathrm{Hom}(\pi^*E, \pi^*F), \tag{32.136}$$

[4] 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.

[5] "Multiplicative" indicates that there exists a cup product such that every group $E^\bullet(M)$ becomes a graded ring.

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 hyperbolic if the the above morphism is invertible for all cotangent vectors.

Definition 32.9.2 (Normally hyperbolic operators). Consider a (pseudo)Riemannian vector bundle E (see Chapter 34). A linear differential operator on E is said to be normally hyperbolic if its principal symbol is proportional to the given metric.

32.9.1 Elliptic complexes

Definition 32.9.3 (Elliptic complex). Consider a collection of vector bundles $\{\pi_n : E_n \to M\}_{n\in\mathbb{N}}$ together with a sequence of differential operators $\big(D_n : C^\infty(E_n) \to C^\infty(E_{n+1})\big)_{n\in\mathbb{N}}$. 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 $\big(\sigma_p(\xi)(D_n)\big)_{n\in\mathbb{N}}$ is exact for all $x \in M$ and $\xi \neq 0$.

?? COMPLETE ??

Chapter 33

Principal Bundles

The main reference for this chapter is [60]. The theory of principal bundles uses the language of (Lie) group theory quite heavily. For all things related to group theory the reader is referred to Sections 3.2 and 3.3. For more information on Lie groups and their associated Lie algebras the reader is referred to Chapter 30.

33.1 Principal bundles

Definition 33.1.1 (Principal bundle). A fibre bundle $\pi : P \to M$ equipped with a right action $\rho : P \times G \to P$ 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 M 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 33.1.2 (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 3.3.15. However, it is possible to locally (i.e. in a neighbourhood of a point $p \in M$) endow the fibres with a group structure by choosing an element of every fibre to be the identity element.

Property 33.1.3. A corollary of the definition is that the bundle $\pi : P \to M$ 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 33.1.4 (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 33.1.5 (Dimension). The dimension of P is given by

$$\dim(P) = \dim(B) + \dim(G). \tag{33.1}$$

Property 33.1.6. Every local trivialization φ_i is G-equivariant:

$$\varphi_i(z \cdot g) = \varphi_i(z) \cdot g. \tag{33.2}$$

Definition 33.1.7 (Principal bundle map). A bundle map $F : P_1 \rightarrow P_2$ 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 31.1.10.
2. f_P is G-equivariant.

The map f_P is said to **cover** f_B.

The following property proves that the equivariance condition on principal bundle maps is in fact a very strong condition:

Property 33.1.8. Every principal bundle map covering the identity is an isomorphism.

Definition 33.1.9 (Vertical automorphism). Consider a principal G-bundle $\pi : P \rightarrow B$. 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.

33.1.1 Associated bundles

Construction 33.1.10 (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 31.1.14, where the left action of G is given by left multiplication in G.

Property 33.1.11. 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.

Example 33.1.12 (Frame bundle). Let V be an n-dimensional vector space and denote the set of frames 20.1.6 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 $\mathrm{GL}(V) \cong \mathrm{GL}(\mathbb{R}^n)$.

Given an n-dimensional vector bundle E, one can construct an associated principal bundle by replacing every fibre $\pi^{-1}(b)$ by $F(\pi^{-1}(b)) \cong \mathrm{GL}(\mathbb{R}^n)$. The right action on this bundle by $g \in \mathrm{GL}(\mathbb{R}^n)$ is given by the basis transformation $\widetilde{e}_j = g^i_j e_i$. This bundle is denoted by FE or FM in the case of the tangent bundle $E = TM$.

Construction 33.1.13 (Associated bundle to a principal bundle). Consider a principal G-bundle $\pi : P \rightarrow B$ 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]This should not be confused with the structure group G, which is also sometimes called the gauge group in physics.

1. Define an equivalence relation $\sim_G$ on the product space $P \times F$ by
$$(p, f) \sim_G (p', f') \iff \exists g \in G : (p', f') = (p \cdot g, g^{-1} \rhd f). \tag{33.3}$$
2. Define the total space of the associated bundle as the following quotient space:
$$P_F := (P \times F)/\sim_G . \tag{33.4}$$
3. Define the projection $\pi_F : P_F \to B$ as follows:
$$\pi_F : [p, f] \mapsto \pi(p), \tag{33.5}$$
where $[p, f]$ is the equivalence class of $(p, f) \in P \times F$ in the quotient space P_F.

Example 33.1.14 (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 $\rhd$ of a matrix group given by
$$\rhd : G \times \mathbb{R}^n \to \mathbb{R}^n : (g \rhd f)^i \mapsto g^i{}_j f^j. \tag{33.6}$$
The tangent bundle is isomorphic to the associated bundle $FM \times_\rhd \mathbb{R}^n$, where the bundle map is defined as $[e, v] \mapsto v^i e_i \in TM$.

Example 33.1.15 (Adjoint bundle). Consider a principal G-bundle P. G acts on itself by conjugation, i.e. the adjoint action:
$$\mathrm{Ad} : G \times G \to G : (g, h) \mapsto g^{-1} h g. \tag{33.7}$$
This action induces an associated bundle $\mathrm{Ad}(P) := P \times_G G$, suitable named the adjoint bundle of P.

Property 33.1.16 (Vertical automorphisms). There exists an isomorphism between the vertical automorphism group $\mathrm{Aut}_V(P)$ and the group of sections of the adjoint bundle $\mathrm{Ad}(P)$.

Construction 33.1.17 (Associated bundle map). Given a principal bundle map (f_P, f_B) 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 $\widetilde{f}_P : P \times_G F \to P \times_{G'} F$ is given by
$$\widetilde{f}_P([p, f]) := [f_P(p), f]. \tag{33.8}$$
- The base space map is simply given by f_B itself:
$$\widetilde{f}_B(b) = f_B(b). \tag{33.9}$$

33.1.2 Sections

Although every vector bundle has at least one global section, namely the zero section, a general principal bundle does not necessarily have a global section. This is made clear by the following property:

Property 33.1.18 (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 $\mathrm{Triv}(P)$.

Corollary 33.1.19. Every local section $\sigma : U \to P$ induces a local trivialization φ by

$$\varphi^{-1} : (m, g) \mapsto \sigma(m) \cdot g. \tag{33.10}$$

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 32.2.6 can now be reformulated as follows:

Property 33.1.20 (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 33.1.21 (Higgs fields). Let (P, B, π, G) be a principal 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

$$\sigma_\phi : B \to P_F : b \mapsto [p, \phi(p)], \tag{33.11}$$

where p is any point in $\pi^{-1}(\{b\})$. This is well-defined due to Equation (33.3). In the other direction one finds

$$\phi_\sigma : P \to F : p \mapsto j_p^{-1} \circ \sigma(\pi(p)), \tag{33.12}$$

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.

33.2 Universal bundle

Definition 33.2.1 (Universal bundle). Consider a topological group G. A universal bundle of G is any principal bundle of the form

$$G \hookrightarrow EG \to BG$$

where EG is weakly contractible. The space BG is called the **classifying space** of G.

Property 33.2.2. A principal G-bundle $EG \to BG$ is universal if and only if EG is weakly contractible.

Definition 33.2.3 (n-universal bundle). A principal bundle with an $(n-1)$-connected total space.

Property 33.2.4 (Delooping). For every topological group one can prove that the loop space of BG is (weakly) homotopy equivalent to G, i.e. $\Omega BG \cong G$. As such it also deserves the name of delooping.

Property 33.2.5 (Groups). Let G be a group (regarded as a discrete toological space). Because the fundamental group of a topological group is Abelian by Property 8.1.16, 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 33.2.6 (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 33.2.7 (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 $[X, B\mathrm{GL}(k)]$. Because $\mathrm{O}(k)$ is the maximal compact subgroup of $\mathrm{GL}(k)$, one also obtains the result that any real vector bundle over a paracompact space admits a *Riemannian structure* (see Chapter 34).

Property 32.2.4 now follows from Eckmann-Hilton duality 8.1.41 together with the above delooping property.

Remark 33.2.8. 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 $[X, BG]$ only classifies numerable principal bundles 31.1.8, but now over arbitrary base spaces X.

An explicit construction of the numerable universal bundle for any topological group G was given by *Milnor*:

Construction 33.2.9 (Milnor ♣). First, consider the infinite join E_∞ equipped with the strong topology. This space is constructed as the direct limit of finite joins 7.3.10:

$$E_n = \underbrace{G \circ \cdots \circ G}_{n \text{ times}},$$

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 33.2.10 (Category theory ♣). Let G be a topological group and consider the delooping (groupoid) $\mathbf{B}G$ from Definition 4.10.2. This groupoid can also be obtained as the *action groupoid* 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 realisation gives us a universal bundle map.

33.3 Connections

33.3.1 Vertical vectors

Because smooth fibre bundles are also smooth manifolds, one can define the traditional notions 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 33.3.1 (Vertical vector). Let $\pi : E \to M$ be a smooth fibre bundle. The subbundle $\mathrm{Vert}(TE) := \ker(\pi_*)$ 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 33.3.2. Consider a smooth principal G-bundle (P, M, π, G). First, construct a map ι_p for every element $p \in P$:

$$\iota_p : G \to P : g \mapsto p \cdot g. \tag{33.13}$$

Then, define a tangent vector $v \in T_pP$ to be vertical if it lies in the image of $\iota_{p,*}$, i.e. $\mathrm{Vert}(T_pP) := \mathrm{im}(\iota_{p,*})$. This definition is equivalent to the previous one because of the short exact sequence

$$0 \longrightarrow \mathfrak{g} \xrightarrow{\iota_{p,*}} T_pP \xrightarrow{\pi_*} T_pM \longrightarrow 0. \tag{33.14}$$

Property 33.3.3 (Dimension of vertical bundle). 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 $\iota_{p,*} : \mathfrak{g} \to \mathrm{Vert}(T_pP)$ is an isomorphism of vector spaces. In particular, it implies that

$$\dim(\mathrm{Vert}(T_pP)) = \dim(\mathfrak{g}) = \dim(G). \tag{33.15}$$

Definition 33.3.4 (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

$$A^\#(p) := \iota_{p,*}(A) \in \mathrm{Vert}(T_pP), \tag{33.16}$$

is called the fundamental vector field associated to A. The action of the vector field $A^\#$ is given by

$$A^\#_p(f) = \left.\frac{d}{dt} f(p \cdot \exp(tA))\right|_{t=0}, \tag{33.17}$$

where $f \in C^\infty(P)$.

Property 33.3.5. The map $(\cdot)^\# : \mathfrak{g} \to \Gamma(TP)$ is a Lie algebra morphism:

$$[A, B]^\# = [A^\#, B^\#], \tag{33.18}$$

where the Lie bracket on the left is the one in $\mathfrak{g}$ and the Lie bracket on the right is the one in $\mathfrak{X}(M)$ given by (32.27).

Property 33.3.6. The vertical bundle satisfies the following equivariance condition:

$$R_{g,*}(\mathrm{Vert}(T_pP)) = \mathrm{Vert}(T_{p\cdot g}P). \tag{33.19}$$

By differentiating the equality

$$R_g \circ \iota_p = \iota_{p\cdot g} \circ \mathrm{ad}_{g^{-1}}$$

and using Example 30.3.2 and Definition 33.3.4, one can obtain the following algebraic reformulation:

$$R_{g,*}\Big(A^\#(p)\Big) = \big(\mathrm{Ad}_{g^{-1}}A\big)^\#(p \cdot g). \tag{33.20}$$

33.3.2 Ehresmann connections

The exact sequence (33.14) does not split canonically. However, one can make a choice of splitting:

Definition 33.3.7 (Ehresmann connection). Consider a smooth fibre bundle E. An (Ehresmann) connection on E is the selection of a subspace $\mathrm{Hor}(T_eE) \leq T_eE$ for every $e \in E$ such that:

1. The horizontal and vertical bundles are complementary: $\mathrm{Vert}(T_eE) \oplus \mathrm{Hor}(T_eE) = T_eE$.
2. The choice of subspace depends smoothly on $e \in E$ in the sense of distributions 32.5.1.

The vectors in $\mathrm{Hor}(T_eE)$ are said to be **horizontal** (with respect to the chosen connection).

Definition 33.3.8 (Horizontal bundle). The horizontal (sub)bundle $\mathrm{Hor}(TE)$ is defined as $\bigsqcup_{e\in E} \mathrm{Hor}(T_eE)$ with the bundle structure induced from TE.

Definition 33.3.9 (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:

$$R_{g,*}(\mathrm{Hor}(T_pP)) = \mathrm{Hor}(T_{p\cdot g}P). \tag{33.21}$$

Remark 33.3.10. Note that this condition was automatically satisfied for vertical bundles as in Equation (33.19).

Property 33.3.11 (Dimension). Properties 33.1.5 and 33.3.3, together with the direct sum decomposition of TP, imply the following relation for all $p \in P$:

$$\dim(\mathrm{Hor}(T_pP)) = \dim(M). \tag{33.22}$$

All dimensional relations between the data of a principal bundle (P, M, π, G) are now summarized:

$$\begin{aligned} \dim(P) &= \dim(M) + \dim(G) \\ \dim(M) &= \dim(\mathrm{Hor}(T_pP)) \\ \dim(G) &= \dim(\mathrm{Vert}(T_pP)) \end{aligned} \tag{33.23}$$

for all $p \in P$.

Definition 33.3.12 (Dual connection). First, define the dual of the horizontal bundle:

$$\mathrm{Hor}(T_p^*P) := \left\{h \in T_p^*P \mid h(v) = 0, v \in \mathrm{Vert}(T_pP)\right\}. \tag{33.24}$$

It is the space of one-forms that vanish on the vertical subspace. A dual connection can then be defined as the selection of a vertical covector bundle $\mathrm{Vert}(T_p^*P)$ satisfying the conditions of Definitions 33.3.7 and 33.3.9, where Vert and Hor should now be interchanged. Note that here the horizontal bundle is canonically defined.

Definition 33.3.13 (Horizontal and vertical forms). Let $\theta \in \Omega^k(P)$ be a differential k-form.

- θ is said to be horizontal if
$$\theta(v_1, \ldots, v_k) = 0 \tag{33.25}$$
whenever at least one of the v_i is in $\mathrm{Vert}(T_pP)$.
- θ is said to be vertical if
$$\theta(v_1, \ldots, v_k) = 0 \tag{33.26}$$
whenever at least one of the v_i is in $\mathrm{Hor}(T_pP)$.

For functions $f \in \Omega^0(P)$ it is vacuously true that they are both vertical and horizontal.

Definition 33.3.14 (Tensorial form). Consider a differential form θ on a principal G-bundle P with values in a vector space V equipped with a representation $\rho : G \to V$. This form is said to be tensorial or **basic of type** (V, ρ) if it is horizontal and if it satisfies the equivariance condition

$$R_g^*\theta = \rho(g^{-1})\theta. \tag{33.27}$$

An equivalent construction exists. 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. The isomorphism is given fibrewise by

$$\phi \mapsto \overline{\phi} := f^{-1}(\pi^*\phi), \tag{33.28}$$

where $f : V \to E_{\pi(u)} \cong (\pi^*E)_u : v \mapsto [u, v]$.

33.3.3 Connection forms

Definition 33.3.15 (Connection one-form). Let P be a principal bundle G-bundle. A connection one-form, associated to a given principal connection, is a $\mathfrak{g}$-valued one-form $\omega \in \Omega^1(P; \mathfrak{g})$ that satisfies the following two conditions:

1. **Cancellation of fundamental vector fields**:

$$\omega(A^\#) = A \tag{33.29}$$

   for all $A \in \mathfrak{g}$, and

2. **Equivariance**:

$$\omega \circ R_{g,*} = \mathrm{Ad}_{g^{-1}} \circ \omega \tag{33.30}$$

   for all $g \in G$.

The horizontal subspaces are recovered as the kernel of the connection one-form: $\mathrm{Hor}(T_pP) = \ker \omega|_p$.

Property 33.3.16 (Connection form from principal connection). Consider a principal G-bundle P. Given a principal connection on P, the associated connection one-form is given by the following map:

$$\omega := (\iota_{p,*})^{-1} \circ \mathrm{pr}_{\mathrm{Vert}}, \tag{33.31}$$

where $\mathrm{pr}_{\mathrm{Vert}}$ is the canonical projection $TP \to \mathrm{Vert}(TP)$.

Property 33.3.17 (Pullback connection). Consider two principal G-bundles P_1, P_2. Let ω be a connection one-form on P_1 and let $F : P_1 \to P_2$ be a bundle map. The pullback $F^*\omega$ defines a principal connection on P_2 called the pullback connection.

33.3.4 Maurer-Cartan form

Definition 33.3.18 (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

$$\Omega := L_{g^{-1},*}. \tag{33.32}$$

Construction 33.3.19. 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 33.3.11 it follows that the horizontal spaces are null-spaces (this trivially defines a smooth distribution and a connection in the sense of Ehresmann 33.3.7) and that the vertical spaces are equal to the tangent spaces, i.e. $\mathrm{Vert}(T_gG) = T_gG$, where the identification $P \cong G$ (as manifolds) is used.

The simplest way to define a connection form ω on this bundle would be the trivial projection $\mathbb{1}_{TP} : TP \to TP = \mathrm{Vert}(TP)$. However, the image of this map would be T_gG and not $\mathfrak{g}$ as required. This can be solved by using the Maurer-Cartan form:

$$\omega(v) := \Omega(v). \tag{33.33}$$

Property 33.3.20. The Maurer-Cartan form is the unique principal connection on the bundle $G \hookrightarrow G \to \{*\}$.

Definition 33.3.21 (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:

$$\omega_f := f^*\Omega. \tag{33.34}$$

The function f is called an **integral** or **primitive** of ω_f.

Property 33.3.22. Let M be a connected manifold. If two functions $f, f' : M \to G$ have the same Darboux derivative, there exists an element $g \in G$ such that $f(p) = g \cdot f'(p)$ for all $p \in M$.

Theorem 33.3.23 (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*

$$d\omega + \frac{1}{2}[\omega \wedge \omega] = 0, \tag{33.35}$$

then (locally) there exists a smooth function $f : M \to G$ such that $\omega = f^*\Omega$.

33.3.5 Local representations

Definition 33.3.24 (Yang-Mills field). Consider a principal G-bundle $P \to M$ 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:

$$\omega^U := \sigma^*\omega. \tag{33.36}$$

Definition 33.3.25 (Local representation). Consider a principal bundle $P \to M$ 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 33.3.26. Consider a principal connection ω on a principal G-bundle $P \to M$. Because of Property 33.1.19 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:

$$\sigma^*\omega|_{(m,g)}(v, A) = \mathrm{Ad}_{g^{-1}}\big(\omega^U_m(v)\big) + \Omega_g(A), \tag{33.37}$$

where $v \in T_mU$, $A \in \mathfrak{g}$ and Ω is the Maurer-Cartan form on G.

Formula 33.3.27 (Compatibility condition). Consider a principal bundle (P, M, π, G) and two open subsets U, V of 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 $U \cap V \subset M$ 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:

$$\omega^V = \mathrm{Ad}_{\xi^{-1}}\omega^U + \xi^*\Omega, \tag{33.38}$$

where Ω is the Maurer-Cartan form on G. This formula holds more generally to (locally) relate the connection one-forms ω and $\xi^*\omega$ for any gauge transformation $\xi \in \mathrm{Aut}_V(P)$.

Example 33.3.28 (General linear group). Let $G = GL(n,\mathbb{R})$. The second term in Equation (33.38) can be written as follows:[2]

$$(\xi^*\Omega)^i{}_j = (\xi(m)^{-1})^i{}_k \frac{\partial}{\partial x^\mu}\xi(p)^k{}_j dx^\mu \tag{33.39}$$

at every point $m \in M$. Formally, this can be written coordinate-independently as

$$\xi^*\Omega = \xi^{-1}d\xi. \tag{33.40}$$

Example 33.3.29 (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 some manifold M, where the sections are induced by a choice 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 32.6.10). Using Equations (33.38) and (33.40) this becomes:

$$\overline{\Gamma}^i{}_{j\mu} = \frac{\partial y^\nu}{\partial x^\mu}\left(\frac{\partial x^i}{\partial y^k}\Gamma^k{}_{l\nu}\frac{\partial y^l}{\partial x^j} + \frac{\partial x^i}{\partial y^k}\frac{\partial^2 y^k}{\partial x^j \partial x^\nu}\right). \tag{33.41}$$

33.3.6 Parallel transport

Definition 33.3.30 (Horizontal lift). Consider a principal bundle (P, M, π, G) and a curve $\gamma : [0,1] \to M$. Given an Ehresmann connection Hor, 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. $\widetilde{\gamma}_{p_0}(0) = p_0$,
2. $\pi \circ \widetilde{\gamma}_{p_0} = \gamma$, and
3. $\widetilde{\gamma}'_{p_0}(t) \in \mathrm{Hor}(T_{\widetilde{\gamma}_{p_0}(t)}P)$ for all $t \in [0,1]$.

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 33.3.31 (Horizontal curve). Curves satisfying the last condition in the above property are said to be horizontal.

Method 33.3.32. Consider a principal bundle (P, M, π, G). Let γ be a curve in M and let ω be a principal connection one-form on P. For general structure groups G, the horizontal lift can be found as follows:

> Let δ be a curve in P that projects onto γ, i.e. $\pi \circ \delta = \gamma$, such that $\widetilde{\gamma}_{p_0}(t) = \delta(t) \cdot g(t)$ for some curve g in G. The latter curve g can be found as the unique solution of the following first-order ODE:
>
> $$\mathrm{Ad}_{g(t)^{-1}}\omega_{\delta(t)}(X_{\delta,\delta(t)}) + \Omega_{g(t)}(Y_{g,g(t)}) = 0, \tag{33.42}$$
>
> where X_δ, Y_g are tangent vectors to the curves δ and g respectively and where Ω is the Maurer-Cartan form on G. The solution is uniquely determined through the initial value condition $\delta(0) \cdot g(0) = p_0$.

[2]A derivation can be found in Lecture 22 of [114].

Remark 33.3.33. When 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 then yields $X_\delta = \sigma_*(X_\gamma)$. Using this one can rewrite the ODE as

$$\mathrm{Ad}_{g(t)^{-1}}\omega_{\delta(t)}(\sigma_* X_{\gamma,\gamma(t)}) + \Omega_{g(t)}(Y_{g,g(t)}) = 0. \tag{33.43}$$

After using the equality $f^*\omega = \omega \circ f_*$ and introducing the Yang-Mills field $A = \sigma^*\omega$, this becomes

$$\mathrm{Ad}_{g(t)^{-1}}A(X_{\gamma,\gamma(t)}) + \Omega_{g(t)}(Y_{g,g(t)}) = 0. \tag{33.44}$$

Example 33.3.34. 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

$$\widetilde{\gamma}(t) = \underbrace{(s \circ \gamma)(t)}_{\delta(t)} \cdot g(t) = (\gamma(t), g(t)),$$

where g is a curve in G. To determine $\widetilde{\gamma}$ it is thus sufficient to find g. The ODE (33.42) then becomes

$$g'(t) = -\omega(\gamma(t), e, \gamma'(t), 0)g(t). \tag{33.45}$$

Using the trivial section s one can further rewrite this formula. First, consider the action of the Yang-Mills field $s^*\omega$ on the derivative $\gamma_* = (\gamma(t), \gamma'(t))$. Using the fact that it is linear in the second argument, it can be rewritten as

$$s^*\omega(\gamma(t), \gamma'(t)) = A(\gamma(t))\gamma'(t),$$

where $A : M \to \mathrm{Hom}(\mathbb{R}^{\dim(M)}, \mathfrak{g})$ gives a linear map for each point $\gamma(t) \in M$. The action can also be rewritten using the relation $f^*\omega = \omega \circ f_*$ as

$$s^*\omega(\gamma(t), \gamma'(t)) = \omega\Big(s_*(\gamma(t), \gamma'(t))\Big) = \omega(\gamma(t), e, \gamma'(t), 0).$$

Combining these relations with the ODE (33.45) gives

$$\left(\frac{d}{dt} + A(\gamma(t))\gamma'(t)\right)g(t) = 0, \tag{33.46}$$

where $\frac{d}{dt}$ is the matrix given by element-wise multiplication of the derivative $\frac{d}{dt}$ and the identity matrix I.

The ODE (33.42) can now be solved. Direct integration and iteration gives

$$g(t) = \left[\mathbb{1} - \int_0^t dt_1 A(\gamma'(t_1)) + \int_0^t dt_1 \int_0^{t_1} dt_2 A(\gamma'(t_1))A(\gamma'(t_2)) - \ldots\right]g(0), \tag{33.47}$$

where A is the Yang-Mills field associated to the local section σ. This can be rewritten using the standard *Dyson trick* (see Formula 58.3.2):

$$g(t) = \left[\mathbb{1} - \int_0^t dt_1 A(\gamma'(t_1)) + \frac{1}{2!}\int_0^t dt_1 \int_0^t dt_2 \mathcal{T}\Big(A(\gamma'(t_1))A(\gamma'(t_2))\Big) - \ldots\right]g(0). \tag{33.48}$$

By noting that this formula is equal to the path-ordered exponential series one finds

$$g(t) = \mathcal{T}\exp\left(-\int_0^t dt' A(\gamma'(t'))\right)g(0). \tag{33.49}$$

Definition 33.3.35 (Parallel transport). The parallel transport map along the curve γ is defined as follows:

$$\mathrm{Par}_t^\gamma : \pi^{-1}(\gamma(0)) \to \pi^{-1}(\gamma(t)) : p_0 \mapsto \widetilde{\gamma}_{p_0}(t). \tag{33.50}$$

This map is G-equivariant and it restricts to an isomorphism on the fibres. The group element given by the path-ordered exponential in Equation (33.49) is generally 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 33.3.36 (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. Let (P, M, π, G) be a principal G-bundle over M and denote the delooping 4.10.2 of G by $\mathbf{B}G$. The assignment of holonomies to smooth paths locally defines a functor

$$\mathrm{hol}_i : \mathcal{P}_1(U_i) \to \mathbf{B}G \tag{33.51}$$

for every chart $U_i \subseteq M$. Globally, these can be glued together using the transition cocycles g_{ij} (in there incarnation as natural isomorphisms) to obtain a functor

$$\mathrm{hol} : \mathcal{P}_1(M) \to \mathbf{Trans}_1(P) \subset G\mathbf{Torsor}, \tag{33.52}$$

where $\mathbf{Trans}_1(P)$ is the full subcategory of the category of G-torsors on the fibres of P (Remark 33.1.2).

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 constructions as given above. ?? COMPLETE ??

33.3.7 Holonomy group

Definition 33.3.37 (Holonomy group). Consider a principal bundle (P, M, π, G) and choose a point $m \in M$. Let $\Omega_m^{ps} M \subset \Omega_m M$ denote the subset of the based loop space consisting of piecewise smooth loops with basepoint $m \in M$. The holonomy group $\mathrm{Hol}_p(\omega)$ based at $p \in \pi^{-1}(m)$ with respect to the connection form ω is given by

$$\mathrm{Hol}_p(\omega) := \{g \in G \mid p \sim p \cdot g\}, \tag{33.53}$$

where two points $p, q \in P$ are identified if there exists a loop $\gamma \in \Omega_m^{ps} M$ such that the horizontal lift $\widetilde{\gamma}$ connects p and q.

Definition 33.3.38 (Reduced holonomy group). The subgroup of the holonomy group induced by contractible loops.

Definition 33.3.39 (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:

$$H(p) := \{q \in P \mid p \sim q\}. \tag{33.54}$$

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 $\sqcup_p H(p)$ is in fact a principal bundle itself. Its structure group is $\mathrm{Hol}_p(\omega)$ for any $p \in P$.

[3] See Definition 42.3.20 for a rigorous exposition.

33.4 Covariant derivatives

33.4.1 Koszul connections

Definition 33.4.1 (Horizontal lifts on associated bundles). Let $P_F := P \times_G F$ be an associated bundle of a principal bundle (P, M, π, G) 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:

$$\widetilde{\gamma}^{P_F}_{[p,f]}(t) := [\widetilde{\gamma}_p(t), f]. \tag{33.55}$$

Although the element f seems to stay constant along the horizontal lift, it in fact changes according to Equation (33.3).

Definition 33.4.2 (Parallel transport). Similar to the case of principal bundles P, the parallel transport map on an associated bundle P_F is defined as

$$\mathrm{Par}^\gamma_t : \pi_F^{-1}(\gamma(0)) \to \pi_F^{-1}(\gamma(t)) : [p, f] \mapsto \widetilde{\gamma}^{P_F}_{[p,f]}(t). \tag{33.56}$$

Example 33.4.3 (Vector bundles). Consider a principal bundle (P, M, π, G) and suppose that the Lie group G acts on a vector space V through a representation $\rho : G \to \mathrm{GL}(V)$. One can construct an associated vector bundle $\pi_1 : P \times_{\mathrm{GL}(V)} V \to M$. Moreover, by working over a chart (U, φ) one can locally write 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 $\widetilde{\gamma}(t) = (\gamma(t), g(t))$ 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:
>
> $$\mathrm{Par}^\gamma_t : \pi_1^{-1}(x_0) \to \pi_1^{-1}(\gamma(t)) : (x_0, v_0) \mapsto \big(\gamma(t), \rho\big(g(t)h^{-1}\big)v_0\big). \tag{33.57}$$
>
> 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 33.4.4. 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 33.4.5 (Covariant derivative). Consider a vector bundle $\pi : E \to M$ with typical fibre V and its associated principal $\mathrm{GL}(V)$-bundle with principal connection ω. Let $\sigma : M \to E$ be a section of the vector bundle and let X be a vector field on M. The covariant derivative of σ with respect to X is defined as follows:

$$\nabla_X\sigma|_{x_0} := \lim_{t\to 0} \frac{(\mathrm{Par}^\gamma_t)^{-1}\sigma(\gamma(t)) - \sigma(x_0)}{t}, \tag{33.58}$$

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:

$$\nabla_X\sigma = \pi_*(\sigma_*X - X^H \circ \sigma). \tag{33.59}$$

One can also rephrase the above definition in terms of the horizontal vector field associated to the lift $\widetilde{\gamma}$ (akin to Definition 32.3.16). By Property 33.1.21 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

$$\phi(\nabla_X\sigma) = X^H(\phi(\sigma)), \tag{33.60}$$

where X^H acts componentwise on V.

Property 33.4.6. The map

$$\Gamma(TM) \times \Gamma(E) \to \Gamma(E) : (X, \sigma) \mapsto \nabla_X \sigma \tag{33.61}$$

defines a Koszul connection 32.6.1. It follows that every principal connection on a principal bundle induces a Koszul connection on all of its associated vector bundles.

33.4.2 Exterior covariant derivative

Definition 33.4.7 (Exterior covariant derivative). Let P be a principal bundle equipped with a principal connection ω and let $\theta \in \Omega^k(P)$ be a differential k-form. The exterior covariant derivative $D\theta$ is defined as follows:

$$D\theta(v_0, \ldots, v_k) := d\theta(v_0^H, \ldots, v_k^H), \tag{33.62}$$

where d is the exterior derivative 32.4.7 and v_i^H is the projection of v_i on the horizontal subspace $\mathrm{Hor}(T_pP)$. From this definition it follows that the exterior covariant derivative $D\theta$ is a horizontal form 33.3.13.

Remark 33.4.8. 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.

For tensorial forms of type (V, ρ) this is given by the following expression:

$$D\theta = d\theta + \omega \mathbin{\overline{\wedge}} \theta, \tag{33.63}$$

where $\overline{\wedge}$ denotes the combination of the wedge product and the action ρ.

Property 33.4.9 (Tensorial). If Φ is an equivariant form, then $D\Phi$ is a tensorial form.

The compatibility condition for connection one-forms (33.38) can be restated in terms of the covariant derivative:

Property 33.4.10 (Gauge transformation). Consider a principal bundle (P, M, π, G) and a connection one-form ω. For every gauge transformation $\xi \in \mathrm{Aut}_V(P)$ one (locally) has the following expression:

$$\xi^*\omega = \omega + \xi^{-1} D\xi, \tag{33.64}$$

where D is the exterior covariant derivative associated to ω.

Formula 33.4.11. Using the Koszul connection on the tangent bundle TP one can rewrite the action of the exterior covariant derivative as follows:

$$\begin{aligned} D\theta(v_0, \ldots, v_k) = \sum_i^k (-1)^i \nabla_{v_i} \theta(v_0, \ldots, \hat{v}_i, \ldots, v_k) \\ + \sum_{i<j}^k (-1)^{i+j} \theta([v_i, v_j], v_0, \ldots, \hat{v}_i, \ldots, \hat{v}_j, \ldots, v_k), \end{aligned} \tag{33.65}$$

where, as usual, $\hat{v}_i$ indicates that this vector is omitted. This formula should remind the reader of the analogous formula for the ordinary exterior derivative (32.43). As an example the formula for a one-form Φ is given:

$$D\Phi(X, Y) = \nabla_X(\Phi(Y)) - \nabla_Y(\Phi(X)) - \Phi([X, Y]). \tag{33.66}$$

Because of Property 33.1.21 one can use the following construction to find an explicit expression for the covariant derivative on an associated vector bundle:

Construction 33.4.12 (Covariant derivative). Let (P, M, π, G) be a principal 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 Equation (33.12). The exterior covariant derivative of ϕ is given by Equation (33.63):

$$D\phi(X) = d\phi(X) + \omega \triangleright \phi(X), \tag{33.67}$$

where $X \in T_pP$. Now, given an additional (local) section $\varphi : U \subseteq M \to P$, one can pull back this derivative to the base manifold M. This gives

$$(\varphi^* D\phi)(Y) = d(\varphi^*\phi)(Y) + \varphi^*\omega \triangleright \varphi^*\phi(Y), \tag{33.68}$$

where $Y = \pi_* X \in T_mM$. After introducing the notations $S := \varphi^*\phi$ and $\nabla_Y S := (\varphi^* D\phi)(Y)$ and remembering the definition of the Yang-Mills field 33.3.24, this becomes

$$\nabla_Y S = dS(Y) + \omega^U(Y) \triangleright S. \tag{33.69}$$

Example 33.4.13. Let $G = \mathrm{GL}(n, \mathbb{R})$. In local coordinates Equation (33.69) can be rewritten as follows:

$$(\nabla_Y S)^i = \frac{\partial S^i}{\partial x^k} Y^k + \Gamma^i{}_{jk} S^j Y^k. \tag{33.70}$$

This is exactly the formula known from classical differential geometry and relativity.

33.4.3 Curvature

Definition 33.4.14 (Curvature). Let ω be a principal connection one-form. The curvature Ω of ω is defined as the exterior covariant derivative $D\omega$.

Property 33.4.9 implies the following important statement:

Property 33.4.15 (Tensorial). In contrast to a connection one-form, the associated curvature is a tensorial $\mathfrak{g}$-valued two-form or, equivalently, an $\mathrm{End}(P)$-valued two-form.

Definition 33.4.16 (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 33.4.17 (Local systems). If the connection ∇ on a vector bundle is flat, the flat sections constitute a (linear) local system 9.3.2. 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.

Formula 33.4.18 (Curvature on associated bundles). The above definition of the curvature, together with Equation (33.63) or, equivalently, Construction 33.4.12, implies that one can express the action of the curvature on sections of associated bundles as follows:

$$D^2\phi = \Omega \triangleright \phi, \tag{33.71}$$

where $\phi \in \Omega^\bullet(M; E)$. This curvature form Ω coincides with the one from Definition 32.6.11.

Example 33.4.19. 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.

Property 33.4.20 (Second Bianchi identity). Let ω be a principal connection one-form with curvature Ω. The curvature is covariantly constant:

$$D\Omega = 0. \tag{33.72}$$

Remark 33.4.21. One should pay attention to the fact that this result does not generalize to arbitrary differential forms. Only the exterior derivative satisfies the coboundary condition $d^2 \equiv 0$, the exterior covariant derivative does not.

Formula 33.4.22 (Cartan structure equation). Let ω be a principal connection one-form and let Ω be its curvature form. The curvature can be expressed in terms of the connection as follows:

$$\Omega = d\omega + \frac{1}{2}[\omega \wedge \omega]. \tag{33.73}$$

The Maurer-Cartan equation in the (geometric) fundamental theorem of calculus 33.3.23 exactly states the vanishing of the algebraic curvature associated to a general $\mathfrak{g}$-valued one-form.

The following property is an immediate consequence of Frobenius's integrability theorem 32.5.5 and the fact that a connection vanishes on the horizontal subbundle:

Property 33.4.23 (Integrability). Let ω be a principal connection one-form. The associated horizontal distribution

$$p \mapsto \mathrm{Hor}(T_pP)$$

is integrable if and only if the connection ω is flat. In contrast, the vertical distribution is always integrable.

Similar to Definition 33.3.24 one can also define the Yang-Mills field strength:

Definition 33.4.24 (Field strength). Let $\pi : P \to M$ be a principal bundle equipped with a principal connection one-form ω and associated curvature Ω. Given a local section $\sigma : U \subseteq M \to P$, one defines the (Yang-Mills) field strength F as the pullback $\sigma^*\Omega$.

Theorem 33.4.25 (Ambrose-Singer). *The Lie algebra of the holonomy group* $\mathrm{Hol}_p(\omega)$ *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.*

33.4.4 Torsion

Definition 33.4.26 (Solder form). Let (P, M, π, G) be a principal bundle and let V be a $\dim(M)$-dimensional vector space equipped with a representation[4] $\rho : G \to \mathrm{GL}(V)$ such that $TM \cong P \times_G V$ as associated bundles. A solder(ing) form θ on P is a tensorial one-form 33.3.14 of type (V, ρ).

Definition 33.4.27 (Torsion). Let (P, M, π, G) be a principal bundle equipped with a principal connection ω and a solder form θ. The torsion Θ is defined as the exterior covariant derivative $D\theta$. This is the content of the **Cartan structure equation**:

$$\Theta = d\theta + \omega \barwedge \theta, \tag{33.74}$$

where the wedge product is defined analogously to the wedge products 32.4.22 and 32.4.27 using the induced representation of $\mathfrak{g}$ on V:

$$\omega \barwedge \theta(v, w) := \omega(v) \triangleright \theta(w) - \omega(w) \triangleright \theta(v). \tag{33.75}$$

Property 33.4.28 (First Bianchi identity). Let ω be a principal connection one-form, Ω its associated curvature, θ a solder form and Θ its associated torsion.

$$D\Theta = \Omega \barwedge \theta \tag{33.76}$$

[4] In general this will be $V = \mathbb{R}^{\dim(M)}$ and $G = \mathrm{GL}(n, \mathbb{R})$.

33.5 Reduction of the structure group

Definition 33.5.1 (Reduction). Consider a principal bundle $G \hookrightarrow P \to M$ 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 $\widetilde{P} \times_H G \to P$. This allows for morphisms besides inclusions, such as covering maps $\lambda : H \to G$. (See for example the definition of spinor bundles in Section 34.3.) As such the name "reduction" is not the best choice of terminology. For covering maps the term **lift(ing)** is sometimes used.

Definition 33.5.2 (G-structure). Consider a manifold M. A G-structure on M is the reduction of the structure group $\mathrm{GL}(n)$ of the frame bundle FM to the group $\iota : G \to \mathrm{GL}(n)$.

Definition 33.5.3 (Integrability). A G-structure P on M is said to be integrable if for every point $p \in M$ there exists a chart $U \ni p$ such that the associated holonomic frame $\{\partial_i\}_{i\leq\dim(M)}$ induces a local section of P.

Property 33.5.4. Consider a smooth manifold M equipped with a G-structure. If this structure is integrable, it admits a torsion-free connection.

Example 33.5.5 (Orientable manifold). An n-dimensional manifold is orientable if and only if the structure group can be reduced to $\mathrm{GL}^+(n)$, the group of invertible matrices with positive determinant. Furthermore, this structure is always integrable if it exists.

Example 33.5.6 (Riemannian manifold). An $\mathrm{O}(n)$-structure turns M into a *Riemannian manifold* 34.1.2. Because the cotangent bundle T^*M transforms under the contragredient representation, which coincides with the regular representation in the case of $\mathrm{O}(n)$, of the transition maps of the tangent bundle TM, these two bundles are equivalent. The isomorphism is given by the musical isomorphism(s) 34.1.3. Riemannian structures are always integrable.

The following property gives a classification of bundle reductions:

Property 33.5.7 (Equivariant morphisms). Consider a principal G-bundle P and let F be a set that admits a transitive action $\varphi : G \to \mathrm{Aut}(F)$. 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

$$P_f := \{p \in P \mid \psi(p) = f\}. \tag{33.77}$$

One can generalize this definition to arbitrary Lie group actions by restricting to the equivariant morphisms that take value in a single orbit.[5]

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 33.1.21, by the (global) sections of the associated coset bundle $P \times_G G/H$.

Corollary 33.5.8. If G is connected, every principal G-bundle is reducible to a maximal compact subgroup of G.

Definition 33.5.9 (Reducible connection). Consider a principal G-bundle P equipped with a connection one-form ω. If a bundle map F induces an H-reduction of P, then the connection ω is said to be reducible (and to be compactible with the given reduction) if $F^*\omega$ takes values in $\mathfrak{h}$.

[5]Since transitive actions have a unique orbit, this is a well-defined generalization.

Property 33.5.10. 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. $D\psi = 0$.

The following two properties characterize bundle reductions in terms of holonomy bundles:

Property 33.5.11 (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 33.5.12. 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 known in the literature:

Property 33.5.13 (Flat connections). A 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.

Remark 33.5.14. The above condition can also be applied to define flatness for topological bundles where the notion of connections does not make sense.

33.6 Characteristic classes

Definition 33.6.1 (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 $c(E) \in H^\bullet(M; R)$ such that if there exists a morphism $f : N \to M$, then $c(f^*E) = f^*c(E) \in H^\bullet(N; R)$. 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. *Stiefel-Whitney classes*).

Using the classification property 33.2.6, one can give a concise construction of characteristic classes in the case of principal bundles:

Construction 33.6.2. Consider a principal bundle (P, M, π, G) with classifying map $\varphi \in [M, BG]$. For every $c \in H^\bullet(BG)$ one defines a characteristic class $c(P) \in H^\bullet(M)$ as the pullback of c under φ.

As the definition implies, both vector bundles and principal bundles admit a theory of characteristic classes. However, in the literature most authors always 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 33.1.10 (see [119] for more information). The characteristic classes of a vector bundle are defined as the ones 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:

[6] Lifts as in the case of Spin-structures do not fall under the holonomy classification.

Definition 33.6.3 (Flag bundle). Let $\pi : E \to M$ be a vector bundle. Using the definition of the flag manifold 20.8.5 one can construct for every fibre E_p a space $\mathrm{Fl}(E_p)$ that has the complete flags of E_p as points (expressed as a sequence of one-dimensional subspaces). Using the bundle construction theorem, one can then obtain the flag bundle $\pi_{\mathrm{Fl}} : \mathrm{Fl}(E) \to M$ that has the flag manifolds as fibres.

Theorem 33.6.4 (Splitting principle). *Consider a vector bundle $\pi : E \to M$. Its flag bundle has the following properties:*

- *The pullback bundle $\pi^*_{\mathrm{Fl}}E$ can be decomposed as a Whitney sum of line bundles.*
- *The induced morphism on cohomology $\pi^*_{\mathrm{Fl}} : H^\bullet(M) \to H^\bullet(\mathrm{Fl}(E))$ is injective.*

For the following form of the splitting principle, see [88, 120].

Theorem 33.6.5 (Splitting principle). *Consider a principal bundle (P, M, π, G) where 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 $S^1 \cong \mathrm{U}(1)$. 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 $\rho : p^*B\iota \to M$ (also with fibre G/T). This fibre bundle has the following properties:*

- *ρ^*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 $H^\bullet(B\mathrm{U}(1))$. Therefore, every characteristic class of ρ^*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.

33.6.1 Chern-Weil theory

The characteristic classes of a vector bundle can be constructed from the connection and curvature forms on the vector bundle. The resulting expressions are polynomial in the curvature forms.

Definition 33.6.6 (Chern-Weil morphism). Let $\pi : E \to M$ be a vector bundle with structure group G and denote the connection one-form and curvature two-form by ω and Ω respectively. There exists a morphism of algebras

$$K[\mathfrak{g}]^G \to \Omega^\bullet(E) : P \mapsto P(\Omega), \tag{33.78}$$

where K is the base field, satisfying:

- $P(\Omega)$ is closed.
- $P(\Omega)$ pulls back uniquely to a (closed) form $\overline{P}(\Omega) := \pi^*P(\Omega)$ on M.
- $\overline{P}(\Omega)$ does not depend on the chosen connection, i.e. for two connection one-forms ω, ω', the difference $\overline{P}(\Omega) - \overline{P}(\Omega')$ is exact.

In the remainder of this section this approach will be followed to find explicit descriptions of characteristic classes of vector bundles and principal bundles.

33.6.2 Complex bundles

In this section only complex bundles are considered. This allows for the choice of $\mathfrak{u}(n)$-valued connection one-forms. See Chapter 37 for more information.

Definition 33.6.7 (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 $c_k(E)$ as follows:

$$\det\left(\mathbb{1} + \frac{it}{2\pi}\Omega\right) =: \sum_{k=1}^{n} c_k(E)t^k. \tag{33.79}$$

The k^{th} Chern class is a cohomology class in $H^{2k}(M)$.

Definition 33.6.8 (Chern polynomial). Let $c_k(E)$ denote the k^{th} Chern class of E. The Chern polynomial is defined as follows:

$$c_t(E) := \sum_{k=1}^{\infty} c_k(E)t^i. \tag{33.80}$$

The **total Chern class** is defined by taking $t = 1$.

Definition 33.6.9 (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:

$$\mathrm{ch}(E) := \mathrm{tr}\left(\exp\left(\frac{i\Omega}{2\pi}\right)\right). \tag{33.81}$$

If $c_i := c_i(E)$ denotes the i^{th} Chern class of E, the Chern character can also be expressed as

$$\mathrm{ch}(E) = \sum_{k=0}^{n} \frac{c_1^k + \cdots + c_n^k}{k!}. \tag{33.82}$$

The term with prefactor $1/k!$ is a homogeneous polynomial of degree k. One sometimes calls this term the k^{th} Chern character. Using Chern-Weil theory, this form is proportional to $\mathrm{tr}(\Omega^k)$.

Formula 33.6.10 (Whitney product formula[7]). The following equality holds for all bundles E_1, E_2:

$$c_t(E_1 \oplus E_2) = c_t(E_1)c_t(E_2). \tag{33.83}$$

Corollary 33.6.11 (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:

$$c_t(E) = \prod_{i=1}^{n}(1 + x_i t), \tag{33.84}$$

where in the case of decomposable vector bundles $E \equiv \oplus_{i=1}^n L_i$ 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, i.e. the k^{th} Chern class, is given by the k^{th} elementary symmetric polynomial:

$$c_k(E) = \sum_{i_1 < \cdots < i_k} x_{i_1} \cdots x_{i_k}. \tag{33.85}$$

[7]This formula is also called the **Whitney sum formula**.

Definition 33.6.12 (Canonical class). Consider a smooth manifold M. The first Chern class of the canonical bundle $\bigwedge^n T^*M$ is called the canonical class of M.

Definition 33.6.13 (Theta characteristic). Consider a smooth manifold M together 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 in cohomology 32.8.6.

After finding the Chern roots of a vector bundle E, one can use them to define various other classes:

Construction 33.6.14 (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]

$$G_f(E) := \det f\left(\frac{it}{2\pi}\Omega\right). \tag{33.86}$$

The coefficients of this power series define characteristic classes of E.

Example 33.6.15 (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 33.6.16 (Todd class). Consider the function

$$Q(x) := \frac{x}{1-e^{-x}} = 1 + \frac{x}{2} + \sum_{i=1}^{\infty} \frac{(-1)^{i-1}B_i}{(2i)!}x^{2i}, \tag{33.87}$$

where B_i is the i^{th} *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

$$\mathrm{td}(E) := \prod_{i=1}^{n} Q(x_i). \tag{33.88}$$

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 n^{th} degree term in $f(x)^{n+1}$ has coefficient 1.

Another genus that is used in the context of the index theorems is the following one:

Example 33.6.17 ($\hat{A}$-genus[9]). The $\hat{A}$-genus is defined through the following function:

$$Q(x) := \frac{\sqrt{x}/2}{\sinh(\sqrt{x}/2)} = 1 - \frac{x}{24} + \frac{7x^2}{5760} - \cdots. \tag{33.89}$$

33.6.3 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 one-forms, one can also define a set of characteristic classes.

[8]In the case that E splits as a sum for line bundles, one simply obtains the product $f(x_1)\cdots f(x_k)$.

[9]This is pronounced as *A-roof genus*.

Definition 33.6.18 (Pontryagin class). Consider a vector bundle $\pi : E \to M$. The Pontryagin classes of E are defined as follows:

$$p_k(E) := (-1)^k c_{2k}(E^{\mathbb{C}}) \in H^{4k}(M), \tag{33.90}$$

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.

$$p_1(E) = c_1^2(E) - 2c_2(E). \tag{33.91}$$

When the vector bundles in question are orientable, the structure group can further be reduced to $\mathrm{SO}(n)$. If the rank is even, one can define the following characteristic class:

Definition 33.6.19 (Euler class). Let $\pi : E \to M$ be an orientable vector bundle of rank $2k$. The Euler class of E is defined as follows:

$$e(E) := p_k(E) \cup p_k(E). \tag{33.92}$$

Property 33.6.20. Using the fact that one can write the total Pontryagin class using Chern-Weil theory as

$$p(E) = \det\left(1 - \frac{1}{2\pi}\Omega\right) \tag{33.93}$$

and that the determinant is the square of the *Pfaffian*, one can equivalently define the Euler class as follows:

$$e(E) := \mathrm{Pf}\left(-\frac{1}{2\pi}\Omega\right). \tag{33.94}$$

33.6.4 Cohomology of Lie groups

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:

$$H^\bullet(BG;\mathbb{Z}) \cong H_c^\bullet(G;\mathbb{Z}). \tag{33.95}$$

?? COMPLETE ??

33.6.5 Chern-Simons forms

By Chern-Weil theory, 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 (locally) one can find a trivialization:

$$\langle \Omega_A, \cdots \rangle_n = d\mathrm{CS}^n(A), \tag{33.96}$$

where $\langle \cdots \rangle_n$ denotes an invariant polynomial of degree $2n$. 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 $\hat{A}$. This connection defines a path between two connections A, A' on P. The relative Chern-Simons form is defined as

$$\mathrm{CS}^n(A, A') := \int_{[0,1]} \langle \Omega_{\hat{A}}, \cdots \rangle_n. \tag{33.97}$$

The differential of this form gives the difference of characteristic forms:

$$d\mathrm{CS}^n(A, A') = \langle \Omega_{A_1}, \cdots \rangle_n - \langle \Omega_{A_0}, \cdots \rangle_n. \tag{33.98}$$

However, note that the Chern-Simons form is only defined up to an exact form.

Example 33.6.21 (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

$$\langle dA, A\rangle + \frac{2}{3}\langle A, [A \wedge A]\rangle. \tag{33.99}$$

This form is the exterior derivative of the second Chern character, which for $\mathfrak{su}(n)$-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.

33.7 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 are able to uniquely characterize the bundle.

33.7.1 Differential characters

In this section all (co)chains, (co)cycles and (co)boundaries are assumed to be smooth. By doing this no generality is lost since every continuous chain is homotopic to a smooth one.

Definition 33.7.1 (Differential character). Consider a positive integer $k \geq 1$ and let M be a manifold. A (**Cheeger-Simons**) differential character of **degree** k is a group homomorphism $\chi : Z_{k-1}(M) \to \mathrm{U}(1)$ that is given by integration on boundaries:[10]

$$\chi(\partial\gamma) = \exp\left(2\pi i \int_\gamma \omega(\chi)\right) \tag{33.100}$$

for some $\omega(\chi) \in \Omega^k(M)$. The group of differential characters of degree k is denoted by $\hat{H}^k(M;\mathbb{Z})$. For $k = 0$ the convention $\hat{H}^0(M;\mathbb{Z}) := H^0(M;\mathbb{Z})$ is used.

Property 33.7.2 (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 33.7.3 (Curvature). Every differential character is represented by a unique, closed and integral k-form. The map $\mathrm{curv} : \hat{H}^k(M;\mathbb{Z}) \to \Omega^k_{\mathrm{int}}(M) : \chi \mapsto \omega(\chi)$ is called the curvature map. If $\mathrm{curv}(\chi) = 0$, the character χ is said to be **flat**.

Property 33.7.4 (Characteristic class). Every differential character gives rise to a characteristic class as follows. The group of cocycles is free and the quotient map $\mathbb{R} \to \mathrm{U}(1)$ is onto, so every differential character lifts to a group homomorphism $\widetilde{\chi} : Z_{k-1}(M) \to \mathbb{R}$ such that $\chi(z) = \exp(2\pi i\widetilde{\chi}(z))$. The map

$$\mathrm{ch}(\chi) : C_k(M) \to \mathbb{Z} : \gamma \mapsto \int_\gamma \mathrm{curv}(\chi) - \widetilde{\chi}(\partial\gamma) \tag{33.101}$$

induces a well-defined map $\mathrm{ch} : \hat{H}^k(M;\mathbb{Z}) \to H^k(M;\mathbb{Z})$. If $\mathrm{ch}(\chi) = 0$, 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. [46]).

[10]Some authors omit the exponential function by working modulo $\mathbb{Z}$. This just replaces the multiplicative U(1)-group by the isomorphic additive $\mathbb{R}/\mathbb{Z}$-group.

Example 33.7.5 (Circle bundles). Consider a U(1)-bundle $\pi : P \to M$ with connection ω. Holonomy around closed curve γ gives a parallel transport map

$$P \to P : p \mapsto p \cdot g(p, \gamma) \tag{33.102}$$

for a smooth function $g : \Omega_p P \to \mathrm{U}(1)$. In fact, g only depends on the homology of γ and the projection $\pi(p)$, so one obtains a map $g \in \hat{H}^2(M;\mathbb{Z})$ with curvature $\frac{-1}{2\pi i}\Omega$ and characteristic class $c_1(P)$. The converse also holds, every different character of degree 2 determines a principal U(1)-bundle with connection (up to connection-preserving isomorphism). This leads to the following equivalence:

$$\hat{H}^2(M;\mathbb{Z}) \cong \{\text{isomorphism classes of } (P,\nabla) \mid P \text{ a circle bundle and } \nabla \text{ a principal connection}\}. \tag{33.103}$$

The curvature and characteristic class maps fit in some exact sequences:

Property 33.7.6 (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:

$$0 \longrightarrow H^{k-1}(M;\mathrm{U}(1)) \longrightarrow \hat{H}^k(M;\mathbb{Z}) \xrightarrow{\text{curv}} \Omega^k_{\text{int}}(M) \longrightarrow 0. \tag{33.104}$$

The first cohomology group classifies flat circle bundles by Property 33.5.13, 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 33.7.7 (Characteristic class exact sequence).

$$0 \longrightarrow \Omega^{k-1}(M)/\Omega^{k-1}_{\text{int}}(M) \longrightarrow \hat{H}^k(M;\mathbb{Z}) \xrightarrow{\text{ch}} H^k(M;\mathbb{Z}) \longrightarrow 0 \tag{33.105}$$

The first map is induced by the holonomy functional

$$\iota : \Omega^{k-1}(M) \to \hat{H}^k(M;\mathbb{Z}) : \omega \to \exp\left(2\pi i \int_{-} \omega\right), \tag{33.106}$$

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).

33.7.2 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 33.7.8 (Differential cocycle). A tuple $(c, h, \omega) \in C^n(M;\mathbb{Z}) \times C^{n-1}(M) \times \Omega^n(M)$ such that

$$\delta c = 0 \tag{33.107}$$

$$d\omega = 0 \tag{33.108}$$

$$\delta h = \omega - c. \tag{33.109}$$

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

$$d : (c, h, \omega) \mapsto (\delta c, \omega - c - \delta h, d\omega) \tag{33.110}$$

defines a cohomology theory $\hat{H}(n)^\bullet(M)$.

Property 33.7.9 (Relation to differential characters). Differential characters and differential cocycles are related as follows:

$$\hat{H}^k(M) \cong \hat{H}(k)^k(M). \tag{33.111}$$

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 Equation (33.101) is exactly the third relation in the definition of cocycles above. The mod $\mathbb{Z}$-reduction of h gives the differential character associated to the cocycle.

Example 33.7.10. The first ordinary differential cohomology group $\hat{H}^1(M;\mathbb{Z})$ is isomorphic to the group of smooth functions $C^\infty(M;\mathrm{U}(1))$.

33.7.3 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 33.7.11 (Simons-Sullivan). *There is an essentially unique functor*

$$\hat{H}^\bullet(-;\mathbb{Z}) : \mathbf{Diff} \to \mathbf{Ab}^{\mathbb{Z}}$$

such that there exist four natural transformations

1. **Flat class**: $j : H^{\bullet-1}(-;\mathrm{U}(1)) \to \hat{H}^\bullet(-;\mathbb{Z})$,
2. **Topological trivialization**: $\iota : \Omega^{\bullet-1}/\Omega^{\bullet-1}_{\mathrm{int}} \to \hat{H}^\bullet(-;\mathbb{Z})$,
3. **Characteristic class**: $\mathrm{ch} : \hat{H}^\bullet(-;\mathbb{Z}) \to H^\bullet(-;\mathbb{Z})$, *and*
4. **Curvature**: $\mathrm{curv} : \hat{H}^\bullet(-;\mathbb{Z}) \to \Omega^\bullet_{\mathrm{int}}$

that fit in the following commutative diagram, where the diagonal sequences are exact:

$$\begin{tikzcd}
0 \arrow[dr] & & & & 0 \\
& H^{\bullet-1}(-;\mathrm{U}(1)) \arrow[rr, "\text{Bockstein}"] \arrow[dr, "j"'] & & H^\bullet(-;\mathbb{Z}) \arrow[ur] \arrow[dr, "\text{de Rham}"] & \\
H^{\bullet-1}_{\mathrm{dR}} \arrow[ur] \arrow[dr, two heads] & & \hat{H}^\bullet(-;\mathbb{Z}) \arrow[ur, "\mathrm{ch}"] \arrow[dr, "\mathrm{curv}"] & & H^\bullet_{\mathrm{dR}} \\
& \Omega^{\bullet-1}/\,\Omega^{\bullet-1}_{\mathrm{int}} \arrow[ur, "\iota"] \arrow[rr, "d"'] & & \Omega^\bullet_{\mathrm{int}} \arrow[ur, hook] \arrow[dr] & \\
0 \arrow[ur] & & & & 0
\end{tikzcd}$$

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 33.7.12 (Deligne complex). Let $\mathbf{B}^k\mathrm{U}(1)_{\mathrm{conn}}$ denote the cochain complex

$$\mathcal{O}^\times_M \xrightarrow{d\log} \Omega^1 \xrightarrow{d} \cdots \xrightarrow{d} \Omega^k. \tag{33.112}$$

(Smooth) Deligne cohomology is defined as follows:

$$H_D^{k+1}(M;\mathbb{Z}) := \check{H}^0(M;\mathbf{B}^k\mathrm{U}(1)_{\mathrm{conn}}), \tag{33.113}$$

where $\check{H}^\bullet$ denotes Čech cohomology 9.3.10 and the cochain complex $\mathbf{B}^k\mathrm{U}(1)_{\mathrm{conn}}$ is turned into a cochain complex by inverting the degrees.

Property 33.7.13 (Deligne-Beilinson product). Consider the Deligne complex for two integers $k,l\in\mathbb{N}$. There exists a cup product

$$\cup:\mathbf{B}^k\mathrm{U}(1)_{\mathrm{conn}}\otimes\mathbf{B}^l\mathrm{U}(1)_{\mathrm{conn}}\rightarrow\mathbf{B}^{k+l+1}\mathrm{U}(1)_{\mathrm{conn}}:x\otimes y\mapsto x\cup y:=\begin{cases}x\wedge dy & \deg(y)=l\\ 0 & \text{otherwise.}\end{cases} \tag{33.114}$$

Example 33.7.14 (Circle bundles). A (Čech-)Deligne cocycle in degree 2 consists of data (A_i,g_{ij}) such that

$$\begin{tikzcd}
A_i \arrow[r, hook] & A_i - A_j = d\log g_{ij} = g_{ij}^{-1}dg_{ij} \\
g_{ij} \arrow[u, "d\log"] \arrow[r, hook] & g_{jk}g_{ki}^{-1}g_{ij} = 1,
\end{tikzcd}$$

where the inclusion arrows denote the restriction to intersections $U_{ij}:=U_i\cap U_j$. Property 33.3.27 and the subsequent example, specialized in the case of U(1)-bundles, show that the above data are exactly the components of a princpal circle bundle with connection.

Remark 33.7.15. As was the case for differential characters, higher Deligne cohomology classes classify higher U(1)-bundles with connection. The main benefit of this approach is that one gets an ‘‘explicit’’ description of the local data. See [117] for a good introduction.

Remark 33.7.16 (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

$$(\delta\beta)_{i_0\ldots i_k} = \alpha_{i_0\ldots i_k}, \tag{33.115}$$

i.e. a trivial k-bundle is equivalent to a twisted $(k-1)$-bundle.

?? COMPLETE ??

33.8 Cartan connections

In the first part of this section a short overview of Klein's **Erlangen program** that 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 [32].

33.8.1 Klein geometry

Definition 33.8.1 (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 and G is called the **principal group**. If the principal space 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 33.8.2. It is clear that every Klein geometry gives a homogeneous space and, hence, a principal bundle of rank $\dim(G) - \dim(H)$.

Definition 33.8.3 (Effective Klein pair). The action of G on G/H is not necessarily effective, i.e. the kernel

$$\ker(\rho) = \{x \in G \mid \forall g \in G : g^{-1}xg \in H\}, \tag{33.116}$$

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 33.8.4 (Reductive Klein pair). A Klein pair $(\mathfrak{g}, \mathfrak{h})$ is said to be reductive if $\mathfrak{g}$ admits a decomposition of the form

$$\mathfrak{g} = \mathfrak{h} + \mathfrak{m}, \tag{33.117}$$

where $\mathfrak{m}$ is an $\mathfrak{h}$-module.

Example 33.8.5 (Euclidean space). Consider the Euclidean group $\mathrm{Euc}(n) := \mathbb{R}^n \rtimes \mathrm{O}(n)$, i.e. the symmetry group of the Euclidean space $\mathbb{R}^n$. This group clearly acts transitively and the subgroup $\mathrm{O}(n)$ 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 $\mathrm{Euc}(n)/\mathrm{O}(n)$.

Definition 33.8.6 (Model geometry). A model geometry consists of the following data:

- an effective Klein pair $(\mathfrak{g}, \mathfrak{h})$,
- a Lie group H such that $\mathrm{Lie}(H) = \mathfrak{h}$, and
- a representation $\mathrm{Ad} : H \to \mathrm{Aut}(\mathfrak{g})$ that restricts to the adjoint representation $\mathrm{Ad}_H : H \to \mathrm{Aut}(\mathfrak{h})$.

Definition 33.8.7 (Local Klein geometry). A local Klein geometry consists of the following data:

- a Lie group G,
- a closed subgroup $H \subset G$, and
- a subgroup $\Gamma \subset G$ acting by covering transformations on G/H such that the left coset space $\Gamma\backslash G/H$ is connected.

33.8.2 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 given:

Alternative Definition 33.8.8 (Lie group). Let M be a smooth manifold and let $\mathfrak{g}$ be a Lie algebra. Assume that M comes equipped with a $\mathfrak{g}$-valued one-form ω satisfying the following conditions:

1. **Maurer-Cartan equation**: $d\omega + \frac{1}{2}[\omega, \omega] = 0$,
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 THIS PROPERTY ??

In a similar way Klein geometries can be characterized as follows:

Property 33.8.9. The bundle $\pi : G \to G/H$ of a Klein geometry G/H admits a one-form $\omega : TG \to \mathfrak{g}$ that satisfies the following conditions:

- ω restricts to an isomorphism on each fibre.
- ω is H-equivariant: $R_h^*\omega = \mathrm{Ad}(h^{-1})\omega$.
- ω cancels $\mathfrak{h}$-fundamental vector fields: $\omega(A^\#) = A$ for all $A \in \mathfrak{h}$.
- ω satisfies the Maurer-Cartan equation.
- ω is complete.

The second and third conditions show that ω defines a principal connection one-form, while the fourth condition states that this connection is flat. In fact this one-form is exactly the Maurer-Cartan form on G, where conditions 3 and 4 are obtained by restricting to the subgroup $H \subset G$.

By dropping the flatness and completeness conditions, one obtain the notion of Cartan connections:

Definition 33.8.10 (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 defined by a one-form $\omega : TP \to \mathfrak{g}$ satisfying the following conditions:

1. ω restricts to an isomorphism on each fibre.
2. ω is H-equivariant.
3. ω cancels $\mathfrak{h}$-fundamental vector fields: $\omega(A^\#) = A$ for all $A \in \mathfrak{h}$.

The form ω is called the Cartan connection.

Definition 33.8.11 (Curvature). By analogy with the Maurer-Cartan condition and the Cartan structure equation 33.4.22, the curvature of a Cartan connection is defined as follows:

$$\Omega := d\omega + \frac{1}{2}[\omega \wedge \omega]. \tag{33.118}$$

By restricting to reductive model spaces an important decomposition of the Cartan connection is obtained:

Property 33.8.12. Consider a Cartan geometry $\pi : P \to M$ with a reductive model space $(\mathfrak{g}, \mathfrak{h})$ such that the Cartan connection can be decomposed 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 $\omega_{\mathfrak{m}}$.

?? COMPLETE ??

33.9 $\mathcal{D}$-geometry ♣

Definition 33.9.1 (Sheaf of differential operators). Let X be a smooth manifold (or variety, see Chapter 11) and denote its structure sheaf by $\mathcal{O}_X$. $\mathcal{D}_X$ denotes the sheaf of $\mathcal{O}_X$-algebras of vector fields (as derivations) on X:

$$\mathcal{D}_X(U) := \{v \in \mathrm{End}(\mathcal{O}_X(U)) \mid v(fg) = v(f)g + fv(g)\}. \tag{33.119}$$

Definition 33.9.2 ($\mathcal{D}_X$-module). An $\mathcal{O}_X$-module equipped with an action of $\mathcal{D}_X$. This is equivalent to a linear map

$$\nabla : \mathcal{D}_X \to \mathrm{End}(\mathcal{O}_X) \tag{33.120}$$

satisfying the following properties:

1. **$\mathcal{O}_X$-linearity**: $\nabla_{fv}\sigma = f\nabla_v\sigma$.
2. **Leibniz rule**: $\nabla_v(f\sigma) = v(f)\sigma + f\nabla_v\sigma$.
3. **Flatness**: $\nabla_{[v,w]}\sigma = [\nabla_v, \nabla_w]\sigma$.

If X is *locally free*, i.e. corresponds to a locally trivial manifold (a bundle), one obtains the algebraic reformulation of a vector bundle with flat connection.

Chapter 34

Riemannian Geometry

The main reference for this chapter is [63].

34.1 Riemannian manifolds

34.1.1 Metric

Definition 34.1.1 (Bundle metric). Consider the bundle of $(0,2)$-tensors. From Definition 21.3.1 it follows that every section g of this bundle defines a bilinear map

$$g_p : T_pM \times T_pM \to \mathbb{R}$$

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 $\{\langle\cdot|\cdot\rangle_p : p \in M\}$ is called a **bundle metric** or **fibre metric**.

Definition 34.1.2 (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 23.2.7). This is given by the *flat* and *sharp* isomorphisms:

Definition 34.1.3 (Musical isomorphisms). Let $g : TM \times_M TM \to \mathbb{R}$ be a Riemannian metric on M. The **flat** isomorphism is defined as follows:

$$\flat : v \mapsto g(v, \cdot). \tag{34.1}$$

The **sharp** isomorphism is defined as the inverse map. For an arbitrary covector field ω it is implicitly given by

$$g(\omega^\sharp, v) = \omega(v). \tag{34.2}$$

These **musical isomorphisms** can be used to raise and lower tensor indices. In index notation they are given by contraction with metric tensor:

$$\begin{aligned} \flat &: v^\mu \mapsto v_\lambda := g_{\lambda\mu}v^\mu \\ \sharp &: \omega_\mu \mapsto \omega^\lambda := g^{\lambda\mu}\omega_\mu. \end{aligned}$$

[1] Recall Section 20.3 about Hermitian forms and metric forms.

Definition 34.1.4 (Codifferential). Using the de Rham differential and the Hodge star operator 21.4.26, one can define a boundary operator $\delta : \Omega^k(M) \to \Omega^{k-1}(M)$:

$$\delta := (-1)^k *^{-1} d* = (-1)^{n(k+1)+1} * d* . \tag{34.3}$$

It is not hard to check that this is indeed a boundary operator in the sense of Definition 5.1.1.

Definition 34.1.5 (Hodge Laplacian[2]). Using the de Rham differential and codifferential one can define a derivation on $\Omega^k(M)$:

$$\Delta := d\delta + \delta d. \tag{34.4}$$

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** $d + \delta$.

Definition 34.1.6 (Harmonic form). An element of the kernel of the Hodge Laplacian Δ. The space of harmonic k-forms is often denoted by $\mathcal{H}^k(M)$.

Theorem 34.1.7 (Hodge decomposition). *Let M be a closed Riemannian manifold. Every differential k-form admits a decomposition of the form*

$$\omega = d\alpha + \delta\beta + h, \tag{34.5}$$

where $h \in \mathcal{H}^k(M)$.

Corollary 34.1.8 (Hodge theorem). The k^{th} de Rham cohomology group is isomorphic (as an Abelian group/vector space) to the space of harmonic k-forms:

$$H^k(M) \cong \mathcal{H}^k(M). \tag{34.6}$$

Remark 34.1.9. The gradient, rotor (curl) and divergence from standard vector calculus (Section 21.1) can be rewritten and generalized using exterior derivatives as follows. Consider a smooth function f and a smooth vector field X.

$$\nabla f = df^{\,\sharp} \tag{34.7}$$

$$\nabla \times X = \left(* \, dX^\flat\right)^\sharp \tag{34.8}$$

$$\nabla \cdot X = \left(*^{-1} d * X^\flat\right)^\sharp, \tag{34.9}$$

Property 21.1.13 now follows from the nilpotency $d^2 = 0$.

Explanation. Looking at Equation (32.37) for the exterior derivative of a smooth function and remembering the definition of the gradient 21.1.1 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 df 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 21.1.8 and rotor 21.1.10. However, here one has to use a different construction because the spaces Λ^1 and Λ^2 have to be used. The Hodge star 21.4.26 can be used to obtain the correct dimensions.

For simplicity we will give a derivation in $\mathbb{R}^3$ with the Euclidean metric. 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 $X^\flat = f_1 dx_1 + f_2 dx_2 + f_3 dx_3$ and a two-form $*X^\flat = f_1 dx_2 \wedge dx_3 + f_2 dx_3 \wedge dx_1 +$

[2] Sometimes called the **Hodge-de Rham** or **Laplace-de Rham** operator.

$f_3dx_1 \wedge dx_2$. After applying the exterior derivative one obtains

$$dX^\flat = \left(\frac{\partial f_3}{\partial x_2} - \frac{\partial f_2}{\partial x_3}\right)dx_2 \wedge dx_3 + \left(\frac{\partial f_1}{\partial x_3} - \frac{\partial f_3}{\partial x_1}\right)dx_3 \wedge dx_1 + \left(\frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2}\right)dx_1 \wedge dx_2,$$
$$d * X^\flat = \left(\frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} + \frac{\partial f_3}{\partial x_3}\right)dx_1 \wedge dx_2 \wedge dx_3.$$

34.1.2 Riemannian manifolds

Definition 34.1.10 (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 34.1.11. Let M be a pseudo-Riemannian manifold. For every $p \in M$ there exists a splitting $T_pM = P \oplus N$ 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 34.1.12 (Index and signature). Let M be a connected pseudo-Riemannian manifold. The dimension of the "negative" subspace N in the above splitting $T_pP = P \oplus N$ is called the index of the manifold. The **signature** is defined as the pair $(\dim(P), \dim(N))$.

Definition 34.1.13 (Hyperbolic manifold). A pseudo-Riemannian manifold with index 1.

Theorem 34.1.14 (Whitney's embedding theorem). *Every smooth paracompact manifold M can be embedded in $\mathbb{R}^{2\dim(M)}$.*

Theorem 34.1.15 (Whitney's immersion theorem). *Every smooth paracompact manifold M can be immersed in $\mathbb{R}^{2\dim(M)-1}$.*

The following theorem is slightly stronger:

Theorem 34.1.16 (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 1's in the binary expansion of n.*

Definition 34.1.17 (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 $t^2g + dt \otimes dt$. This manifold is called the Riemannian cone or **metric cone** of (M, g).

34.1.3 Levi-Civita connection

Definition 34.1.18 (Riemannian connection). An affine connection ∇ on (M, g) is said to be Riemannian if it satisfies the following two conditions:

1. **metric-compatibility**):

$$\nabla g = 0 \tag{34.10}$$

or, equivalently,

$$X(g(Y, Z)) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z) \tag{34.11}$$

for all vector fields X, Y and Z.

2. **torsion-freeness**:

$$\nabla_X Y - \nabla_Y X = [X, Y] \tag{34.12}$$

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 $O(n)$-structure by Example 33.5.6, Properties 33.5.4 and 33.5.10 imply that there exists a torsion-free connection that preserves this structure. The following theorem gives an even sharper result:

Theorem 34.1.19 (Fundamental theorem). *Every pseudo-Riemannian manifold admits a unique Levi-Civita connection.*

Formula 34.1.20 (Koszul formula). The Levi-Civita connection ∇ on a pseudo-Riemannian manifold (M, g) is implicitly defined by the following formulae:

$$\begin{aligned} 2g(\nabla_X Y, Z) &= \mathcal{L}_X g(Y, Z) + d(\iota_X g)(Y, Z) && (34.13)\\ &= X(g(Y, Z)) + Y(g(Z, X)) - Z(g(X, Y)) \\ &\qquad + g([X, Y], Z) - g([Z, X], Y) - g([Y, Z], X). && (34.14)\end{aligned}$$

The following (local) formula is often useful in calculations (especially in general relativity, see Chapter 66):

Formula 34.1.21 (Divergence). Consider the Riemannian volume form

$$\mathrm{Vol} = \sqrt{\det(g)}dx^1 \wedge \cdots \wedge dx^n. \tag{34.15}$$

The divergence of a vector field X is defined as follows:

$$\mathcal{L}_X \,\mathrm{Vol} =: \mathrm{div}(X)\,\mathrm{Vol}\,. \tag{34.16}$$

With respect to an orthonormal frame field this is equivalent to

$$\mathrm{div}(X) = \frac{1}{2}\sum_{i=1}^{n}(\mathcal{L}_X g)(e_i, e_i). \tag{34.17}$$

Let ∇ be the Levi-Civita connection. Using the Koszul formula one can show that the above formula implies the following equality

$$\mathrm{div}(X) = \mathrm{tr}(Y \mapsto \nabla_Y X) \equiv \nabla_\mu X^\mu, \tag{34.18}$$

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 21.1.8 from vector calculus.

Using the metric determinant one can locally write the divergence in terms of ordinary partial derivatives:

$$\nabla_\mu V^\mu = \frac{1}{\sqrt{\det(g)}}\partial_\mu(\sqrt{\det(g)}V^\mu). \tag{34.19}$$

Definition 34.1.22 (Laplace-Beltrami operator). Consider a Riemannian manifold (M, g). The Laplace-Beltrami operator on M is defined as the Laplace operator 21.1.20, i.e. as the divergence of the gradient:

$$\Delta := \mathrm{tr}\,\nabla^2. \tag{34.20}$$

This is, up to a sign[3], equal to the **Bochner Laplacian**:

$$\Delta = -\nabla^*\nabla. \tag{34.21}$$

[3]Some authors already include a minus sign in the definition of the Laplace-Beltrami operator.

The geodesic equation 28.2.47 can be generalized as follows:

Definition 34.1.23 (Geodesic). A curve γ on a Riemannian manifold (M, g) that is autoparallel with respect to the Levi-Civita connection:

$$\nabla_{\dot{\gamma}}\dot{\gamma} = 0. \tag{34.22}$$

?? COMPLETE (HESSIAN, ...) ??

34.1.4 Killing vectors

Definition 34.1.24 (Killing vector). Let (M, g) be a pseudo-Riemannian manifold. A vector field X satisfying

$$\mathcal{L}_X g = 0 \tag{34.23}$$

is called a Killing vector field.

Formula 34.1.25. Given a Levi-Civita connection ∇ on (M, g) one can rewrite the Killing condition as follows:

$$\nabla_{(\mu}X_{\nu)} = 0. \tag{34.24}$$

Definition 34.1.26 (Killing tensor). Let ∇ be the Levi-Civita connection on (M, g). A tensor T satisfying

$$\nabla_{(\mu_N}T_{\mu_1\ldots\mu_{N-1})} = 0 \tag{34.25}$$

is called a Killing tensor.

34.2 Curvature

Formula 34.2.1 (Riemann curvature tensor). The Riemann (curvature) tensor R is defined as the following $(1, 3)$-tensor:

$$R(v, w)z := [\nabla_v, \nabla_w]z - \nabla_{[v,w]}z, \tag{34.26}$$

where ∇ is the Levi-Civita connection. Locally it is given by

$$R^i_{jkl}e_i = R(e_k, e_l)e_j. \tag{34.27}$$

Property 34.2.2 (Bianchi identity). The first (or **algebraic**) Bianchi identity reads

$$R(X, Y)Z + R(Y, Z)X + R(Z, X)Y = 0. \tag{34.28}$$

The second (or **differential**) Bianchi identity is a similar identity for the covariant derivative:

$$(\nabla_Z R)_{X,Y}W + (\nabla_X R)_{Y,Z}W + (\nabla_Y R)_{Z,X}W = 0. \tag{34.29}$$

Formula 34.2.3 (Directional curvature operator[4]).

$$R_v(w) := R(w, v)v \tag{34.30}$$

[4]Also called the **tidal force operator** (mostly in physics).

Formula 34.2.4 (Sectional curvature).

$$\sec(v,w) := \frac{g(R(w,v)v,w)}{g(v,v)g(w,w) - g(v,w)^2} = \frac{g(R_v(w),w)}{g(v\wedge w, v\wedge w)} \tag{34.31}$$

An important result states that the sectional curvature only depends on the span of v, w.

Remark 34.2.5. For surfaces the sectional curvature coincides with the Gaussian curvature K by Gauss's Theorema Egregium 28.2.51. In general, the sectional curvature gives the Gaussian curvature of the plane spanned by the vectors v, w.

Formula 34.2.6 (Ricci tensor). In coordinate-free notation the Ricci tensor is defined as the trace of the Riemann tensor:

$$\mathrm{Ric}(v,w) := \mathrm{tr}\left(x \mapsto R(x,v)w\right) \tag{34.32}$$

Equivalently

$$\mathrm{Ric}(v,w) = \sum_{i=1}^{n} g(R(e_i,v)w, e_i). \tag{34.33}$$

Locally this becomes:

$$R_{\mu\nu} \equiv \mathrm{Ric}_{\mu\nu} := R^\lambda{}_{\mu\nu\lambda}. \tag{34.34}$$

Property 34.2.7. The Ricci tensor can also be rewritten in terms of the sectional curvature whenever $\|v\| = 1$. Let $\{e_1, \ldots, e_{n-1}\}$ be a set of orthonormal vectors such that $\{e_1, \ldots, e_{n-1}, v\}$ forms an orthonormal basis.

$$\mathrm{Ric}(v,v) = \sum_{i=1}^{n-1} \sec(v, e_i). \tag{34.35}$$

It follows that the Ricci tensor can be interpreted as an averaged (sectional) curvature.

Formula 34.2.8 (Ricci scalar).

$$R := R^\mu{}_\mu \tag{34.36}$$

This (scalar) quantity is also called the **scalar curvature**.

Formula 34.2.9 (Einstein tensor).

$$G_{\mu\nu} := R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R \tag{34.37}$$

Property 34.2.10. For 4-manifolds the Einstein tensor $G_{\mu\nu}$ is the only tensor containing at most second derivatives of the metric that satisfies

$$\nabla_\mu G^{\mu\nu} = 0. \tag{34.38}$$

Definition 34.2.11 (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 (66.2) (with a cosmological constant).

34.3 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 [10].

34.3.1 Spin structures

From Definition 20.4.28 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. This section was not placed in the chapter on Clifford algebras because the natural setting for spinors is Riemannian geometry.

Definition 34.3.1 (Spinor). Let V be a vector space equipped with a metric g of signature (p, q). Consider the set $\mathbb{C}^{2^k} \times F_{\mathrm{SO}}V \times \mathrm{Spin}(p, q)$, where $k = \lfloor \frac{p+q}{2} \rfloor$ and $F_{\mathrm{SO}}V$ 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

$$c_2 = \Lambda c_1 \qquad \mathfrak{b}_1 = L\mathfrak{b}_2 \qquad \Lambda = \Lambda_2\Lambda_1^{-1} \qquad \rho(\Lambda) = L, \tag{34.39}$$

where ρ is the 2-to-1 covering map $\mathrm{Spin}(p, q) \to \mathrm{SO}(p, q)$. 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 $\mathrm{Spin}(p, q)$. 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 34.3.2 (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 $F_{\mathrm{SO}}M$ 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 $F_{\mathrm{SO}}M$ 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:

$$\rho(\Lambda_i(x)\Lambda_j^{-1}(x)) = L_{ij}(x) \tag{34.40}$$

whenever

$$\mathfrak{b}_i(x) = L_{ij}(x)\mathfrak{b}_j(x) \tag{34.41}$$

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 34.3.3 (Spin structure). Consider the orthonormal frame bundle

$$\pi_{\mathrm{SO}} : F_{\mathrm{SO}}M \to M$$

that is obtained by reducing the structure group of the frame bundle FM from $\mathrm{GL}(n)$ to $\mathrm{SO}(n)$.

The smooth manifold M is said to have a spin structure (cf. Definition 33.5.2) if there exists a principal Spin(n)-bundle $\pi_{\text{Spin}} : P_{\text{Spin}} \to M$ and an equivariant 2-fold lifting of F_{SO} to P_{Spin}, i.e. a morphism $\xi : P_{\text{Spin}} \to F_{\text{SO}}M$ together with the 2-fold covering map $\rho : \text{Spin}(n) \to \text{SO}(n)$ that satisfy

- $\pi_{\text{SO}} \circ \xi = \pi_{\text{Spin}}$, and
- $\xi(p \lhd g) = \xi(p) \cdot \rho(g)$

for all $g \in \text{Spin}(n)$, 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 Spin(n)-bundle P_{Spin} is called the **spin frame bundle** (of M).

Definition 34.3.4 (Spinor bundle). The vector bundle associated to the spin frame bundle induced by the (fundamental) spinor representation 25.4.15. 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**.

Example 34.3.5. A 3-manifold is spin if it is compact and orientable.

Example 34.3.6. Any (stably) parallelizable manifold is spin.

34.3.2 Stiefel-Whitney classes

Definition 34.3.7 (Stiefel-Whitney classes). Consider a rank-n vector bundle E over a paracompact space M (hence admitting an 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 product formula**: $w_k(E \oplus E') = \bigoplus_{i=1}^k w_i(E) \smile w_{k-i}(E')$ 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 34.3.8 (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 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 $g_{ij} : U_i \cap U_j \to \text{O}(n)$. Using the two-fold covering map $\rho : \text{Pin}(n) \to \text{O}(n)$, this cocycle can be used to define a function

$$p_{ijk} : U_i \cap U_j \cap U_k \to \text{Spin}(n) : x \mapsto \gamma_{jk}(x)\gamma_{ik}^{-1}(x)\gamma_{ij}(x), \tag{34.42}$$

where γ_{ij} is a lift of the cocycle g_{ij} to 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 33.5.1.

Property 34.3.9 (Stiefel-Whitney classes). For a Pin$^+(n)$-extension of an O(n)-bundle, the cocycle p satisfies $[p] = w_2$. However, for a Pin$^-(n)$-extension of an 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 34.3.10. Consider the orientation line bundle 32.7.4 of an $\mathrm{O}(n)$-bundle. The cocycles 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 34.3.11 (Orientability). A smooth manifold is orientable if and only if its first Stiefel-Whitney class vanishes.

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

Property 34.3.12 (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 34.3.13 (Integral classes). Consider the Bockstein morphism $\beta : H^i(M;\mathbb{Z}_2) \rightarrow H^{i+1}(M;\mathbb{Z})$ associated to the mod-2 reduction $\mathbb{Z} \rightarrow \mathbb{Z}_2$. The integral Stiefel-Whitney classes are defined as the images $\beta w_i \in H^{i+1}(M;\mathbb{Z})$.

Property 34.3.14 (Spin$^{\mathbb{C}}$-structures). An orientable vector bundle admits a Spin$^{\mathbb{C}}$-structure if and only if its third integral Stiefel-Whitney class vanishes.

34.3.3 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 34.3.15 (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 34.3.16 (Plymen's criterion). A Riemannian manifold admits a 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. (Note that Morita-equivalence of C^*-algebras is more involved than the similar notion for ordinary rings.)

The spinor bundle 34.3.4 is a particular instance of the following notion:

Definition 34.3.17 (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

$$\partial := \sum_{i=1}^n e_i\partial_i \tag{34.43}$$

[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*.

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 $C\ell_n(\mathbb{R})$:

$$\underline{\partial} := \sum_{i=1}^{n} \gamma_i \partial_i. \tag{34.44}$$

To extend this definition to Clifford modules one simply needs to replace partial derivatives by covariant derivatives as usual:

Property 34.3.18 (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:

$$\nabla^E(\iota_{C\ell}(X) \cdot \sigma) = \iota_{C\ell}(\nabla X) \cdot \sigma + \iota_{C\ell}(X) \cdot \nabla^E \sigma, \tag{34.45}$$

where $\iota_{C\ell} : TM \to C\ell(TM)$ is the canonical map that embeds a vector field in $C\ell(TM)$.

Definition 34.3.19 (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:

$$\underline{D} := \sum_{i=1}^{n} \gamma_i \cdot \nabla^E_{e_i} \sigma, \tag{34.46}$$

where $\{e_i\}_{i \leq n}$ is a local (orthonormal) frame field.

Property 34.3.20 (Ellipticity). The Dirac operator is a self-adjoint elliptic differential operator 32.9.1.

Definition 34.3.21 (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

$$S \cong S^+ \oplus S^- \tag{34.47}$$

under Clifford multiplication, which in turn induces a decomposition of E. The Dirac operator $\underline{D}$ interchanges these spaces:

$$\underline{D}S^+ = S^- \qquad\qquad \underline{D}S^- = S^+. \tag{34.48}$$

By the previous property, $\underline{D}$ is elliptic and, if M is compact, one can show that this implies that it is also Fredholm 23.4.37. The index of $\underline{D}$ is defined as the Fredholm index of $\underline{D}|_{E^+}$:

$$\mathrm{ind}(\underline{D}) := \dim\ker(\underline{D}|_{E^+}) - \dim\mathrm{coker}(\underline{D}|_{E^+}). \tag{34.49}$$

The symbol $\sigma(D) : \pi^*E \to \pi^*F$, where $\pi : T^*M \to M$ is the cotangent bundle projection, of every elliptic differential operator gives 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 39.2.45 to obtain a (compactly supported[6]) cohomology class on T^*M and apply the Thom isomorphism 32.8.24 to obtain a (compactly supported) cohomology class on 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

[6]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 32.8.13 of T^*M.

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:

$$\mathrm{ch}(\mathcal{T}(1)) \equiv \mathrm{ch}(\mathrm{th}(E)) = \mathcal{T}(\mathrm{td}(E)), \tag{34.50}$$

where 1 denotes the multiplicative unit in K-theory and td is the Todd class 33.6.16. This also implies that

$$\mathrm{ch}(\mathcal{T}_K(x)) = \mathcal{T}(\mathrm{ch}(x) \cup \mathrm{td}(E)), \tag{34.51}$$

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 34.3.22 (Topological index). Consider a compact n-dimensional manifold M together with an elliptic differential operator D.

$$\begin{aligned}\mathrm{ind}(D) :=&(-1)^{n(n+1)/2}\int_M \mathcal{T}^{-1}(\mathrm{ch}(\sigma(D))) \cup \mathrm{td}(T_{\mathbb{C}}M)\\ =&(-1)^n\int_{TM} \mathrm{ch}(\sigma(D)) \cup \mathrm{td}(T_{\mathbb{C}}M),\end{aligned} \tag{34.52}$$

where $\mathcal{T}$ is the Thom isomorphism on TM.

Corollary 34.3.23. The index of an elliptic differential operator on an odd-dimensional manifold vanishes.

Property 34.3.24. The $\hat{A}$-genus 33.6.17 of a spin manifold is an integer.

Formula 34.3.25 (Lichnerowicz formula). Let $\underline{D}$ be the Dirac operator on a spinor bundle S. Unlike in the Euclidean case, the Dirac operator does not square to the (Bochner) Laplacian:

$$\underline{D}^2 = \nabla^*\nabla + \frac{1}{4}R, \tag{34.53}$$

where R is the scalar (Ricci) curvature.

?? COMPLETE ??

34.4 Conformal structures

Definition 34.4.1 (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 28.2.12):

$$f^*g' = \Omega g \tag{34.54}$$

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 34.4.2 (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

$$\mathcal{L}_X g = \kappa g. \tag{34.55}$$

In local coordinates this amounts to

$$\nabla_\mu X_\nu + \nabla_\nu X_\mu = \kappa g_{\mu\nu}, \tag{34.56}$$

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 34.4.3 (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 34.4.4 (Conformal group). The conformal group $\mathrm{Conf}(M)$ is the connected component of the identity of the conformal diffeomorphism group of the conformal compactification of M.

Property 34.4.5. The conformal group of a (pseudo-)Euclidean space of signature (p, q) is isomorphic to $\mathrm{SO}(p + 1, q + 1)$.

Chapter 35

Symplectic Geometry

References for this chapter are [39, 52].

35.1 Symplectic manifolds

Definition 35.1.1 (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**: $d\omega = 0$, and
2. **Nondegeneracy**: $\iota_X\omega = 0 \implies X = 0$.

If the closedness condition is dropped, one obtains an **almost symplectic form**.

Definition 35.1.2 (Symplectic manifold). A manifold equipped with a symplectic form.

Property 35.1.3 (Dimension). From the antisymmetry and the nondegeneracy of the symplectic form it follows that the manifold is necessarily even-dimensional.

Theorem 35.1.4 (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*

$$\omega|_\Omega = \sum_i dp^i \wedge dq^i. \tag{35.1}$$

The adapted charts in this theorem are called **Darboux charts**. Any symplectic manifold admits a covering by Darboux charts.

Remark 35.1.5. 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.

Formula 35.1.6. In Darboux coordinates the components of the symplectic form ω are given by

$$\omega = \begin{pmatrix} 0 & -\mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix}. \tag{35.2}$$

Using the nondegeneracy condition one can define the "dual" or inverse $\omega^\sharp$ as

$$\omega^\sharp = \begin{pmatrix} 0 & \mathbb{1} \\ -\mathbb{1} & 0 \end{pmatrix}. \tag{35.3}$$

Note that the literature is very divided on what convention to use. Some authors use the opposite of the above convention.

Property 35.1.7. In the language of G-structures (Section 33.5) one can restate the definition of symplectic manifolds. A smooth $2n$-dimensional manifold is almost symplectic exactly if it admits an $\mathrm{Sp}(2n)$-structure. It is symplectic exactly if the $\mathrm{Sp}(2n)$-structure is integrable, which by Darboux's theorem is equivalent to first-order integrability, i.e. $d\omega = 0$.

Definition 35.1.8 (Symplectic potential). By the Poincaré lemma 32.8.8 the symplectic form ω locally defines a one-form:

$$\omega = d\theta. \tag{35.4}$$

This one-form is sometimes called the sympletic potential of (M, ω).

Construction 35.1.9 (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

$$\alpha := p_i dq^i.$$

In coordinate-free notation this can be written as follows:

$$\alpha|_y(X) = y(\pi_*X), \tag{35.5}$$

where $X \in T_yT^*M$. This one-form serves as a symplectic potential for the cotangent bundle:

$$\omega = d\alpha.$$

Definition 35.1.10 (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:

$$\mathcal{L}_X\omega = \omega. \tag{35.6}$$

By Cartan's magic formula a global symplectic potential is then given by

$$\theta := X \lrcorner\, \omega. \tag{35.7}$$

It follows that the existence of a Liouville vector field implies the exactness of M.

Definition 35.1.11 (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**: $d\omega = 0$, and
2. **Nondegeneracy**: $\iota_X\omega = 0 \implies X = 0$.

Construction 35.1.12 (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 Equation (35.5):

$$\alpha|_y(X_1, \ldots, X_n) := y\big(\pi_*X_1, \ldots, \pi_*X_n\big), \tag{35.8}$$

where $X_1, \ldots, X_n \in T_y\Lambda^nT^*M$.

Example 35.1.13 (Killing form). The transgression 3-cocycle 30.6.11 induced by the Killing form of a compact simple Lie group turns this group into a 2-plectic manifold.

Definition 35.1.14 (Symplectomorphism). A symplectomorphism is an isomorphism of symplectic manifolds, i.e. a diffeomorphism $f : (M, \omega_M) \to (N, \omega_N)$ satisfying

$$f^*\omega_N = \omega_M. \tag{35.9}$$

These functions form an infinite-dimensional Lie group called the **symplectomorphism group**. This should not be confused with the symplectic group $\mathrm{Sp}(n)$.

Definition 35.1.15 (Symplectic vector field). A vector field is said to be symplectic if its flow preserves the symplectic form ω:

$$\mathcal{L}_X\omega = 0. \tag{35.10}$$

Equivalently, a vector field is symplectic if its flow is a symplectomorphism. These vector fields form a Lie subalgebra of $\mathfrak{X}(M)$.

Theorem 35.1.16 (Gromov's nonsqueezing theorem[1]). *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 invariants:

Definition 35.1.17 (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:

1. **Homogeneity**: If $\alpha \neq 0$, $C(M, \alpha\omega) = |\alpha| C(M, \omega)$.
2. **Normalization**: $C(B^{2n}(r)) = C(Z^{2n}(r)) = \pi r^2$.

Example 35.1.18. 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.

35.2 Hamiltonian vector fields

Definition 35.2.1 (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[2]:

$$\omega(X_f, \cdot) = -df(\cdot). \tag{35.11}$$

This can be rewritten using $\omega^\sharp$ as

$$X_f(\cdot) = \omega^\sharp(-df, \cdot). \tag{35.12}$$

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**.[3]

[1] Also called the theorem of the **symplectic camel** (after a paper by *I. Stewart*).

[2] A lot of different conventions exist in the literature. Here the one that is compatible with the Hamiltonian equations 48.3.3 (which are universally accepted) is used.

[3] The fact that the Hamiltonian flow indeed preserves the symplectic form follows from the closedness of ω.

Definition 35.2.2 (Poisson bracket). Let (M, ω) be a symplectic manifold. The Poisson bracket of two functions $f, g \in C^\infty(M)$ is defined as

$$\{f, g\} := X_f(g) \tag{35.13}$$

or, equivalently, as

$$X_{\{f,g\}} := [X_f, X_g], \tag{35.14}$$

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 35.2.3. 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[4]) 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 the structure becomes a Poisson algebra 30.2.3.

Definition 35.2.4 (Poisson manifold). A smooth manifold on which the algebra of smooth functions can be equipped with a Poisson algebra structure. This is equivalently encoded in a bivector field $\Pi \in \Gamma(\Lambda^2 TM)$ such that

$$\{f, g\} = df \wedge dg(\Pi). \tag{35.15}$$

Property 35.2.5. Every symplectic manifold is a Poisson manifold. The converse, however, is not true.

By generalizing this structure even more, one obtains Jacobi manifolds:

Definition 35.2.6 (Jacobi manifold). A smooth manifold on which the algebra of smooth functions can be equipped with a Lie algebra structure $\{\cdot, \cdot\}$ such that

$$\mathrm{supp}(\{f, g\}) \subseteq \mathrm{supp}(f) \cap \mathrm{supp}(g). \tag{35.16}$$

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

$$\{f, g\} = (f dg - g df)(v) + df \wedge dg(\Pi). \tag{35.17}$$

Remark 35.2.7 (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 a $(n-1)$-form ζ there might not exist a vector field X_ζ that satisfies Equation (35.11). If one restricts to the subspace $\mathrm{Ham}(M)$ of the **Hamiltonian forms** that do induce a Hamiltonian vector field, one can define a generalized Poisson structure as follows:

$$\{\zeta, \xi\} := \mathcal{L}_{X_\xi} \zeta. \tag{35.18}$$

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 definition 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 [86] this will be called the **hemi-bracket**.

[4]The kernel is given by the constant functions and, hence, it is not a bijection.

For $n = 1$ one can equivalently define the Poisson (semi)bracket as

$$\{f, g\}_s := \iota_{X_g}\iota_{X_f}\omega. \tag{35.19}$$

For $n > 1$, however, this structure differs from the hemi-bracket by an exact form:

$$\{\zeta, \xi\} = \{\zeta, \xi\}_s + d\iota_{X_\zeta}\xi. \tag{35.20}$$

The **semi-bracket** $\{\cdot,\cdot\}_s$ 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 27.7.6 where $L_1 := C^\infty(M)$ and $L_0 := \mathrm{Ham}(M)$. They are isomorphic as Lie 2-algebras.

Remark 35.2.8 (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 35.2.3. 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 $X = X_h$.

35.3 Lagrangian submanifolds

Definition 35.3.1 (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 $T_p^\omega S$) at the point $p \in S$ is defined as the subspace

$$T_p^\perp S := \{v \in T_pM \mid \omega(v, \iota_* w) = 0, \forall w \in T_pS\}. \tag{35.21}$$

Definition 35.3.2 (Isotropic submanifold). Let (M, ω) be a symplectic manifold. An embedded submanifold $\iota : S \hookrightarrow M$ is said to be isotropic if $\iota_* T_pS \subset T_p^\perp S$ or, equivalently, if $\omega|_S \equiv 0$. It is said to be coisotropic if $T_p^\perp S \subset \iota_* T_pS$.

Property 35.3.3 (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 35.3.4 (Lagrangian submanifold). Let (M, ω) be a symplectic manifold. An embedded submanifold $\iota : S \hookrightarrow M$ is said to be Lagrangian if $\iota_* T_pS = T_p^\perp S$. 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 35.3.5 (Closed sections). Consider a closed section of a cotangent bundle T^*M, i.e. a map $\sigma : M \to T^*M$ such that $d\sigma = 0$. The graph of σ is a Lagrangian submanifold.

Example 35.3.6 (Conormal bundle). Consider a submanifold $\iota : S \hookrightarrow M$ of a symplectic manifold. The conormal bundle

$$N^*S := \{(p, \alpha) \mid \forall v \in T_pS : \alpha(v) = 0\} \tag{35.22}$$

is a Lagrangian submanifold of T^*M.

Theorem 35.3.7 (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*

$$\frac{\partial W}{\partial q} : M \times \mathbb{R}^k \to \mathbb{R}^k,$$

the subset $\Lambda \subset T^*M$, locally defined by the equations

$$\frac{\partial W}{\partial u^i} = 0 \qquad\qquad p_i = \frac{\partial W}{\partial q^i}, \tag{35.23}$$

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

$$k_0 \geq \dim(M) - \mathrm{rk}(D(\pi \circ \iota)|_{\lambda_0})$$

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 35.3.8 (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 $[\theta] \in H^1_{\mathrm{dR}}(L)$.[5] For M a cotangent bundle and θ the Liouville one-form, this is called the Liouville class of L.

Definition 35.3.9 (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's theorem. In fact, this implies that TM is locally spanned by Hamiltonian vector fields.

More generally one can define a (complex) polarization:

Definition 35.3.10 (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 35.3.11. 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 35.3.12 (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

$$\ker(\pi_*) = T_\alpha T^*_{\pi(\alpha)}M \otimes \mathbb{C} = \mathrm{span}_{\mathbb{C}}\left\{ \frac{\partial}{\partial p_i} \,\middle|\, p_i \text{ is a Darboux coordinate on } T^*M \right\},$$

where $\pi : T^*M^{\mathbb{C}} \to M$ is the (complexified) cotangent bundle projection. It can be shown that this polarization is real.

[5] This class depends on the choice of potential.

Definition 35.3.13 (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 $E := (\mathcal{P} + \overline{\mathcal{P}}) \cap TM$ which are each others 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 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 35.3.14 (Kähler polarization). A polarization $\mathcal{P}$ such that $\mathcal{P} \cap \overline{\mathcal{P}} = \emptyset$. In this case $\mathcal{P}$ is always admissible and $E = TM$. Every *Kähler manifold* (see Chapter 37) 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 bundle is given by the polarization. More generally, the (anti)holomorphic tangent bundles of any complex bundle are called the (anti)holomorphic polarization.

Definition 35.3.15 (Lagrangian Grassmannian). The set of isomorphism classes of Lagrangian subspaces of $\mathbb{R}^{2n}$. It is often denoted by $\mathrm{LGr}(n)$. It can be shown to be isomorphic to the homogeneous space $\mathrm{U}(n)/\mathrm{O}(n)$.

Property 35.3.16 (Maslov class). The square of the determinant function gives an isomorphism $\pi_1(\mathrm{LGr}(n)) \cong \pi_1(S^1)$, 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 $H^1(\mathrm{LGr}(n)) \cong \mathbb{Z}$, 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 Maslov index (for loops) of a Lagrangian submanifold $L \subset M$ is defined analogously. Every loop in L defines a loop in $\mathrm{LGr}(n)$, 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 $L_0 \in \mathrm{LGr}(n)$, one can stratify the Grassmannian by the sets

$$\Lambda_k := \{L \in \mathrm{LGr}(n) \mid \dim(L \cap L_0) = k\}. \tag{35.24}$$

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 to 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 $H_1(\mathrm{LGr}(n), \Lambda_0) \cong H_1(\mathrm{LGr}(n)) \cong \mathbb{Z}$. 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 35.3.17. There exists a generalization [66], 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:

$$\mu + \frac{k_0 - k_1}{2} \in \mathbb{Z}. \tag{35.25}$$

Property 35.3.18 (Homotopy invariance). The Maslov index is invariant under homotopies rel endpoints.

Property 35.3.19 (Additivity). If γ_1 and γ_2 are two paths that can be concatenated, then

$$\mu(\gamma_2 * \gamma_1) = \mu(\gamma_1) + \mu(\gamma_2). \tag{35.26}$$

Alternative Definition 35.3.20 (Maslov index). The above two properties give rise to an axiomatic characterization of the Maslov index:[6]

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**: $\mu(t \mapsto e^{2\pi it}) = 1$.

Formula 35.3.21. Consider a loop $\gamma : [0,1] \to \mathrm{LGr}(n)$. An explicit formula for its Maslov index is given by

$$\mu(L) = \frac{\alpha(1) - \alpha(0)}{\pi}, \tag{35.27}$$

where $\exp(i\alpha(t))$ is the determinant of any unitary lift of γ.

Definition 35.3.22 (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:

$$\mu(\gamma_1, \gamma_2) := \mu(\gamma_1 \times \gamma_2; \Delta), \tag{35.28}$$

where the right-hand side denotes the Maslov index with respect to the Δ-stratification of $\mathrm{LGr}(2n)$. 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, one should choose a *compatible almost complex structure* J (Chapter 37). Then, the Maslov index is defined as follows

$$\mu(\gamma_1, \gamma_2) := \mu(\gamma_1, e^{-\theta J}\gamma_2), \tag{35.29}$$

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.

35.4 Hamiltonian dynamics

35.4.1 Dynamical systems

Definition 35.4.1 (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[7]

$$\dot{F} := \{H, F\}, \tag{35.30}$$

where $\{\cdot,\cdot\}$ is the Poisson bracket on M. The time evolution is completely governed by the Hamiltonian flow $\exp(tX_H)$.

Definition 35.4.2 (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 $\dot{F} \equiv \{H, F\} = 0$.

Property 35.4.3 (Noether's theorem). Every function whose Poisson bracket leaves the Hamiltonian invariant is a conserved quantity:

$$\{H, Q\} = 0 \iff \{Q, H\} = 0. \tag{35.31}$$

[6]Some authors introduce additional factors of 2 in these conditions to account for nonorientablity.
[7]Note that this construction can in fact be generalized to Poisson manifolds.

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 $M := T^*Q \xrightarrow{\pi} Q$ with its canonical symplectic structure. The Hamiltonians that will be considered are of the form (in local Darboux coordinates)

$$H(q,p) = \frac{1}{2}g(p,p) + V(q), \tag{35.32}$$

where V is smooth. These Hamiltonians have two types of symmetries (conserved quantities):

Definition 35.4.4 (Kinematical symmetry). Consider a conserved quantity C. The symmetry is said to be kinematical if $\pi_*(X_C) \in \Gamma(TQ)$ exists and satisfies $\mathcal{L}_{\pi_*(X_C)}g = 0$.

Remark. The second condition says that $\pi_*(X_C)$ is a Killing vector 34.1.24.

Definition 35.4.5 (Dynamical symmetry). Any symmetry that is not a kinematical symmetry.

The following algorithm gives a way to find conditions to see whether a given observable is conserved:

Method 35.4.6 (Van Holten's algorithm). Let the conserved quantity be analytic, i.e.

$$C(q,p) = \sum_{k=0}^{N} \frac{1}{k!} a^{(n_1\ldots n_k)}(q) p_{n_1}\ldots p_{n_k}$$

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

$$\sum_{n=1}^{N}\left[\frac{1}{(k-1)!}a^{n_1\ldots n_{k-1}i}p_{n_1}\ldots p_{n_{k-1}}\frac{\partial V}{\partial q^i} - \frac{2}{k!}\frac{\partial}{\partial q^i}a^{n_1\ldots n_k}p_{n_1}\cdots p_{n_k}g^{im}p_m\right] = 0.$$

Because two polynomials are equal if and only if their corresponding coefficients are equal, one obtains the following equations:

- 0^{th} order:

$$a^k\frac{\partial V}{\partial q^k} = 0, \tag{35.33}$$

- 1^{st} order:

$$a^{(n_1 i)}\frac{\partial V}{\partial q^i} - 2\frac{\partial a}{\partial q^i}g^{in_1} = 0, \tag{35.34}$$

- $\cdots$, and
- N^{th} order:

$$\frac{1}{N!}a^{(n_1\ldots n_N i)}\frac{\partial V}{\partial q^i} - \frac{2}{(N-1)!}\frac{\partial}{\partial q^i}a^{(n_1\ldots n_{N-1}}g^{i)n_N} = 0, \tag{35.35}$$

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

$$a_{(m_1\ldots m_N)}{}^i\partial_i V - 2N\partial_{(m_N}a_{m_1\ldots m_{N-1})} = 0. \tag{35.36}$$

The upper bound N in the series expansion is determined by the generalized Killing condition (34.25):

$$\partial_{(m_{N+1}}a_{m_1\ldots m_N)} = 0 \implies a_{(m_1\ldots m_{N+1})} = 0. \tag{35.37}$$

Remark 35.4.7. The above algorithm still holds for curved manifolds upon replacing all partial derivatives ∂_i by (Levi-Civita) covariant derivatives ∇_i.

35.4.2 Hamilton-Jacobi equation

Definition 35.4.8 (Hamilton-Jacobi equation). Consider a smooth manifold M such that its cotangent bundle comes equipped with a Hamiltonian function $H : T^*M \to \mathbb{R}$. The Hamilton-Jacobi equation for H is the differential equation for a (smooth) function $S : M \to \mathbb{R}$ of the following form:

$$H \circ dS = 0. \tag{35.38}$$

This can be rewritten as

$$H\left(q, \frac{\partial S}{\partial q}\right) = 0. \tag{35.39}$$

Because of Property 35.3.5 one immediately sees that the solutions of the Hamilton-Jacobi equation define a Lagrangian submanifold of the cotangent bundle. Furthermore, they have the property that they are transversal to the fibres of the projection $\pi : T^*M \to M$. By relaxing this transversality condition one obtains the following more general notion:

Definition 35.4.9 (Geometric solution). A Lagrangian submanifold of the level set $H^{-1}(0)$ of a smooth function $H : T^*M \to \mathbb{R}$.

Remark 35.4.10. It can be proven that geometric solutions can locally be described by a solution of the Hamilton-Jacobi equation.

35.4.3 Integrability

Definition 35.4.11 (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 $i, j \leq n : \{F_i, F_j\} = 0$.

Property 35.4.12. Because the Poisson brackets 35.2.2 are related to the commutator of Hamiltonian vector fields, a CIS gives rise to a maximal set of mutually commuting vector fields. Frobenius's theorem 32.5.5 then says that a CIS also gives rise to a maximally integrable distribution and, hence, an n-dimensional regular *foliation*.

35.5 Symplectic reduction

35.5.1 Hamiltonian actions

Alternative Definition 35.5.1 (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 $G = T^n$, 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 the concept of a moment map:

Definition 35.5.2 (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
$$\mu^\xi : M \to \mathbb{R} : p \mapsto \langle \mu(p), \xi \rangle \tag{35.40}$$
is the Hamiltonian function for the fundamental vector field 33.3.4 associated to ξ, i.e. the vector field X_ξ generated by one-parameter group $e^{t\xi}$.

2. μ is an intertwiner between the G-action on M and the coadjoint representation 30.3.10 on $\mathfrak{g}^*$.

2*. Equivalently, this says that 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 35.5.3. 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 is Hamiltonian.

Property 35.5.4 (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 30.6). Let $H : \mathfrak{g} \to C^\infty(M) : \xi \mapsto H_\xi$ be a map assigning Hamiltonian functions to infinitesimal generators. The obstruction
$$\tau(\xi, \zeta) := \{H_\xi, H_\zeta\} - H_{[\xi,\zeta]} \tag{35.41}$$
satisfies the 2-cocycle condition
$$\tau([\xi, \zeta], \theta) + \tau([\zeta, \theta], \xi) + \tau([\theta, \xi], \zeta) = 0, \tag{35.42}$$
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 35.5.5 (Toric manifold). A closed, connected, symplectic $2n$-manifold equipped with a faithful, Hamiltonian torus action of $\mathbb{T}^n$.

Theorem 35.5.6 (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 35.5.7 (Delzant polytope). A polytope $\Delta \subset \mathbb{R}^n$ satisfying the following conditions:

1. **Symplicity**: 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 35.5.8 (Delzant). *The symplectomorphism classes of toric manifolds are in bijection with the isomorphism classes of Delzant polytopes.*

Theorem 35.5.9 (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:*

$$\int_M (f\circ\mu)\frac{\omega^n}{n!} = \int_{\mathfrak{g}^*} fPd\lambda \tag{35.43}$$

for all $f \in L^1(\mathfrak{g}^*)$, where P is piecewise polynomial and λ is the Lebesgue measure on $\mathfrak{g}^*$.

Corollary 35.5.10 (Localization). Consider a circle action on a closed, symplectic n-manifold (M,ω), generated by a Morse function 29.5.1. The Fourier transform of the Liouville measure is a piecewise polynomial function:

$$\int_M e^{-\lambda f}\frac{\omega^n}{n!} = \sum_{p\in\mathrm{Crit}(f)} \frac{e^{-\lambda f(p)}}{\lambda^n e(p)}, \tag{35.44}$$

where $e : M \to \mathbb{Z}$ is the product of the weights obtained by considering the tangent space T_pM as an S^1-representation.

?? COMPLETE ??

35.5.2 Reduction

Property 35.5.11 (Coisotropic reduction). If the leaf space of the characteristic distribution 35.3.3 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 35.5.12 (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 33.1.4. Now, suppose that G acts as a symmetry group on some vector field $X \in \mathfrak{X}(M)$, i.e. $\Phi_g^*X = X$ for all $g \in G$. The reduced vector field $\overline{X} \in \mathfrak{X}(M/G)$ is defined through the following equation:

$$\overline{X}(\pi(m)) := \pi_*X(m), \tag{35.45}$$

where $\pi : M \to M/G$ is the quotient projection.

?? CHECK (cursus Antwerpen) ??

Theorem 35.5.13 (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:*

$$\iota^*\omega = \pi^*\overline{\omega}, \tag{35.46}$$

where $\iota : M_0 \to M$ is the canonical inclusion and $\pi : M_0 \to \overline{M_0}$ the quotient projection.

Remark 35.5.14. 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.*

35.5.3 Poisson reduction

Definition 35.5.15 (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 $f, g \in C^\infty(N)$:

$$\Phi^*[f, g] = \{\Phi^* f, \Phi^* g\}. \tag{35.47}$$

Definition 35.5.16 (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 35.5.17 (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 35.5.18. 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.

35.5.4 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. It is not too hard to show that $T^*G/G \cong \mathfrak{g}^*$.

Formula 35.5.19 (Lie-Poisson equations). First, assign to any vector field $X : Q \to TQ$ a linear function $\mu_X : T^*Q \to \mathbb{R}$ by the following formula:

$$\mu_X(\alpha|_q) := \alpha(X)|_q. \tag{35.48}$$

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:

$$\mu_i \circ \pi : \alpha_k(E^k)_L \mapsto \alpha_i, \tag{35.49}$$

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:

$$\dot{\mu}_i = \{\mu_i, h\}_{\mathfrak{g}^*} = -C^k_{ij}\mu_k \frac{\partial h}{\partial \mu_j}, \tag{35.50}$$

where the Poisson bracket on $\mathfrak{g}^*$ is defined by applying the Poisson reduction theorem to T^*G.

35.6 Metaplectic structures

Definition 35.6.1 (Metaplectic group). Consider the symplectic group $\mathrm{Sp}(2n, \mathbb{R})$ as defined in 20.4.60. This group admits a double covering called the metaplectic group $\mathrm{Mp}(2n, \mathbb{R})$.

Remark 35.6.2. In contrast to the Spin-groups that are the double covers of $\mathrm{SO}(n)$ by Property 25.4.13, the metaplectic groups are not matrix groups, i.e. they do not admit a faithful finite-dimensional representation.

Definition 35.6.3 (Metaplectic structure). Consider a symplectic $2n$-manifold (M, ω). By Property 35.1.7 the frame bundle FM can be reduced to a $\mathrm{Sp}(2n)$-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 $\mathrm{Mp}(2n)$-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 $\rho : \mathrm{Mp}(2n) \to \mathrm{Sp}(2n)$ that satisfies:

- $\pi_{\mathrm{Sp}} \circ \xi = \pi_{\mathrm{Meta}}$, and
- $\xi(p \lhd g) = \xi(p) \cdot \rho(g)$

for all $g \in \mathrm{Mp}(2n)$, where $\lhd$ and $\cdot$ denote the right actions of the respective structure groups.

?? COMPLETE ??

Chapter 36

Contact Geometry

36.1 Contact structures

36.1.1 Contact form

Definition 36.1.1 (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 36.1.2. 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:

$$PT^*M = (T^*M \backslash \{0_M\})/\sim, \tag{36.1}$$

where the equivalence relation $\sim$ is defined by $\omega \sim \rho \iff \exists \lambda \in \mathbb{R}_0 : \omega = \lambda\rho$.

Definition 36.1.3 (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 non-integrability condition[1]:

$$\alpha \wedge (d\alpha)^n \neq 0. \tag{36.2}$$

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 Equation (32.58) explains why contact manifolds are sometimes said to be ‘‘maximally nonintegrable’’.

Property 36.1.4 (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. the line bundle TM/ξ is trivial (or orientable).

36.1.2 Reeb vector fields

Definition 36.1.5 (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 $\iota_X d\alpha = 0$.

Property 36.1.6. Given a contact manifold, there exists a unique Reeb vector field associated to it.

?? COMPLETE ??

[1] In fact it is maximally non-integrable. (Compare with Frobenius's theorem ?? TODO (FORM VERSION) ??.)

Chapter 37

Complex Geometry

37.1 Complex structures

Definition 37.1.1 (Almost complex structure). Let M be a smooth manifold. An almost complex structure on M is a (complexified) smooth $(1,1)$-tensor field $J : TM \rightarrow TM$ such that $J|_p : T_pM \rightarrow T_pM$ satisfies $J|_p^2 = -1$ for all $p \in M$. Such a structure allows to treat the tangent spaces as complex vector spaces by

$$(a+ib)X := aX + bJX. \tag{37.1}$$

A general 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 $E_{\mathbb{R}}$.

This definition implies the following property:

Property 37.1.2. An almost complex manifold is even-dimensional and orientable.

An almost complex structure induces a decomposition of the 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

$$\left\{\frac{\partial}{\partial z^k} := \frac{1}{2}\left(\frac{\partial}{\partial x^{2k-1}} - i\frac{\partial}{\partial x^{2k}}\right)\right\}_{k\leq n}$$

and

$$\left\{\frac{\partial}{\partial \bar{z}^k} := \frac{1}{2}\left(\frac{\partial}{\partial x^{2k-1}} + i\frac{\partial}{\partial x^{2k}}\right)\right\}_{k\leq n},$$

respectively.

Remark 37.1.3. The reason that the almost complex structure is defined on the complexified tangent bundle has to do with the fact that J is only diagonalizable on a complex vector space (because it squares to a negative value).

Example 37.1.4 (Complex vector spaces). Consider a complex vector space V. By looking at Property 21.3.23 and using the canonical isomorphism $V \cong T_vV$ 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 37.1.5 (Reduction of structure group). A $2m$-dimensional manifold M admits an almost complex structure if and only if the structure group of the tangent bundle TM can be reduced from $\mathrm{GL}(\mathbb{R}^{2n})$ to $\mathrm{GL}(\mathbb{C}^n)$.

Moreover, the set of of almost complex structures on $\mathbb{C}^n$ is given by the homogeneous space $\mathrm{GL}(\mathbb{R}^{2n})/\mathrm{GL}(\mathbb{C}^n)$. By globalizing this one obtains that the set of almost complex structures on a complex vector bundle (E, J_0) is given by $\mathrm{Aut}(E_\mathbb{R})/\mathrm{Aut}(E, J_0)$.

Definition 37.1.6 (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.

Definition 37.1.7 (Complex dimension). The integer n in previous definition is called the complex dimension of M. It is denoted by $\dim_\mathbb{C}(M)$.

Property 37.1.8. An almost complex manifold is complex if and only if the $\mathrm{GL}(\mathbb{C}^n)$-structure is integrable.

The integrability condition can be rephrased algebraically as follows:

Theorem 37.1.9 (Newlander-Nirenberg). *An almost complex manifold is complex if and only if the **Nijenhuis tensor** N_J vanishes for all vector fields:*

$$N_J(X, Y) = [JX, JY] - J[JX, Y] - J[X, JY] - [X, Y] = 0. \tag{37.2}$$

In a local coordinate-induced basis this becomes

$$J^\nu_\rho \partial_\nu J^\mu_\sigma - J^\nu_\sigma \partial_\nu J^\mu_\rho - J^\mu_\nu \partial_\rho J^\nu_\sigma + J^\mu_\nu \partial_\sigma J^\mu_\rho = 0. \tag{37.3}$$

Property 37.1.10 (Stiefel-Whitney classes). Let E be a complex vector bundle and denote its real underlying bundle by $E_\mathbb{R}$.

$$c_1(E) \bmod 2 = w_2(E_\mathbb{R}) \tag{37.4}$$

Definition 37.1.11 (Metalinear structure). Consider the complex linear group $\mathrm{GL}(n, \mathbb{C})$ together with the morphism $\det : \mathrm{GL}(n, \mathbb{C}) \to \mathbb{C}^\times$. The metalinear group can be considered as the domain of the holomorphic square root of det:

$$\mathrm{ML}(n, \mathbb{C}) := \left\{(A, z) \in \mathrm{GL}(n, \mathbb{C}) \times \mathbb{C}^\times \,\middle|\, \det(A) = z^2\right\}. \tag{37.5}$$

An equivalent definition, which will be used in the remainder of the text, makes use of the special linear group:

$$\mathrm{ML}(n, \mathbb{C}) = \frac{\mathrm{SL}(n, \mathbb{C}) \times \mathbb{C}}{2\mathbb{Z}}, \tag{37.6}$$

where $\mathbb{Z}$ acts on the product group as $k : (A, z) \mapsto (e^{-2\pi ik/n})A, z + 2\pi ik/n)$. This group is the double cover of $\mathrm{GL}(n, \mathbb{C})$.

Similar to the definition of spinor and metasymplectic structures (Definitions 34.3.3 and 35.6.3), 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 $\mathrm{ML}(n, \mathbb{C}) \to \mathrm{GL}(n, \mathbb{C})$ such that it "commutes" with the bundle map $F_{\mathrm{ML}}M \to FM$.

Property 37.1.12 (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 37.1.13. The above definitions can be restricted to real manifolds and real metalinear structures.

Definition 37.1.14 (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 defined by the action $(g,\lambda) \mapsto z\lambda$, where $g \equiv (A, z) \in \mathrm{ML}(n,\mathbb{C})$.

This can be seen as a (holomorphic) square root of the determinant line bundle. Consider the bundle of 1-densities $|\Omega^1|(M)$ from Definition 32.7.10. 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 $\mu\overline{\nu}$ along the covering map $F_{\mathrm{ML}}M \to FM$. If one does not use the conjugation, a section of the ordinary n-form bundle $\Omega^n(M)$ is obtained. heta

Property 37.1.15 (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.

37.2 Complex differential forms

Property 37.2.1. On a complex manifold there exist coordinates $\{z^\mu\}_{\mu\leq n}$ such that the almost complex structure J can be written as (when extended to $TM^{\mathbb{C}}$)

$$J = i\partial_\mu \otimes dz^\mu - i\partial_{\overline{\mu}} \otimes d\overline{z}^\mu. \tag{37.7}$$

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 37.2.2. Note that on the complexified tangent bundle their exist two kind of imaginary units: i and J. The differential forms dz and $d\overline{z}$ are both linear with respect to the scalar i, but only dz is linear with respect to J, i.e. $dz \circ J = dz$ and $d\overline{z} \circ J = -d\overline{z}$.

Using the basis forms $dz^\mu, d\overline{z}^\mu$ one can also define complex Grassmann spaces $\Omega^{p,q}(M)$, analogous to $\Omega^k(X)$ for smooth manifolds:

$$\Omega^{1,0}(M) := \mathrm{span}_{C^\infty(M,\mathbb{C})}\{dz^\mu\} \tag{37.8}$$

$$\Omega^{0,1}(M) := \mathrm{span}_{C^\infty(M,\mathbb{C})}\{d\overline{z}^\mu\} \tag{37.9}$$

$$\Omega^{p,q}(M) := \left(\bigwedge_{i=1}^{p} \Omega^{1,0}(M)\right) \wedge \left(\bigwedge_{j=1}^{q} \Omega^{0,1}(M)\right). \tag{37.10}$$

Property 37.2.3. 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

$$\Omega^k(M) = \bigoplus_{p+q=k} \Omega^{p,q}(M)$$

of forms of total degree k one can then define the canonical projection maps $\pi^{p,q} : \Omega^k \to \Omega^{p,q}$.

Definition 37.2.4 (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:

$$\partial := \pi^{p+1,q} \circ d \tag{37.11}$$

$$\overline{\partial} := \pi^{p,q+1} \circ d. \tag{37.12}$$

The latter is called the Dolbeault operator. A form is said to be **holomorphic** if it satisfies $\overline{\partial}\omega = 0$, in analogy with the classical Cauchy-Riemann condition (15.7).

Property 37.2.5. By explicitly writing out the action of the de Rham differential d on a general (p,q)-form one obtains the following decomposition:

$$d = \partial + \overline{\partial}. \tag{37.13}$$

Note that for an almost complex manifold this relation in general does not hold. An almost complex manifold is integrable if and only if this expression holds. By using the coboundary property of d one also obtains

$$\partial^2 = \overline{\partial}^2 = 0 \tag{37.14}$$

$$\partial\overline{\partial} + \overline{\partial}\partial = 0. \tag{37.15}$$

Remark 37.2.6 (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.

More generally, a complex vector bundle E is **holomorphic**, i.e. admits a trivialization with holomorphic transition functions, if it admits a Dolbeault operator $\overline{\partial} : \Omega^{\bullet,\bullet}(M;E) \rightarrow \Omega^{\bullet,\bullet+1}(M;E)$ 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.

Theorem 37.2.7 (Koszul-Malgrange). *Let $E \rightarrow M$ be a holomorphic vector bundle. There exists a unique connection ∇ on E such that $\nabla^{0,1} = \overline{\partial}$.*

Formula 37.2.8. Analogous to the definition of the de Rham codifferential (34.3), one can define the adjoints of the Dolbeault operators:

$$\partial^\dagger := -*\partial* \tag{37.16}$$

$$\overline{\partial}^\dagger := -*\overline{\partial}*, \tag{37.17}$$

where the fact that the real dimension of a complex manifold is even is used: $(-1)^{n(k+1)+1} = -1$.

Corollary 37.2.9. Using these definitions one can write the Hodge Laplacian 34.1.5 as:

$$\Delta = 2(\partial\partial^\dagger + \partial^\dagger\partial) = 2(\overline{\partial}\overline{\partial}^\dagger + \overline{\partial}^\dagger\overline{\partial}). \tag{37.18}$$

37.3 Kähler manifolds

In analogy with the definition of Riemannian manifolds 34.1.2 one can also define metrics for complex vector bundles:

Definition 37.3.1 (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 37.3.2 (Kähler manifold). Consider a smooth manifold equipped with a Riemannian structure (M,g), a symplectic structure (M,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[1], and

[1] If not, the manifold is said to be almost Kähler.

2. The symplectic form is **compatible** with the almost complex structure:

$$\omega(v,w) = \omega(Jv,Jw) \tag{37.19}$$

and

$$\omega(v,Jv) > 0; \tag{37.20}$$

or

1. M is Hermitian with metric $h(v,w) := g(v,w) + ig(v,Jw)$, and
2. The fundamental two-form $\omega(v,w) := g(v,Jw)$ is closed and hence symplectic[2];

or[3]

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):

$$\nabla_X J = 0. \tag{37.21}$$

Remark 37.3.3. 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 on a Kähler manifold.

Property 37.3.4 (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-tamed**.

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).

Remark 37.3.5. 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 37.3.6 (Kähler form). When any of the equivalent pairs of conditions in Definition 37.3.2 is satisified, the central object is the Kähler form (also called the **fundamental form**):

$$\omega(v,w) := g(v,Jw). \tag{37.22}$$

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 37.3.7. The metric $g \equiv g_{\mu\nu}dx^\mu \otimes dx^\nu$ can be rewritten as

$$g = g_{\mu\overline{\nu}}\big(dz^\mu \otimes d\overline{z}^\nu + d\overline{z}^\nu \otimes dz^\mu\big). \tag{37.23}$$

The Kähler form can then be written as

$$\omega = ig_{\mu\overline{\nu}}dz^\mu \wedge d\overline{z}^\nu. \tag{37.24}$$

[2]The nondegeneracy condition is automatically satisfied because of the nondegeneracy of the metric.
[3]Here the symplectic structure can be recovered using the Kähler form defined below.

Definition 37.3.8 (Kähler potential). Using the $\partial\overline{\partial}$-lemma 37.5.4 one can locally write the Kähler form as

$$\omega = i\partial\overline{\partial}K(z,\overline{z}), \tag{37.25}$$

where the real function $K \in \Omega^0(M)$ is called the **Kähler potential**.

Corollary 37.3.9. Expression (37.24) implies that one can locally rewrite the metric as

$$g_{\mu\overline{\nu}} = \partial_\mu\partial_{\overline{\nu}}K(z,\overline{z}). \tag{37.26}$$

Property 37.3.10. 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 $\Gamma_{\overline{\mu\nu}}{}^{\overline{\lambda}}$ components do not vanish. They are given by

$$\Gamma_{\mu\nu}{}^{\lambda} = g^{\lambda\overline{\rho}}\partial_\mu g_{\nu\overline{\rho}} \tag{37.27}$$

$$\Gamma_{\overline{\mu\nu}}{}^{\overline{\lambda}} = g^{\overline{\lambda}\rho}\partial_{\overline{\mu}} g_{\overline{\nu}\rho}. \tag{37.28}$$

Accordingly, the only nonvanishing component of the Riemann curvature tensor is

$$R_{\overline{\mu}\nu\overline{\lambda}\rho} = g_{\overline{\lambda}\kappa}\partial_{\overline{\mu}}\Gamma_{\nu\rho}{}^{\kappa}. \tag{37.29}$$

Definition 37.3.11 (Kähler transformation). From Definition 37.3.8 one can conclude that the Kähler potential is not unambiguously defined. The following transformation leaves the Kähler form invariant:

$$K'(z,\overline{z}) = K(z,\overline{z}) + f(z) + \overline{f}(\overline{z}). \tag{37.30}$$

On overlapping coordinate charts the transformation between Kähler potentials is exactly of this form.

Definition 37.3.12 (Hyperkähler manifold). A manifold is said to be hyperkähler if it is *hypercomplex* and if it admits a (Riemannian) metric that is Kähler with respect to all complex structures. Explicitly this means that:

1. there exist distinct complex structures I, J, K such that $I^2 = J^2 = K^2 = IJK = -1$, and
2. the Kähler forms induced by I, J and K are closed.

Definition 37.3.13 (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 $\mathrm{SU}(n)$.[4] This implies for example that the manifold is Ricci flat.

Property 37.3.14 (Calabi-Yau conjecture). For a compact Calabi-Yau manifold, the first Chern class vanishes.

37.3.1 Killing vectors

Definition 37.3.15 (Holomorphic Killing vector). Consider the set of Killing vector fields $\{X_A\}_{A\in I}$ associated to the metric g. Within this set of vector fields one can consider those k_A that satisfy

$$\mathcal{L}_{k_A}J = 0. \tag{37.31}$$

[4]The original definition by *Yau* was that of a compact, Ricci-flat Kähler manifold with vanishing first Chern class.

or, equivalently by the Kähler condition,

$$\mathcal{L}_{k_A}\omega = 0. \tag{37.32}$$

These are called holomorphic Killing vector fields because their components are holomorphic in the sense of complex analysis. This can easily be shown by writing the Killing condition in terms of covariant derivatives and by using Equation (37.7).

Definition 37.3.16 (Moment map). Let k be a holomorphic Killing vector field. From $d\omega = 0$ one can, using Cartan's magic formula 32.4.17 and the above condition, derive that $\iota_k\omega$ is closed. Poincaré's lemma then implies that there exists a real function $\mathcal{P}(z, \overline{z})$ such that

$$\iota_k\omega = d\mathcal{P}. \tag{37.33}$$

Using Equation (37.24) one can then find the following expression for the Killing vector fields:

$$k^\mu = -ig^{\mu\overline{\nu}}\partial_{\overline{\nu}}\mathcal{P}. \tag{37.34}$$

37.3.2 Dirac operators

Property 37.3.17. 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 37.1.10.

Property 37.3.18 (Spin$^{\mathbb{C}}$-structures). By Properties 34.3.14 and 37.1.10, every almost complex manifold admits a Spin$^{\mathbb{C}}$-structure. If there exists a line bundle L, such that $c_1(L) = w_2(M)$ mod 2, then M admits a Spin$^{\mathbb{C}}$-structure with L as its determinant line bundle.

Definition 37.3.19 (Kähler-Dirac operator). Consider a Riemannian manifold (M, g). The Kähler-Dirac operator is a square root of the Hodge Laplacian 34.1.5:

$$D := d + \delta. \tag{37.35}$$

Definition 37.3.20 (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

$$S \cong \Omega^{0,n}(M) \otimes \sqrt{\Omega^{n,0}} \tag{37.36}$$

and that the Dirac operator can be identified with $\overline{\partial} + \overline{\partial}^*$, where $\overline{\partial}$ is the Dolbeault operator 37.2.4. For the action on the θ-characteristic, choose a connection ∇:

$$\overline{\partial}(\omega \otimes \psi) := (\overline{\partial}\omega) \otimes \psi + \sum_{i=1}^{2n} \pi^{0,k+1}(dx^i \wedge \omega) \otimes \nabla_i\psi \tag{37.37}$$

$$\overline{\partial}^*(\omega \otimes \psi) := (\overline{\partial}^*\omega) \otimes \psi - \sum_{i=1}^{2n} (\partial_i \lrcorner\, \omega) \otimes \nabla_i\psi, \tag{37.38}$$

where $\omega \in \Omega^{0,k}(M)$.

37.4 Complex curves

37.4.1 Riemann surfaces and orbifolds

Definition 37.4.1 (Riemann surface). A complex manifold of (complex) dimension one. It is often assumed to be connected.

Property 37.4.2 (Genus classification). Compact Riemann surfaces are, topologically, characterized by their genus.

Example 37.4.3 (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 Möbius group $\mathrm{PSL}(2,\mathbb{C})$, which acts as follows:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} z := \frac{az+b}{cz+d}. \tag{37.39}$$

It can be proven that given any two triples of points, there exists a unique Möbius transformation mapping them onto each other.

Theorem 37.4.4 (Uniformization theorem). *Every simply-connected Riemann surface is biholomorphically equivalent to either the Riemann sphere, the complex plane or the upper half plane.*

Corollary 37.4.5. 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 $S^2, \mathbb{C}$ or $\overline{\mathbb{H}}$ by a freely-acting discrete group.

Moreover, S^2 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 $\overline{\mathbb{H}}$.

Definition 37.4.6 (Family of Riemann surfaces). A family of Riemann surfaces of genus g with n marked points is a function $p:\mathcal{C}\rightarrow U$ that admits n disjoint sections $s_i: U\rightarrow\mathcal{C}$ 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.

Given two such families p, p' and a subset $V\subset U'$, the restriction $p'|_V$ is called a **pullback** of p if there exists a function $\varphi: V\rightarrow U$ such that $\mathcal{C}'|_V \cong \varphi^*\mathcal{C}$.

Definition 37.4.7 (Stable Riemann surface). A Riemann surface Σ of genus g with n marked points is said to be stable if its punctured Euler characteristic

$$\chi(\Sigma\backslash\{z_1,\dots,z_n\}) = 2-2g-n \tag{37.40}$$

is negative.

Property 37.4.8 (Automorphism group). For a stable Riemann surface (Σ, J), every element of $\mathrm{Aut}_n(\Sigma, J)$ 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 is always a smooth Lie group.

Construction 37.4.9 (Moduli space). Denote by $\mathcal{M}_{g,n}$ denote the set of isomorphism classes of Riemann surfaces of genus g with n marked points.

Let Σ be an oriented, closed surface with n marked points. $\mathcal{M}_n(\Sigma)$ denotes the moduli space of almost complex structures up to automorphisms that preserves the order of the marked points:

$$\mathcal{M}_n(\Sigma) := \mathcal{J}(\Sigma)/\mathrm{Diff}_{+,n}(\Sigma). \tag{37.41}$$

In general the action of diffeomorphisms is not free or proper, since $\mathrm{Aut}_n(\Sigma, J)$ fixes J, so the quotient is not necessarily a smooth manifold. By restricting the action to the identity-component and the surfaces to stable surfaces, one obtains the **Teichmüller space** $\mathcal{T}(\Sigma, n)$. $\mathcal{M}_n(\Sigma)$ can then be obtained by further quotienting out the action of the mapping class group 8.1.10.

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 (Chapter 29):

Definition 37.4.10 (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 homomorphism $G' \to G$ 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 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 37.4.11 (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:

$$\varphi_y^{-1} \circ F = f \circ \varphi_x^{-1}. \tag{37.42}$$

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:

$$\int_C \omega := \frac{1}{|G|} \int_{\varphi^{-1}(C)} \omega_U, \tag{37.43}$$

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 37.4.12 (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:

$$\chi(X) := \frac{1}{|G|} \sum_{g \in G} \chi(X^g). \tag{37.44}$$

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".

Property 37.4.13 (Universal curve over $\mathcal{M}_{g,n}$). Let C be Riemann surface of genus g with n marked points (with negative Euler characteristic). Denote its (finite) automorphism group by G. There exist

- an open, bounded, connected set $U \subset \mathbb{C}^{3g-3+n}$,

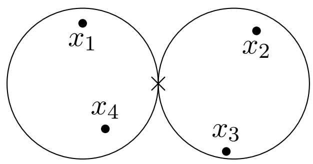
- a family $p : \mathcal{C} \to U$ of Riemann surfaces of genus g with n marked points, and
- a G-action on $\mathcal{C}$ descending to a G-action on U

such that the following conditions are satisfied:

1. The fiber $C_0 := p^{-1}(0)$ is (isomorphic to) 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 with 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 cases of $g = 0, n = 4$. By Example 37.4.3 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 $c \in \mathbb{CP}^1 \backslash \{0, 1, \infty\}$. (This number is also 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 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 37.4.14 (Stable curve). A compact, complex algebraic curve with n marked points that satisfies the following conditions:

1. It has only simple **nodal singularities** (i.e. a singularity of the form $xy = 0$).
2. The marked points are distinct and do not coincide with the nodes.
3. The curve has a finite number of automorphisms.

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 37.4.15. 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 number of automorphisms.

Property 37.4.16 (Deligne-Mumford compactification). There exists a morphism of compact, complex orbifolds $\overline{\pi} : \overline{\mathcal{C}}_{g,n} \rightarrow \overline{\mathcal{M}}_{g,n}$ 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 with 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.

37.4.2 Pseudoholomorphic maps

Definition 37.4.17 (Cauchy-Riemann operator). Consider a complex vector bundle (E, J) and an almost complex manifold (M, j). A Cauchy-Riemann operator on E is a complex-linear map

$$D : \Gamma(E) \rightarrow \Gamma(\overline{\mathrm{Hom}}(TM, E)) \tag{37.45}$$

satisfying the Leibniz rule

$$D(f\sigma) = (\overline{\partial} f)\sigma + f(D\sigma), \tag{37.46}$$

where $\overline{\mathrm{Hom}}$ 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 32.6.1. The following property shows that this is no coincidence:

Property 37.4.18. For every Cauchy-Riemann operator D on a Hermitian vector bundle (E, J) over an almost integrable manifold (M, j) there exists a Hermitian connection ∇ such that

$$D\sigma = \nabla\sigma + J \circ \nabla\sigma \circ j \tag{37.47}$$

for all sections $\sigma \in \Gamma(E)$.

Definition 37.4.19 (Pseudoholomorphic function). Consider two almost complex manifolds (M, J) and (N, j). A function $f : (M, J) \rightarrow (N, j)$ is said to be pseudoholomorphic if it satisfies the **Cauchy-Riemann equations**

$$f_* + j \circ f_* \circ J = 0 \tag{37.48}$$

or, equivalently,

$$j \circ f_* = f_* \circ J, \tag{37.49}$$

i.e. the differential is complex-linear. This can rephrased in terms of the Cauchy-Riemann-like operator

$$\overline{\partial}_J : C^\infty(M, N) \rightarrow \Gamma(\overline{\mathrm{Hom}}(TM, TM)) : f \mapsto f_* + j \circ f_* \circ J, \tag{37.50}$$

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 $\Gamma(\overline{\mathrm{Hom}}(TM, f^*TN))$, since the tangent space at a point $f \in C^\infty(M, N)$ is modelled on $\Gamma(f^*TN)$.[5] By Remark 32.4.23 this operator can also be written in terms of the differential df.

[5] To make this statement precise one needs to endow $C^\infty(M, N)$ with the right topology, since this manifold is infinite-dimensional.

Remark 37.4.20 (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 2D 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 2D manifolds in the domain, because in general there are no nonconstant pseudoholomorphic functions (even locally) when the domain has a higher dimension.

Property 37.4.21 (Regularity). If a function of class C^1 satisfies the Cauchy-Riemann equations, it is automatically smooth.

Definition 37.4.22 (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 37.4.23 (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 37.4.24 (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:

$$\mathrm{E}(f) := \int_\Sigma f^*\omega = \frac{1}{2}\int_\Sigma |df|^2 \mathrm{Vol}_\Sigma. \tag{37.51}$$

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 37.4.25 (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 $(\Sigma, J_0, f, z_1, \ldots, z_n)$, 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 biholomorphic or conformal diffeomorphism) that preserves the order of the marked points.

Example 37.4.26. When $M = \{*\}$, the moduli space reduces to $\mathcal{M}_{g,n}$, the moduli space of Riemann surfaces of genus g with n marked points.

37.5 Cohomology

37.5.1 Dolbeault cohomology

Theorem 37.5.1 (Hodge decomposition). *Let M be a compact Kähler manifold.*

$$H^k_{\mathrm{dR}}(M) \cong \bigoplus_{p+q=k} H^{p,q}(M) \tag{37.52}$$

for all $k \in \mathbb{N}$

By analogy with the Poincaré lemma for smooth manifolds one can prove the following theorems:

Theorem 37.5.2 (∂-lemma). *Let $\alpha \in \Omega^{p,q}(M)$. If $\partial\alpha = 0$, locally there exists a complex form $\beta \in \Omega^{p-1,q}$ such that $\alpha = \partial\beta$.*

Theorem 37.5.3 ($\overline{\partial}$-lemma). *Let $\alpha \in \Omega^{p,q}(M)$. If $\overline{\partial}\alpha = 0$, locally there exists a complex form $\beta \in \Omega^{p,q-1}$ such that $\alpha = \overline{\partial}\beta$.*

Theorem 37.5.4 ($\partial\overline{\partial}$-lemma). *Let $\alpha \in \Omega^{p,q}(M)$. If $d\alpha = 0$, locally there exists a complex form $\beta \in \Omega^{p-1,q-1}$ such that $\alpha = \partial\overline{\partial}\beta$.*

37.5.2 Lagrangian Floer homology ♣

In this section a (co)homology theory is constructed from the intersection theory of Lagrangian submanifolds (see Chapter 35 for an introduction).

Recall the pseudoholomorphic polygons from Section 37.4.2. 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 37.5.5 (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 Figure 37.1. 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.

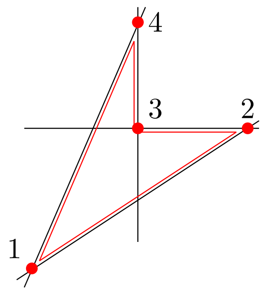


Figure 37.1: Pseudoholomorphic 4-gon.

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.* 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 n^{th} 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 $(f_n)_{n\in\mathbb{N}}$, here 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 Figure 37.1.

- **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 *montone Lagrangian submanifolds* or by requiring that $[\omega]\cdot\pi_2(M,L)=0=[\omega]\cdot\pi_2(M,L')=0$.

Definition 37.5.6 (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:

$$\partial\langle p\rangle := \sum_{q\in L\cap L'\backslash\{p\}} n(p,q)\langle q\rangle, \tag{37.53}$$

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:

$$\mu([f]) := \mathrm{ind}(D_{\overline{\partial}_J}), \tag{37.54}$$

where $D_{\overline{\partial}_j}$ 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 $\mathrm{LGr}(TM)$ 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 37.5.7 (Maslov covering). A **Maslov covering** of M is a $\mathbb{Z}$-covering $\mathcal{L}\to\mathrm{LGr}(TM)$ such that the fibre over every Lagrangian is given by the $\mathbb{Z}$-covering of $\mathrm{LGr}(\mathbb{R}^{\dim(M)})$. This corresponds to patching together the universal covers of the Lagrangian Grassmannians at every point of M.

A Maslov covering exists if and only if the first Chern class is 2-torsion:

$$2c_1(M)=0. \tag{37.55}$$

[6]The domain, i.e. the disk with two boundary punctures, has as automorphism group $\mathbb{R}$. It scales the distance between the punctures.

(This can be restated in cohomological conditions involving the Chern class [67], $c_1(M)$ 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.)

$$\partial\langle p\rangle = \sum_{\substack{q\in L\cap L' \\ \mathrm{ind}([f])=1}} |\mathcal{M}(p,q;[f])| T^{\omega([f])}\langle q\rangle, \tag{37.56}$$

where T is the generator of the *Novikov field* Λ and $\omega([f])$ denotes the symplectic energy of $[f]$.

Property 37.5.8 (Hamiltonian isotopy). Consider two Lagrangian submanifolds L, L'. If there exists a Hamiltonian isotopy $L' \to L''$, then

$$HF(L, L') \cong HF(L, L''). \tag{37.57}$$

If there exists a Hamiltonian isotopy $L \to L'$, then

$$HF(L, L') \cong H^\bullet(L), \tag{37.58}$$

where the left-hand side denotes singular cohomology.

Definition 37.5.9 (Symplectic action functional). For every two Lagrangian submanifolds $L, L' \subset M$, one can consider the space of smooth paths connecting L and L':

$$\mathcal{P}(L, L') := \{\gamma \in C^1([0,1], M) \mid \gamma(0) \in L, \gamma(1) \in L'\}. \tag{37.59}$$

The symplectic action functional is defined as follows:

$$A : \widetilde{\mathcal{P}}(L, L') \to \mathbb{R} : [\gamma, h] \mapsto \int_{[0,1]\times[0,1]} h^*\omega, \tag{37.60}$$

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 trivialisable on h, Stokes's theorem implies that this integral is (up to a constant depending on the choice of base point) equal to

$$\int_{[0,1]} \gamma^*\theta, \tag{37.61}$$

where θ is the symplectic potential.

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:

$$\mathrm{grad} A([\gamma, h]) = J_t \frac{\partial\gamma}{\partial t}. \tag{37.62}$$

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 29.5) 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.

[7] To be correct one should use the so-called *Novikov covering*.

Definition 37.5.10 (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

$$\mu : \mathrm{Hom}(L', L'') \otimes \mathrm{Hom}(L, L') \to \mathrm{Hom}(L, L'') \tag{37.63}$$

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 pseudo-holomorphic 3-gons with boundary on the Lagrangians and marked points q_1, q_2 and q.

Remark 37.5.11 (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 $m_0 \neq 0$.

?? COMPLETE ??

Chapter 38

Calculus of Variations

The standard references for global variational calculus are [65, 74].

38.1 Constrained systems

Definition 38.1.1 (Holonomic constraint). A constraint $f(q,t) = 0$ is said to be holonomic if it only depends on the coordinates q^i and t and not on the derivatives.

Method 38.1.2 (Holonomic constraints). The Euler-Lagrange equations of a system with k holonomic constraints $f_k(q,t) = 0$ can be obtained from the generalized action functional

$$I_{\text{ext}}[q,\lambda] := \int_a^b \left[L(q,\dot{q},t) + \sum_{i=1}^k \lambda_i(t) f_i(q,t)\right] dt, \tag{38.1}$$

where the λ_i are are called **Lagrange multipliers**. Extrimizing with respect to these multipliers induce the constraints:

$$\frac{\delta I_{\text{ext}}}{\delta \lambda_i} = 0. \tag{38.2}$$

38.2 Introduction

Definition 38.2.1 (Variational symmetry). Consider an integral quantity I defined by a Lagrangian function:

$$I_M = \int_M L(q,u,u_I)dq, \tag{38.3}$$

where u are (analytic) functions of the variables q. A transformation $(x,u) \longrightarrow (x',u')$ of the variables[1] is called a variational (or Noether) symmetry if it satisfies

$$\int_{M'} L(q',u',u'_I)dq' = \int_M L(q,u,u_I)dq \tag{38.4}$$

for arbitrary M.

Following *Lie*, the notion of a group of transformations is introduced:

[1] The transformations of the derivatives ∂u are induced by the ones for u.

Definition 38.2.2 (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 is the number of independent parameters.

Remark 38.2.3. It should be clear that this is the same as a finite-dimensional Lie group 30.1.1.

Instead of a parameters, one can also generalize to functions:

Definition 38.2.4 (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 is the number of independent functions.

Remark 38.2.5. In physics terminology the infinite groups would be the symmetry groups obtained by "gauging" a global symmetry $\mathfrak{G}_k$.

Theorem 38.2.6 (Noether). *Consider an integral quantity I that is invariant under a continuous group $\mathfrak{G}$.*

- *If $\mathfrak{G}$ is finite continuous of the form $\mathfrak{G}_k$, there exist k 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 $\mathfrak{G}_k$.*
- *If $\mathfrak{G}$ is infinite continuous of the form $\mathfrak{G}_{\infty,k}$, there exist k independent relations among the Lagrangian expressions and their derivatives[2]. Conversely, if k such relations exist, then I is invariant under a group of the form $\mathfrak{G}_{\infty,k}$.*

Remark 38.2.7. In fact, the first theorem is also valid in the limit of an infinite number of parameters.

Property 38.2.8 (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 38.2.9 (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 38.2.10 (One dimension). Consider the following infinitesimal transformations

$$\begin{aligned} q^i &\longrightarrow q^i + \varepsilon\xi^i(q^k,t)\\ t &\longrightarrow t+\varepsilon\tau(q^k,t)\\ \dot{q}^i &\longrightarrow \dot{q}^i+\varepsilon(\dot{\xi}^i-\dot{q}^i\dot{\tau}), \end{aligned}\tag{38.5}$$

where the transformation of the "velocities" on the last line, induced by the coordinate transformations, is called a **prolongation** (see also Definition 38.4.3 further below). These transformations

[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.

generate Noether symmetries if they leave the Lagrangian invariant up to a total derivative (in first order) for every subinterval $[t_0, t_1] \subseteq [a, b]$ and for some function $f(q, t)$:

$$\int_{\tilde{t}_0}^{\tilde{t}_1} L(\tilde{q}, \dot{\tilde{q}}, \tilde{t}) d\tilde{t} = \int_{t_0}^{t_1} L(q, \dot{q}, t) dt + \varepsilon \int_{t_0}^{t_1} \frac{df}{dt} dt + O(\varepsilon^2). \tag{38.6}$$

This is equivalent to requiring that the transformation is a solution of the following differential equation (sometimes called the **Rund-Trautman identity**):

$$\frac{\partial L}{\partial t}\tau + \frac{\partial L}{\partial q^i}\xi^i + \frac{\partial L}{\partial \dot{q}^i}(\dot{\xi}^i - \dot{q}^i\dot{\tau}) + L\dot{\tau} = \dot{f}. \tag{38.7}$$

By Noether's (first) theorem one obtains for every such symmetry a conserved quantity of the following form

$$F := f - \left[L\tau + \frac{\partial L}{\partial \dot{q}^i}(\xi^i - \dot{q}^i\tau)\right]. \tag{38.8}$$

It is important to note that the left-hand side of Equation (38.7) is simply the Lie derivative of the Lagrangian form Ldt with respect to the vector field that generates the transformations (38.5). (This will turn out to be an important concept in the calculus of variations. See Definition 38.4.39.)

38.3 Jet bundles

38.3.1 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 38.3.1 (Jet). Consider two smooth manifolds M, 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

$$\left.\frac{\partial^\alpha f^i}{\partial x^\alpha}\right|_p = \left.\frac{\partial^\alpha g^i}{\partial x^\alpha}\right|_p \tag{38.9}$$

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 $j_p^r f$. The number 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 38.3.2 (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:

$$\pi_r : J^r(M, N) \to M : j_p^r f \mapsto p \tag{38.10}$$

$$\pi_{r,0} : J^r(M, N) \to N : j_p^r f \mapsto f(p). \tag{38.11}$$

One can also define a **k-jet projection** $\pi_{r,k}$ as the map

$$\pi_{r,k} : J^r(M, N) \to J^k(M, N) : j_p^r f \mapsto j_p^k f, \tag{38.12}$$

where $k \leq r$. The k-jet projections satisfy the transitivity property $j_{k,m} = j_{r,m} \circ j_{k,r}$.

Definition 38.3.3 (Prolongation). Let $f : M \to N$ be a smooth function. The r-jet prolongation $j^r f$ is defined as the following map:

$$j^r f : M \to J^r(M, N) : p \mapsto j^r_p f. \tag{38.13}$$

Property 38.3.4 (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 $J^r(\varphi^{-1}, \psi) := J^r(\varphi^{-1}, V) \circ J^r(\varphi(U), \psi) : J^r(U, V) \to J^r(\varphi(U), \psi(V))$. 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 38.3.5 (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

$$S^k(U) := \{f \in C^\infty(M, N) \mid j^k f \in U\}, \tag{38.14}$$

where U is open in $J^k(M, N)$.

Property 38.3.6. 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 38.3.7 (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

$$j^r_{\sigma(p)}\pi \circ j^r_p\sigma = j^r_p \mathbb{1}_M. \tag{38.15}$$

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:

$$U^r_i := \{j^r_p\sigma \mid \sigma(p) \in U_i\} \tag{38.16}$$

$$\varphi^r_i := \left(x^k, u^\alpha, \left.\frac{\partial^I u^\alpha}{\partial x^I}\right|_p\right), \tag{38.17}$$

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 38.3.8 (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 38.3.9 (Infinite jet bundle). The inverse limit 3.7.4 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.

38.3.2 Contact structure

Every jet manifold J^k carries a natural constact structure 36.1.3:

Definition 38.3.10 (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 this in turn defines a distribution, called the **Cartan distribution**. For finite-order jet spaces this distribution is completely nonintegrable, but 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:

$$\theta_I^\alpha := \mathbf{d}u_I^\alpha - \sum_\mu u_{I,\mu}^\alpha \mathbf{d}x^\mu, \tag{38.18}$$

where the x^μ and u_I^α are the independent and dependent variables, respectively.

Property 38.3.11 (Cartan connection). There exists a connection on the infinity-jet bundle $J^\infty(E)$ where the horizontal subbundle is exactly given by the Cartan distribution. Furthermore, this connection is flat.

38.3.3 Differential operators

Alternative Definition 38.3.12 (Differential operator). Let E_1, E_2 be two 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 if it factors through a finite 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**.

Property 38.3.13 (Formal adjoints). Consider two differential operators $D, D^\dagger : \Gamma(E) \to \Gamma(E^* \otimes \Lambda^{\dim(M)}(M))$. 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:

$$\langle D(\sigma_1), \sigma_2\rangle - \langle \sigma_1, D^\dagger(\sigma_2)\rangle = dK(\sigma_1, \sigma_2). \tag{38.19}$$

This formula can be interpreted as a generalization of Green's identities. In the case where M is compact, Stokes's theorem 32.7.23 shows that D and $D^\dagger$ are related through integration by parts.

Definition 38.3.14 (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 $X(\sigma) \in T_{\pi_k(\sigma)}M$.

38.4 Variational bicomplex ♣

In this section the language of jet bundles as introduced in the previous section wil be heavily used to rephrase the classical theory of variations 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.

38.4.1 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:

$$\mathbf{d} = d + \delta.$$

The **horizontal derivative** d lifts the de Rham differential on M to E (hence the name). On smooth functions $C^\infty(J^\infty(E))$ it acts as follows:

$$df := D_\mu f dx^\mu, \tag{38.20}$$

where the **total derivative**[4]

$$D_\mu f := \frac{\partial f}{\partial x^\mu} + \frac{\partial f}{\partial u^\alpha} u^\alpha_{,\mu} + \frac{\partial f}{\partial u^\alpha_{,\nu}} u^\alpha_{,\nu\mu} + \cdots \tag{38.21}$$

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

$$d \circ \delta = -\delta \circ d \tag{38.22}$$

which follows from the nilpotency of all differentials. The differentials d, δ turn $\Omega^\bullet(J^\infty(E))$ into a bigraded complex called the **variational bicomplex**.

Remark 38.4.1. Some authors use the term variational bicomplex for the bicomplex of **local forms** $\Omega^\bullet_{\text{loc}}(M \times \Gamma(E))$, 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 32.7.10).

Definition 38.4.2 (Local Lagrangian). A top-degree horizontal form on $J^\infty(E)$. Because $\Omega^{m,0}(J^\infty(E))$ is one-dimensional, such forms are proportional to the volume form:

$$\mathbf{L} = L\,\mathrm{Vol}, \tag{38.23}$$

where L is a smooth function on the infinite jet bundle (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:

$$S : \phi \mapsto \int_M (j^\infty \phi)^* \mathbf{L}. \tag{38.24}$$

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 38.4.3 (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:

[3] In fact, the fibre bundle can be replaced with any fibred manifold.

[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.

1. X and j^kX coincide on $C^\infty(E)$.
2. j^kX preserves the contact ideal, i.e. if θ is a contact form, then $\mathcal{L}_{j^kX}\theta$ is also a contact form.

Locally, the prolongation of a vector field

$$X = X^\mu\partial_\mu + X^\alpha\partial_\alpha \tag{38.25}$$

is given by

$$(j^kX)^\alpha_I = D_I(X^\alpha - u^\alpha_\mu X^\mu) + u^\alpha_{\mu I}X^\mu \tag{38.26}$$

for all $|I| \leq k$.

A similar definitions exists for vector fields on the base manifold:

Definition 38.4.4 (Total vector field). Given a generalized vector field 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:

$$X = X^\mu\partial_\mu \longrightarrow \mathrm{tot}X = X^\mu D_\mu. \tag{38.27}$$

In particular, the total vector fields associated to a coordinate-induced basis ∂_μ are exactly the total derivatives D_μ.

Definition 38.4.5 (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 $\mathrm{Ev}(J^\infty(E))$.

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

$$X = X^\alpha_I[u]\partial^I_\alpha. \tag{38.28}$$

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:

$$j^\infty X = \sum_{|I|=0}^{\infty}(D_IX^\alpha)\partial^I_\alpha. \tag{38.29}$$

Furthermore, by writing out Cartan's magic formula 32.4.17 (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 38.4.6 (Evolutionary decomposition). Consider a generalized vector field X on E. By extending the tot-construction to generalized vector fields as

$$\mathrm{tot}X := \mathrm{tot}(\pi_*X), \tag{38.30}$$

one can define the evolutionary part of X as follows:

$$X_{\mathrm{ev}} := X - (\pi_{\infty,0})_*(\mathrm{tot}X). \tag{38.31}$$

Locally this can be written as

$$X_{\mathrm{ev}} = \big(X^\alpha - u^\alpha_\mu X^\mu\big)\partial_\alpha \tag{38.32}$$

for

$$X = X^\mu\partial_\mu + X^\alpha\partial_\alpha.$$

Using this definition the prolongation of X can be decomposed as follows:

$$j^\infty X = j^\infty X_{\mathrm{ev}} + \mathrm{tot}X. \tag{38.33}$$

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 38.4.7 (Cartan calculus). Let X, Y denote a generalized vector field and an evolutionary vector field, respectively.

1. Exterior derivatives:

$$dx^\mu = \mathbf{d}x^\mu, \tag{38.34}$$
$$\delta u^\alpha_I = \mathbf{d}u^\alpha_I - u^\alpha_{I\mu}dx^\mu. \tag{38.35}$$

2. Interior product:

$$\{d, \iota_{j^\infty Y}\}_+ = 0. \tag{38.36}$$

3. Lie derivatives:

$$\mathcal{L}_{j^\infty X} = \mathbf{d}\iota_{j^\infty X} + \iota_{j^\infty X}\mathbf{d}, \tag{38.37}$$
$$\mathcal{L}_{j^\infty Y} = \delta\iota_{j^\infty Y} + \iota_{j^\infty Y}\delta, \tag{38.38}$$
$$[d, \mathcal{L}_{j^\infty Y}] = [\delta, \mathcal{L}_{j^\infty Y}] = 0. \tag{38.39}$$

38.4.2 Euler operators

Property 38.4.8 (Total differential operators). A differential operator $P : \mathrm{Ev}(J^\infty(E)) \rightarrow \Omega^\bullet(J^\infty(E))$ that can locally be written as

$$P(X) = \sum_{|I|=0}^{k}(D_I X^\alpha)P^I_\alpha \tag{38.40}$$

for some differential forms P^I_α. By (formally) integrating by-parts, this can locally be written as

$$P(X) = \sum_{|I|=0}^{k} D_I\big(X^\alpha Q^I_\alpha\big), \tag{38.41}$$

where the smooth forms Q^I_α can be expressed as follows:

$$Q^I_\alpha = \sum_{|J|=0}^{k-|I|}\binom{|I|+|J|}{|J|}(-1)^{|J|}D_J P^{IJ}_\alpha. \tag{38.42}$$

The zeroth-order part Q_α also defines a total differential operator:

$$E_P(X) := X^\alpha Q_\alpha. \tag{38.43}$$

This operator is called the **Euler operator** (associated to P).

The following property shows that the Euler-operator through a formal "integration-by-parts" formula:

Property 38.4.9 (Decomposition of total differential operators). Consider a total differential operator $P : \mathrm{Ev}(J^\infty(E)) \to \Omega^{\dim(M),k}(J^\infty(E))$. The Euler operator E_P is th 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:

$$P(X) = E_P(X) + dR_U(X). \tag{38.44}$$

The operator R_U can locally be expressed as follows:

$$R_U(X) = \sum_{|I|=0}^{k-1} D_I\big(X^\alpha D_\mu \lrcorner Q^{\mu I}_\alpha\big). \tag{38.45}$$

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 38.4.10 (Euler-Lagrange operator). Consider a local Lagrangian $\mathbf{L}$. This form induces a total differential operator as follows:

$$P_{\mathbf{L}}(X) := \mathcal{L}_{j^\infty X}\mathbf{L} = \sum_{|I|=0}^{k} (D_I X^\alpha)(\partial^I_\alpha L)\,\mathrm{Vol}\,. \tag{38.46}$$

The coefficients (38.42) associated to this operator are given by the following formula (to turn these coefficients into true forms one should multiple by the volume form):

$$E^I_\alpha(L) := \sum_{|J|=0}^{k-|I|} \binom{|I|+|J|}{|J|}(-1)^{|J|} D_J\big(\partial^{IJ}_\alpha L\big). \tag{38.47}$$

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

$$\delta_{\mathrm{EL}}\mathbf{L} := E_\alpha(L)\delta u^\alpha \wedge \mathrm{Vol}\,. \tag{38.48}$$

An explicit formula for the Euler-Lagrange derivative is given by the following formula (this is just Equation (38.47) for $|I| = 0$):

$$\delta_{\mathrm{EL}}L := \left(\frac{\partial L}{\partial u^\alpha} - D_\mu \frac{\partial L}{\partial u^\alpha_\mu} + \cdots\right)\delta u^\alpha \tag{38.49}$$

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 38.4.11 (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

$$\delta_{\mathrm{EL}}((j^\infty\phi)^*\mathbf{L}) = (j^\infty\phi)^*(\delta_{\mathrm{EL}}(\mathbf{L})). \tag{38.50}$$

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 38.4.12 (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 38.4.13 (Local variational formula). Decomposition 38.4.9 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

$$\mathcal{L}_{j^\infty X}\mathbf{L} = j^\infty X \lrcorner \delta_{\mathrm{EL}}(\mathbf{L}) + d(j^\infty X \lrcorner \gamma) \tag{38.51}$$

for some $\gamma \in \Omega^{\dim(M)-1,1}(J^\infty(E))$.

38.4.3 Functional complex

Example 38.4.14 (Interior Euler operator). For every smooth form $\omega \in \Omega^{p,q}(J^\infty(E))$ one can define a total differential operator as follows:

$$P_\omega(X) := j^\infty X \lrcorner \omega. \tag{38.52}$$

As for the previous example this operator induces (higher) Euler operators:

$$F^I_\alpha(\omega) := \sum_{|J|=0}^{k-|I|} \binom{|I|+|J|}{|J|}(-1)^{|J|} D_J\big(\partial^{IJ}_\alpha \lrcorner \omega\big). \tag{38.53}$$

Since in this case they arise from interior multiplication, they are called interior Euler operators. For $p = \dim(M)$ one again obtains a globally defined Euler operator (also called the interior Euler operator):

$$I(\omega) := \frac{1}{q}\delta u^\alpha \wedge F_\alpha(\omega). \tag{38.54}$$

The interior Euler operator defines a sequence of spaces, the so-called spaces of **functional forms**, as follows:

$$\mathcal{F}^q(J^\infty(E)) := \Big\{\omega \in \Omega^{\dim(M),q}(J^\infty(E)) \,\Big|\, I(\omega) = \omega\Big\}. \tag{38.55}$$

Property 38.4.15. The interior Euler operator has the following important properties:

- I is a projector $I^2 = I$.
- 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 38.4.16. 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 38.4.17 (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

$$A^I_{\alpha\beta} = (-1)^{|I|+1} A^I_{\beta\alpha}. \tag{38.56}$$

The local $(\dim(M), 2)$-form

$$w^{|I|} := \delta u^\alpha \wedge \left(A^I_{\alpha\beta} \delta u^\beta_I + D_I(A^I_{\beta\alpha} \delta u^\beta) \right) \wedge \mathrm{Vol} \tag{38.57}$$

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

$$\omega = w^0 + w^1 + w^2 + \cdots . \tag{38.58}$$

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 38.4.18 (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

$$P : \mathrm{Ev}(J^\infty(E)) \to \Omega^{\dim(M),q-1}(J^\infty(E))$$

such that

$$\omega = \delta u^\alpha \wedge P_\alpha. \tag{38.59}$$

This representation is unique if it exists.

The forms in $\mathcal{F}^q$ are said to be functional due to the following property:

Property 38.4.19 (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:

$$W_\omega(X_1, \ldots, X_q)[\sigma] := \int_K (j^\infty\sigma)^* \omega(j^\infty X_1, \ldots, j^\infty X_q). \tag{38.61}$$

The number 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 38.4.20. The differential δ_V on $\mathcal{F}^\bullet$ induces a differential on the space of functionals of the above form:

$$\delta W_\omega := W_{\delta_V \omega}. \tag{38.62}$$

An equivalent definition can be given by a formula similar to (32.43), where an evolutionary vector field X acts on a degree-0 functional by Lie derivation:

$$X(W_\omega[\sigma]) := \int_V (j^\infty\sigma)^* (\mathcal{L}_{j^\infty X}\omega). \tag{38.63}$$

[5] This is a generalization of 38.3.13: P is said to be formally skew-adjoint if there exists an $(m-1, q-1)$-form K such that

$$j^\infty X \lrcorner P(Y) + j^\infty Y \lrcorner P(X) = dK \tag{38.60}$$

for all $X, Y \in \mathrm{Ev}(J^\infty(E))$.

38.4.4 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:

$$\Phi^\sharp : \Omega^{p,q}(J^\infty(E')) \to \Omega^{p,q}(J^\infty(E)) : \omega \mapsto \pi^{p,q}(\Phi^*\omega). \tag{38.64}$$

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 38.4.21. The projection $\pi^{p,q} : \Omega^{p+q}(J^\infty(E)) \to \Omega^{p,q}(J^\infty(E))$ is defined by substituting $\mathbf{d}u^\alpha_I \longrightarrow \delta u^\alpha_I + u^\alpha_{\mu I}dx^\mu$ 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 (38.61) only care about (m, q)-forms for a specific q (in particular **action functionals**, i.e. $q = 0$).

Formula 38.4.22 (Local Lagrangians). For local Lagrangians $\mathbf{L} \equiv L\,\mathrm{Vol}_{E'}$, the projected pullback acts as follows:

$$\Phi^\sharp\mathbf{L} = (L \circ \Phi)\det(D_\mu f^\mu)\,\mathrm{Vol}_E, \tag{38.65}$$

where $\pi'(\Phi[u]) = (f^\mu)$. So we obtain 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 32.4.9. The following property states the "naturality" of the different operators on the variational bicomplex with respect to the projected pullback:

Property 38.4.23. 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 38.4.24. 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:

$$\mathcal{L}^\sharp_X : \Omega^{p,q}(J^\infty(E)) \to \Omega^{p,q}(J^\infty(E)) : \omega \mapsto \pi^{p,q}(\mathcal{L}_X\omega). \tag{38.66}$$

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 (38.55): $(\mathcal{F}^\bullet, \delta_V)$. 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 38.4.25. Let 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
$$Y = -\frac{\partial S}{\partial u_\mu}\partial_\mu + \left(S - u_\mu \frac{\partial S}{\partial u_\mu}\right)\partial_u, \tag{38.67}$$
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:
$$\Phi^\natural : \Omega^{m,q}(J^\infty(E)) \to \mathcal{F}^q(J^\infty(E')) : \omega \mapsto (I \circ \Phi^\sharp)\omega \tag{38.68}$$
$$\mathcal{L}_X^\natural : \Omega^{m,q}(J^\infty(E)) \to \mathcal{F}^q(J^\infty(E)) : \omega \mapsto (I \circ \mathcal{L}_X^\sharp)\omega. \tag{38.69}$$

Property 38.4.26. 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 38.4.27 (Euler-Lagrange operator). Consider a local Lagrangian $\mathbf{L}$. If Φ is a contact transformation, then
$$\delta_{\text{EL}}(\Phi^\sharp \mathbf{L}) = \Phi^\natural(\delta_{\text{EL}}\mathbf{L}). \tag{38.70}$$
If X is a generalized vector field, then
$$\delta_{\text{EL}}(\mathcal{L}_{j^\infty X}^\sharp \mathbf{L}) = \mathcal{L}_{j^\infty X}^\natural(\delta_{\text{EL}}\mathbf{L}). \tag{38.71}$$

Formula 38.4.28 (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.
$$\mathcal{L}_{j^\infty X}^\natural \omega = \delta_V(j^\infty X_{\text{ev}} \lrcorner\, \omega) + I(j^\infty X_{\text{ev}} \lrcorner\, \delta_V\omega) \tag{38.72}$$
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 (Property 5.1.7) will be followed, i.e. a cochain map will be found $h : C_\bullet \to C_\bullet$ such that
$$\mathbb{1} = d \circ h + h \circ d. \tag{38.73}$$

Property 38.4.29 (Vertical complex is exact). Homotopy operators $h_V^{p,q}:\Omega^{p,q}\rightarrow\Omega^{p,q-1}$ for the vertical complex

$$0\longrightarrow\Omega^p(M)\xrightarrow{\pi_\infty^*}\Omega^{p,0}\xrightarrow{\delta}\Omega^{p,1}\xrightarrow{\delta}\cdots \tag{38.74}$$

are given by the following formula

$$h_V^{p,q}(\omega)=\int_0^1\frac{1}{t}\Phi^*_{\log t}(j^\infty R\lrcorner\,\omega)dt, \tag{38.75}$$

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 Equation (38.95).

The analogous statement for the horizontal complex is a bit more involved:

Property 38.4.30 (Augmented horizontal complex is exact). The homotopy operators $h_H^{p,q}:\Omega^{p,q}\rightarrow\Omega^{p-1,q}$ for the augmented horizontal complex

$$0\longrightarrow\Omega^{0,q}\xrightarrow{d}\Omega^{1,q}\xrightarrow{d}\cdots\xrightarrow{d}\Omega^{\dim(M),q}\xrightarrow{I}\mathcal{F}^q\xrightarrow{I}0 \tag{38.76}$$

are given by the following formula

$$h_H^{p,q}(\omega)=\frac{1}{q}\sum_{|I|=0}^{k-1}\frac{|I|+1}{m-p+|I|+1}D_I\big(\delta u^\alpha\wedge F^{I\mu}_\alpha(D_\mu\lrcorner\,\omega)\big), \tag{38.77}$$

where the F^I_α are the interior Euler operators (38.53).

Corollary 38.4.31 (Functional decomposition). The de Rham spaces on $J^\infty(E)$ admit the following decomposition:

$$\Omega^{p,q}(J^\infty(E))=d\Omega^{p-1,q}(J^\infty(E))\oplus\mathcal{F}^q(J^\infty(E)), \tag{38.78}$$

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):

$$0\longrightarrow\mathbb{R}\longrightarrow\Omega^{0,0}\xrightarrow{d}\Omega^{1,0}\xrightarrow{d}\cdots\xrightarrow{d}\Omega^{\dim(M),0}\xrightarrow{\delta_{\mathrm{EL}}}\mathcal{F}^1\xrightarrow{\delta_V}\mathcal{F}^2\xrightarrow{\delta_V}\cdots. \tag{38.79}$$

Explicit formulas are not shown since these are too complicated for the current objective. See [65] for a full account.

Remark 38.4.32 (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. differentials forms of the form $\delta_{\mathrm{EL}}\mathbf{L}$. From the form of the homotopy operator $\mathcal{F}^1\rightarrow\Omega^{m,0}$ 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 result in a Lagrangians of minimal order.

38.4.5 Variational problems

The forms in $\mathcal{F}^1$ admit the following characterization:

Definition 38.4.33 (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:

$$\Omega^{\dim(M),1}_{\text{source}}(E) := \delta C^\infty(E) \wedge \Omega^{\dim(M),0}(E). \tag{38.80}$$

Locally it can be written as

$$\omega = \omega_\alpha(x, u, u_I)\delta u^\alpha \wedge \mathrm{Vol}\,. \tag{38.81}$$

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 38.4.34 (First variational formula). In Property 38.4.13 it was shown that the Euler-Lagrange operator locally satisfies

$$\mathcal{L}_{j^\infty X}\mathbf{L} = j^\infty X \lrcorner\, \delta_{\text{EL}}\mathbf{L} + d(j^\infty X \lrcorner\, \gamma) \tag{38.82}$$

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 38.4.31, one can find a globally defined form that still satisfies the same formula (with the disadvantage that it does not admit a canonical expression):

$$\delta\mathbf{L} = \delta_{\text{EL}}\mathbf{L} + d\sigma. \tag{38.83}$$

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

$$\forall X \in \mathrm{Ev}(J^\infty(E)) : \mathcal{L}_{j^\infty X}\mathbf{L} = 0. \tag{38.84}$$

Equation (38.83) then becomes:

$$\delta_{\text{EL}}\mathbf{L} + d\sigma = 0. \tag{38.85}$$

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 38.4.35 (Global variational formula). The variational formula can also be extended to all generalized vector fields:

$$\mathcal{L}^\sharp_{j^\infty X}\mathbf{L} = X_{\text{ev}} \lrcorner\, \delta_{\text{EL}}(\mathbf{L}) + d\sigma \tag{38.86}$$

for some $\sigma \in \Omega^{\dim(M),0}(J^\infty(E))$.

Definition 38.4.36 (Lepage form). A $\dim(M)$-form $\rho \in \Omega^\bullet(J^\infty(E))$ such that

$$\pi^{\dim(M),0}(X \lrcorner\, \mathbf{d}\rho) = 0 \tag{38.87}$$

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 38.4.37 (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 38.4.38 (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

$$\mathcal{L}^\natural_{j^\infty X}\Delta = 0. \tag{38.88}$$

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 38.4.27 and the fact that δ_{EL} annihilates d-exact forms 38.4.16:

Definition 38.4.39 (Bessel-Hagen symmetry). A Bessel-Hagen or **divergence** symmetry of a Euler-Lagrange form $\delta_{\text{EL}}(\mathbf{L})$ is a generalized vector field X such that

$$\mathcal{L}^\sharp_{j^\infty X}\mathbf{L} = d\eta \tag{38.89}$$

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 38.4.40 (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

$$\delta_{\text{EL}}(j^\infty X \lrcorner \Delta) = 0 \tag{38.90}$$

or, again by 38.4.16,

$$X \lrcorner \Delta = d\eta \tag{38.91}$$

for some local $\eta \in \Omega^{\dim(M),0}(J^\infty(E))$. Corollary 38.4.35 shows that to every global (generalized) symmetry, there corresponds a global conservation law.

Remark 38.4.41. 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 38.4.42 (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 38.4.43 (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.*

38.4.6 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 $\delta_{\text{EL}}\mathbf{L}$, 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 38.4.44 (Helmholtz conditions). By applying the Euler-Lagrange operator δ_{EL} to its defining variational formula 38.4.13 and using the (infinitesimal) naturality condition (38.50) and the fact that it annihilates d-exact forms 38.4.16, it can be seen that a source form Δ can be obtained from a local Lagrangian if

$$\mathcal{L}_{j^\infty X}\Delta = \delta_{\text{EL}}(X \lrcorner \Delta) \tag{38.92}$$

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]

$$\delta_V\Delta = 0. \tag{38.93}$$

Source forms satisfying this condition are said to be **locally variational**.

Formula 38.4.45 (Local expression). Consider a source form that admits the local expression

$$\Delta = P_\alpha \delta u^\alpha \wedge \text{Vol}\,.$$

The Helmholtz conditions can locally be expressed as follows:

$$(-1)^{|I|}\partial^I_\alpha P_\beta = E^I_\beta(P_\alpha), \tag{38.94}$$

where the E^I_β are the Lie-Euler operators (38.47).

Example 38.4.46 (Volterra-Vainberg formula). If E is trivial and $\Delta = F_\alpha\delta u^\alpha \wedge \text{Vol}$ satisfies the Helmholtz conditions, then

$$L := \int_0^1 u^\alpha F_\alpha[tu]dt \tag{38.95}$$

satisfies $\Delta = \delta_{\text{EL}}\mathbf{L}$. If Δ is homogeneous of degree k in u, this can be expressed as

$$\mathbf{L} = \frac{1}{k+1}\iota_R\Delta, \tag{38.96}$$

where $R := u^\alpha\partial_\alpha$ is the (vertically) radial vector field on E.

38.4.7 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 are mapped to forms of order $k+1$. Therefore, attention will have to be restricted to a specific subcomplex of $\Omega^\bullet(J^\infty(E))$:

$$\Omega^\bullet_k(E) \subset \Omega^\bullet(J^{k+1}(E)) := \delta\text{-closure of } \Omega^\bullet(J^k(E)). \tag{38.97}$$

From the basic definitions and properties of the two differentials d,δ, it follows that $\Omega^\bullet_k$ is generated by $C^\infty(J^k(E))$, the horizontal basis $\{dx^\mu\}_{\mu\leq\dim(M)}$ and the contact basis $\{\delta u^\alpha_I\}^{|I|\leq k}_{\mu\leq\dim(M)}$.

[6] This expression also immediately follows from Property 38.4.15.

The next step is to further restrict to a horizontally closed subcomplex. To this end, consider forms $\omega \in \Omega_k^{p,q}(E)$ of the form

$$\omega = \left[du^{\alpha_1}_{I_1} \wedge \ldots \wedge du^{\alpha_r}_{I_r} \wedge d\delta u^{\beta_1}_{J_1} \wedge \ldots \wedge d\delta u^{\beta_s}_{J_s}\right] \wedge f dx^{\kappa_1} \wedge \ldots \wedge dx^{\kappa_{p-r-s}}, \tag{38.98}$$

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:

$$J = \frac{1}{(r+s)!}\frac{D(u^{\alpha_1}_{I_1}, \ldots, u^{\alpha_r}_{I_r}, \delta u^{\beta_1}_{J_1}, \ldots, \delta u^{\beta_s}_{J_s})}{D(x^{\mu_1}, \ldots, x^{\mu_r}, x^{\nu_1}, \ldots, x^{\nu_s})} dx^{\mu_1} \wedge \ldots \wedge dx^{\mu_r} \wedge dx^{\nu_1} \wedge \ldots \wedge dx^{\nu_s}. \tag{38.99}$$

This factor has the form of a Jacobian determinant (with respect to total derivatives) and as such the subcomplex spanned by the above forms the is called the **Jacobian (sub)complex** $\mathcal{J}_k^\bullet(E)$.

Property 38.4.47 (Alternative characterizations). The Jacobian complexes 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 38.4.21).

$$\mathcal{J}_k^{p,q}(E) = \Omega^{p,q}(J^\infty(E)) \cap \delta\text{-closure of } \pi^{\bullet,0}\Big(\Omega^\bullet(J^{k-1}(E))\Big). \tag{38.100}$$

- For $p < m$ the Jacobian complex contains those forms for which d does not increase the order:

$$\mathcal{J}_k^{p,q}(E) = \{\omega \in \Omega_k^{p,q} \mid d\omega \in \Omega_k^{p+1,q}(E)\}. \tag{38.101}$$

- If $\omega \in \mathcal{J}_k^{p,q}(E)$, 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 38.4.48 (Exactness). Let E be trivial. If $\delta_{\mathrm{EL}}\omega = 0$ for $\omega \in \Omega_k^{m,q=0}(E)$ or $I(\omega) = 0$ for $\omega \in \Omega_k^{m,q\geq 1}(E)$, then $\omega \in \mathcal{J}_k^{m,q}(E)$ and $\omega = d\eta$ for $\eta \in \mathcal{J}_k^{m-1,q}(E)$.

Property 38.4.49 (Functional dependence of Lagrangians). If Δ is a locally variational source form of order k, it is polynomial of degree m in k^{th}-order derivatives of the u^α.

38.5 Partial differential equations

In this section the content of Chapter 19 is generalized using the language of differential geometry.

38.5.1 Algebraic formulation

In this section partial differential equations of the form

$$f(x, u, u_I) = 0 \tag{38.102}$$

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 is the order of the PDE and m 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 defines a subbundle Σ^0 of the finite jet bundle $J^k(E)$ and a solution is simply a section of Σ^0.

Remark 38.5.1. For every $l \in \mathbb{N}$, define the subspace

$$\Sigma^l := \{(x, u, u_I) \in J^{k+l}(E) \mid \forall |J| \leq l : D_J f(x, u, u_I) = 0\} = J^l(\Sigma_0) \cap J^{k+l}(E),$$

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 38.5.2 (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 38.5.3 (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 38.5.4. 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 38.5.5 (Exterior system). Consider a general PDE of the form (38.102). This PDE can be turned into a set of differential forms on the zero locus of f, i.e. Σ_0:

$$\begin{aligned}\theta &:= du - u_\mu dx^\mu \\ &\ \vdots\end{aligned} \tag{38.103}$$

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 32.5.7 or, combined with the PDE $f = 0$ and its differential consequences, it gives a characteristic system 32.5.9. The associated distribution becomes integrable (sometimes called **completely integrable**) if and only if the ideal is a differential ideal by Frobrenius's theorem 32.5.6. The integral manifolds of the distribution then give a solution of the PDE.

Formula 38.5.6 (Lagrange-Charpit equations). Let $f = 0$ be a first-order PDE on an n-dimensional manifold and consider the differential closure of the exterior system of $f = 0$:

$$\begin{cases} f &= 0 \\ df &= \left(\dfrac{\partial f}{\partial x^\mu} + u_\mu \dfrac{\partial f}{\partial u}\right) dx^\mu + \dfrac{\partial f}{\partial u_\mu} du_\mu = 0 \\ \theta &= 0 \\ d\theta &= 0. \end{cases} \tag{38.104}$$

The characteristic system of these equations is given by the PDE itself together with the Lagrange-Charpit equations:

$$\frac{dx^1}{\partial f/\partial u_1} = \cdots = \frac{dx^n}{\partial f/\partial u_n} = \frac{-du_1}{\partial f/\partial x^1 + u_1(\partial f/\partial u)} = \cdots = \frac{du_n}{\partial f/\partial x^n + u_n(\partial f/\partial u)} = \frac{du}{\sum_\mu u_\mu(\partial f/\partial u_\mu)}. \tag{38.105}$$

Definition 38.5.7 (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 n-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**.

38.5.2 Symmetries

Property 38.5.8. A Lie group G with Lie algebra $\mathfrak{g}$ is the symmetry group of a (nondegenerate) system F of PDEs if and only if

$$F = 0 \implies j^\infty X(F) = 0 \tag{38.106}$$

for all generators $X \in \mathfrak{g}$.

?? COMPLETE ??

38.5.3 Pseudogroups ♣

Example 38.5.9 (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 7.2.19.

By the inverse function theorem 32.1.10 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 4.10.1. 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 38.5.10 (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 an $n \in \mathbb{N}_0$, called the **order**, such that for all $r \geq n$ the jets $\pi_r : \mathcal{G}^{(r)} \to M$ form an embedded submanifold of $\Pi_r : \mathcal{D}^{(r)} \to M$ and such that the jet projections $\pi_{r+1,r} : \mathcal{G}^{(r+1)} \to \mathcal{G}^{(r)}$ are fibrations, then $\mathcal{G}$ is called a regular pseudogroup.

Definition 38.5.11 (Lie pseudogroup). Let $\mathcal{G} \subset \mathcal{D}^\omega$ be a regular analytic pseudogroup of order k. 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 38.5.12. Let $\mathcal{G}$ be a Lie pseudogroup of order k. The regularity condition implies that for all $r \geq k$ the jet bundle $\mathcal{G}^{(r)}$ is described by a set of r^{th}-order PDEs

$$F\left(z, \phi^{(r)}\right) = 0.$$

The (local) solutions to these equations are exactly the analytic functions that have $(z_0, \phi_0^{(r)})$ as local coordinates of their r-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 38.5.13 (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 39

K-theory ♣

The main reference for this chapter is [57].

In this chapter all topological (base) 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 8.5.1. In general all statements will given for an arbitrary base field k. When it is necessary to restrict to the specific cases $k = \mathbb{R}, \mathbb{C}$, this will be explicitly stated.

39.1 Preliminaries

Definition 39.1.1 (Stable general linear group). Let R be a (unital) ring. For every two integers $m < n$ there exists a canonical inclusion $\mathrm{GL}(m, R) \hookrightarrow \mathrm{GL}(n, R)$ through extension (direct sum) by $\mathbb{1}_{n-m}$. This allows to define the stable general linear group (or infinite general linear group) as a direct limit:

$$\mathrm{GL}(R) := \varinjlim_{n\in\mathbb{N}} \mathrm{GL}(n, R). \tag{39.1}$$

Remark 39.1.2. A similar construction leads to the stable orthogonal and stable unitary groups O and U. Sometimes the notations $\mathrm{O}(\infty)$ and $\mathrm{U}(\infty)$ are also used when $R = \mathbb{R}$ and $R = \mathbb{C}$, respectively.

39.2 Topological K-theory

39.2.1 Introduction

Definition 39.2.1 (K-theory). Let $\mathrm{Vect}(X)/\sim$ 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 structure $(\mathrm{Vect}(X)/\sim, \oplus)$ forms an Abelian monoid. The Grothendieck completion 3.2.6 of this monoid is called the (real) K-theory of X.

Notation 39.2.2. The K-theory of a space X is denoted by $K^0(X)$. If one wants to emphasize the base field k, the notation $K^0_k(X)$ 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 39.2.3 (Point). Let $*$ be the one-point space. The K-theory $K^0(*)$ is isomorphic to the additive group of integers $(\mathbb{Z}, +)$.

Definition 39.2.4 (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 39.2.5 (Virtual rank). The virtual rank of the virtual bundle $([E],[E'])$ is defined as follows:

$$\mathrm{rk}([E],[E']) := \mathrm{rk}(E) - \mathrm{rk}(E'). \tag{39.2}$$

Property 39.2.6. Property 32.2.11 implies that every virtual bundle is of the form $[E]-[X\times\mathbb{R}^n]$ for some vector bundle E and integer $n\in\mathbb{N}$.

Definition 39.2.7 (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 39.2.8. The reduced K-theory $\widetilde{K}(X)$ can equivalently be defined as follows. Consider the stable isomorphism classes 32.2.12 of vector bundles over X. Under Whitney sums these define a commutative group $(\mathrm{Vect}(X)/\sim_{\text{stable}},\oplus)$, which is (naturally) isomorphic to $\widetilde{K}^0(X)$.

The following construction is very similar to 25.3.7. In fact it is the one obtained for the Banach functor that restricts vector bundles to subspaces.

Definition 39.2.9 (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 said to be 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 $\mathrm{Aut}(F|_Y)$.

The relative K-theory $K^0(X,Y)$ is defined as the quotient of $\mathscr{V}(X,Y)$ by the following equivalence relation:

$$x\sim x' \iff \exists e,e'\in\mathscr{E}(X,Y) : x+e\cong_{\mathscr{V}} x'+e'. \tag{39.3}$$

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, $K^0(X,\emptyset)\equiv K^0(X)$.

Alternative Definition 39.2.10. Consider a space X with a closed subspace Y.

$$K^0(X,Y) := \ker\bigl(K^0(X)\to K^0(Y)\bigr). \tag{39.4}$$

Property 39.2.11 (Excision). Consider a space X together with a closed subspace Y. The relative K-theory is related to the reduced K-theory as follows:

$$K^0(X,Y)\cong\widetilde{K}^0(X/Y). \tag{39.5}$$

39.2.2 Classification

Property 39.2.12 (Orthogonal groups). Consider the classifying space 33.2.1 of the orthogonal group $\mathrm{O}(n)$ and recall the Grassmannian $\mathrm{Gr}(n,\mathbb{R}^N)$ of n-dimensional subspaces in $\mathbb{R}^k$. There exists a canonical inclusion of Grassmannians:

$$\iota_k : \mathrm{Gr}(n,\mathbb{R}^k)\hookrightarrow\mathrm{Gr}(n,\mathbb{R}^{k+1}) : W\mapsto W. \tag{39.6}$$

By taking the direct limit of these inclusions, one obtains the infinite Grassmannian:

$$BO(n) := \varinjlim_{k\in\mathbb{N}} \mathrm{Gr}(n, \mathbb{R}^k). \tag{39.7}$$

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

$$\iota_{n,k} : \mathrm{Gr}(n, \mathbb{R}^k) \hookrightarrow \mathrm{Gr}(n+1, \mathbb{R}^{k+1}) : W \mapsto W \oplus \mathrm{span}\{e_{k+1}\}. \tag{39.8}$$

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 39.2.13. A similar construction allows to construct the classifying spaces $BU(n)$ and BU by starting from complex Grassmannians.

Remark 39.2.14. It should be noted that neither BO nor BU can be expressed as classifying spaces of a group over some infinite-dimensional Hilbert space. 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 39.2.15 (Homotopy classification). For all spaces X the K-theory can be represented as follows:

$$K_k^0(X) = [X, B\mathrm{GL}(k) \times \mathbb{Z}]. \tag{39.9}$$

When specializing to $k = \mathbb{R}, \mathbb{C}$ this gives

$$KO^0(X) = [X, BO \times \mathbb{Z}] \tag{39.10}$$

and

$$K^0(X) = [X, BU \times \mathbb{Z}] \tag{39.11}$$

due to homotopy invariance. Reduced K-theory can be obtained by considering basepoint-preserving homotopies, which for connected spaces gives $[X, B\mathrm{GL}]$.

Remark 39.2.16 (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 the Grothendieck K-theory to the representable K-theory.

The following theorem should be compared to Remark 39.2.14 above (in fact this theorem can be proven through *Kuiper's theorem*):

Theorem 39.2.17 (Atiyah-Jänich). *The space of Fredholm operators 23.4.37 on a separable, infinite-dimensional Hilbert space forms a classifying space for K-theory.*

39.2.3 Negative degree

Definition 39.2.18 (K^{-1}). For every field k the relative K-functor K^{-1} is defined as follows (this can again be obtained as a property when using a different definition):

$$K^{-1}(X, Y) := [X/Y, \mathrm{GL}(k)]_*, \tag{39.12}$$

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 relation (7.11):

$$K_k^{-1}(X) := [X, \mathrm{GL}(k)]. \tag{39.13}$$

For $k = \mathbb{R}, \mathbb{C}$ one can use homotopy invariance to obtain

$$KO^{-1}(X) := [X, \mathrm{O}] \tag{39.14}$$
$$K^{-1}(X) := [X, \mathrm{U}]. \tag{39.15}$$

To define lower degree groups it will be useful to extend K-theory to locally compact spaces:

Definition 39.2.19 (K-theory of locally compact spaces). Let X be a locally compact space and denote its one-point compactification 7.5.27 by $\widehat{X}$. The groups $K^0(X)$ and $K^{-1}(X)$ are defined as follows:

$$K^0(X) := \ker\Big(K^0(\widehat{X}) \to K^0(\{\infty\})\Big) \tag{39.16}$$
$$K^{-1}(X) := \ker\Big(K^{-1}(\widehat{X}) \to K^{-1}(\{\infty\})\Big) \tag{39.17}$$

So the K-theory of a locally compact space is defined as the reduced K-theory of its one-point compactification.

Corollary 39.2.20 (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

$$K^0(X\backslash Y) = \widetilde{K}^0(X/Y). \tag{39.18}$$

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 $X\backslash Y$:

$$K^0(X, Y) \cong K^0(X\backslash Y). \tag{39.19}$$

Definition 39.2.21 (K^{-n}). Lower-degree relative K-groups are defined as follows:

$$K^{-n}(X, Y) := K^0((X\backslash Y) \times \mathbb{R}^n). \tag{39.20}$$

By taking $Y = \emptyset$ one obtains the groups $K^{-n}(X)$ as K^0-groups of trivial line bundles:

$$K^{-n}(X) := K^0(X \times \mathbb{R}^n). \tag{39.21}$$

Before relating this to reduced K-theory, the reader should be warned about 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 39.2.19.

It can be shown through a series of homeomorphisms that the above definition is equivalent to the following one:

$$K^{-n}(X, Y) \cong \widetilde{K}^0(\Sigma^n(X/Y)), \tag{39.22}$$

where Σ denotes the reduced suspension functor 7.3.7. As such, the reduced suspension functor gives a way to move down in the tower of K-groups.

39.2.4 Bott periodicity

Definition 39.2.22 (Cup product). First the Whitney sum and tensor product constructions are generalized to vector bundles over different base spaces. Let E, E' be vector bundles over the base spaces B, B' and consider the projection maps $\pi : B \times B' \to B$ and $\pi : B \times B' \to B'$. The exterior sum $E \oplus E' \to B \times B'$ 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 39.2.4 it follows that every element $x \in K^0(X)$ can be written as 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:

$$([E] - [E']) \cup ([F] - [F']) := [E \otimes F] + [E' \otimes F'] - [E \otimes F'] - [E' \otimes F]. \tag{39.23}$$

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 defined above. By restricting to $Y \times Y'$ one can then obtain an element of $K^0(Y \times Y')$:

$$K^0(Y) \times K^0(Y') \xrightarrow{\iota \times \iota'} K^0(\widehat{Y}) \times K^0(\widehat{Y'}) \xrightarrow{\cup} K^0(\widehat{Y} \times \widehat{Y'}) \xrightarrow{\text{res}} K^0(Y \times Y'), \tag{39.24}$$

where ι, ι' are the inclusions induced by (39.16).

Property 39.2.23 (Ring structure). By precomposing with the diagonal morphism $K^0(X \times X) \to K^0(X)$ 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 39.2.24. 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 27.1.5.

By using the isomorphism (39.19) one can also extend the cup product to an operation on relative K-theory:

$$K^{-m}(X, Y) \to K^{-n}(X', Y') \to K^{-m-n}(X \times X', X \times Y' \cup X' \times Y). \tag{39.25}$$

Notation 39.2.25 (Bott element). Consider the complex relative K-group

$$K^0_{\mathbb{C}}(D^2, S^1) \cong \widetilde{K}^0_{\mathbb{C}}(S^2) \cong K^0_{\mathbb{C}}(\mathbb{R}^2)$$

of the unit disk with respect to its boundary. The Bott element represented by the triple $\big(D^2 \times \mathbb{C}^2, D^2 \times \mathbb{C}^2, \alpha : (x, v) \mapsto xv\big)$ will be denoted by β. 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 39.2.26 (Complex Bott periodicity). *The cup product with the Bott element β 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 β gives an isomorphism $K^0(X) \cong K^0(X \times \mathbb{R}^2)$.*

Corollary 39.2.27. Applying Bott periodicity to the case $X = *, Y = \emptyset$ and comparing to Example 39.2.3, one obtains $K^0(D^2, S^1) \cong \mathbb{Z}$. One can also conclude that β is a generator of $K^0(D^2, S^1)$.

Corollary 39.2.28 (Spheres). Bott periodicity and Equation (39.22) also allow to compute the reduced K-theory of spheres:

$$\widetilde{K}^0(S^n) = \begin{cases} 0 & n \text{ odd} \\ \mathbb{Z} & n \text{ even.} \end{cases} \tag{39.26}$$

For n even one can see that the generator is given by $\beta^{n/2}$.

Property 39.2.29 (Homotopy groups of unitary group). Property 32.2.4 can be generalized to the stable linear group to obtain an isomorphism $\widetilde{K}^0_k(S^n) \cong \pi_{n-1}(\mathrm{GL}(k))$. By specializing to $k = \mathbb{C}$, recalling that $\mathrm{GL}(m,\mathbb{C})$ deformation retracts onto $\mathrm{U}(m)$ 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

$$\mathrm{U}(n) \to \mathrm{U}(n+1) \to S^{2n+1}, \tag{39.27}$$

one can show that for $n > i/2$ the homotopy groups $\pi_i(\mathrm{U}(n))$ satisfy the same periodic relation.

Theorem 39.2.30 (Real Bott periodicity). *The cup product with the real Bott element, i.e. the generator of $KO^{-8}(*) \cong \mathbb{Z}$, gives an isomorphism*

$$KO^{-n}(X,Y) \to KO^{-n-8}(X,Y). \tag{39.28}$$

Theorem 39.2.31 (Weak Bott periodicity). *The following spaces are homotopy equivalent:*[1]

$$\mathrm{GL}(\mathbb{R}) \sim \Omega^8\,\mathrm{GL}(\mathbb{R}) \tag{39.29}$$
$$\mathrm{O} \sim \Omega^8\mathrm{O}$$
$$\mathrm{U} \sim \Omega(\mathbb{Z} \times B\mathrm{U}) \tag{39.30}$$
$$\mathbb{Z} \times B\mathrm{U} \sim \Omega\mathrm{U}. \tag{39.31}$$

Corollary 39.2.32. Through Eckmann-Hilton duality this implies the periodicity in the homotopy groups of the stable orthogonal and unitary groups (cf. Property 39.2.29).

Property 39.2.33. One can also relate real and quaternionic K-theory through the following homotopy equivalences:

$$\mathbb{Z} \times B\mathrm{GL}(\mathbb{R}) \sim \Omega^4(B\mathrm{GL}(\mathbb{H})) \tag{39.32}$$
$$\mathbb{Z} \times B\mathrm{GL}(\mathbb{H}) \sim \Omega^4(B\mathrm{GL}(\mathbb{R})). \tag{39.33}$$

Remark 39.2.34 (Positive degree). Bott periodicity also allows to define K-groups in positive degree.

39.2.5 Clifford modules

One can restate the previous sections in terms of Clifford modules 34.3.17, i.e. vector bundles for which the fibres carry a representation of a Clifford algebra. The content of Section 25.3 will be used here.

From Definition 25.3.7 and Example 25.3.9 it should be clear that what was called $K^0(X)$ is in fact equivalent to $K^{0,0}(\mathbf{Vect}(X))$. In a similar vein one can prove that $K^{-1}(X)$ is equivalent to $K^{1,0}(\mathbf{Vect}(X))$. This relation is in fact generalizable to all values of p,q:

Property 39.2.35. The $K^0(X)$-modules $K^{p,q}(\mathbf{Vect}(X))$ and $K^{q-p}(X)$ are isomorphic.

Property 25.3.6 then implies the Bott periodicity for the groups $K^{q-p}(X)$. For complex K-theory one can use Remark 25.3.10. 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 list see [57].

39.2.6 Cohomology theory

Property 39.2.36 (Excision). It can be shown that the excision property 39.2.11 holds at every degree $n \in \mathbb{Z}$:

$$K^n(X/Y, \{y_0\}) \cong K^n(X, Y). \tag{39.34}$$

More generally this can be stated as

$$K^n(X\backslash U, Y\backslash U) \cong K^n(X, Y), \tag{39.35}$$

where $\overline{U} \subset \mathring{Y}$.

Property 39.2.37 (Homotopy invariance). Homotopic maps induce equal morphisms in K-theory at every degree. In particular this implies that homotopy equivalences induce isomorphisms in K-theory.

Remark 39.2.38 (Generalized cohomology). The above properties imply that the complex of K-groups satisfies the Eilenberg-Steenrod axioms 8.5.1 except for the dimension axiom. As such it is a generalized cohomology theory.

39.2.7 Applications

?? REFER TO PHYSICS ??

Property 39.2.39 (Degree). Recall Definition 8.2.38 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. In fact, 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) 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.

Definition 39.2.40 (H-space). A sphere S^n is said to be an H-space if it admits a continuous function $\mu : S^n \times S^n \to S^n$ of bidegree $(1, 1)$, i.e. it admits such a function for which the pointwise maps are homotopic to the identity. (This second formulation can also be used for other topological spaces, e.g. the H-structure on loop groups 8.1.13.)

Property 39.2.41 (Puppe sequence). The mapping cone C_f from Definition 8.1.67 fits in an exact sequence:

$$X \longrightarrow Y \longrightarrow C_f \longrightarrow \Sigma X \longrightarrow \Sigma Y. \tag{39.36}$$

Reduced K-theory maps this to an induced exact sequence (whenever the reduced K-theory is defined for X, Y):

$$\widetilde{K}^0(\Sigma X) \longrightarrow \widetilde{K}^0(\Sigma Y) \longrightarrow \widetilde{K}^0(C_f) \longrightarrow \widetilde{K}^0(X) \longrightarrow \widetilde{K}^0(Y). \tag{39.37}$$

The Puppe sequence in K-theory allows to prove an important theorem in the case of functions between spheres:

Definition 39.2.42 (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]

[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.

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 39.2.43 (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 39.2.44. 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 33.6.1) one can define a morphism to ordinary cohomology:

Construction 39.2.45 (Chern character). The Chern character of vector bundles 33.6.9 induces a Chern character on K-theory:

$$\mathrm{ch} : K(X) \mapsto H^\bullet(M) : [E] - [F] \mapsto \mathrm{ch}(\nabla^E) - \mathrm{ch}(\nabla^F), \tag{39.38}$$

where ∇^E, ∇^F are any two 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 as $\nabla^E|_N = \alpha^*(\nabla^F|_N)$.

Theorem 39.2.46 (Atiyah-Hirzebruch). *Consider a space X. The Chern character induces the following isomorphisms:*

$$K^0_{\mathbb{C}}(X) \otimes \mathbb{Q} \cong \bigoplus_{i\in\mathbb{Z}} H^{2i}(X; \mathbb{Q}) \tag{39.39}$$

$$K^1_{\mathbb{C}}(X) \otimes \mathbb{Q} \cong \bigoplus_{i\in\mathbb{Z}} H^{2i+1}(X; \mathbb{Q}). \tag{39.40}$$

For noncompact spaces one needs to work with rational K-theory.

Property 39.2.47 (KK-theory). Topological K-theory and Kasparov K-theory 24.1.48 are related as follows:

$$KK(\mathbb{C}, C(X)) \cong K^0(X) \tag{39.41}$$

$$KK(C(\mathbb{R}), C(X)) \cong K^1(X). \tag{39.42}$$

39.3 Twisted K-theory

The most straightforward way to introduce twisted K-theory is similar to that of twisted de Rham theory. Namely by introducing a degree-3 integral cohomology class or, equivalently, by twisting by a U(1)-bundle gerbe.

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 PGL($\mathbb{C}$)-valued transition functions to GL($\mathbb{C}$)-valued functions, one obtains a U(1)-valued 3-cocycle $f_{ijk} := \widetilde{g}_{ij}\widetilde{g}_{jk}\widetilde{g}_{ki}$. This defines a class in $\check{H}^2(M; \mathrm{U}(1)) \cong H^3(M; \mathbb{Z})$. 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 $\mathrm{Fred}(\mathcal{H}) \hookrightarrow \mathrm{Fred}(P) \to M$.

Definition 39.3.1 (Twisted K-theory). Consider a cohomology class $\alpha \in H^3(M;\mathbb{Z})$ together with its associated bundle of Fredholm operators $\mathrm{Fred}(P) \to M$. The α-twisted K-theory on M is defined as the set of homotopy classes of sections of this bundle:

$$K^0_\alpha(M) := \Gamma(\mathrm{Fred}(P))/\sim . \tag{39.43}$$

39.4 Differential K-theory

Just as in the case of ordinary cohomology (Section 33.7), one can also generalize ordinary K-theory to a differential cohomology theory by incorporating connection data.

One of the definitions of K-theory makes use of (virtual) vector bundles. 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 39.4.1 (Simons-Sullivan mdodel). A vector bundle equipped with an equivalence class of connections, where two connections are equivalent if there 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 $\mathrm{Struct}(M)$ of isomorphism classes of structured bundles over M. This set becomes a commutative monoid under Whitney sums. The Grothendieck completion 3.2.6 of $\mathrm{Struct}(M)$ gives a model for differential K-theory:

$$\widehat{K}^0(M) := G(\mathrm{Struct}(M)). \tag{39.44}$$

39.5 Algebraic K-theory

39.5.1 Determinant

Over noncommutative rings R the determinant of a matrix is not as easily defined as over commutative rings such as field. 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 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 20.4.39).
- 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 $\mathrm{GL}(n,R)$ to its stable version 39.1.1. 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 20.4.51. It can also be shown that $E(R)$ is equal to the commutator subgroup $[\mathrm{GL}(R),\mathrm{GL}(R)]$.

The determinant map is then abstractly defined as the quotient map from the following definition:

Definition 39.5.1 (K_1). The first algebraic K-group of a ring R is defined as the Abelianization of its stable general linear group:

$$K_1(R) := \mathrm{GL}(R)/[\mathrm{GL}(R),\mathrm{GL}(R)]. \tag{39.45}$$

The quotient map $\pi : \mathrm{GL}(R) \to K_1(R)$ is called the **determinant map**.

To obtain lower K-groups a "suspension functor" needs to be defined:

Definition 39.5.2 (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:

$$\Sigma R := M(R)/M_{\text{fin}}(R). \tag{39.46}$$

Definition 39.5.3 (Lower K-groups). For all integers $n \geq 1$ one defines the K-groups as follows:

$$K_n(R) := K_{1-n}(\Sigma^n R). \tag{39.47}$$

Example 39.5.4 (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 32.2.7:

$$K_0(C(X)) \cong K^0(X). \tag{39.48}$$

39.6 Operator K-theory

Definition 39.6.1 (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 39.6.2. 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

$$\begin{pmatrix} p & 0 \\ 0 & 0 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} q & 0 \\ 0 & 0 \end{pmatrix} \tag{39.49}$$

are homotopy-equivalent.

Property 39.6.3 (Topological K-theory). If X is a compact Hausdorff space, then

$$K^0(X) \cong K_0(C(X)). \tag{39.50}$$

This isomorphism is induced by the Serre-Swan theorem 32.2.7.

Definition 39.6.4 (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 39.6.5 (Bott periodicity). First, define the suspension of a C^*-algebra A as follows:

$$\Sigma A := \{f \in C([0,1], A) \mid f(0) = f(1) = 0\}. \tag{39.51}$$

Then, $K_1(A) \cong K_0(\Sigma A)$. Moreover, one defines all higher homology groups recursively as[3]

$$K_n(A) \cong K_{n-1}(\Sigma A). \tag{39.52}$$

Property 39.6.6 (KK-theory). If A is a C^*-algebra, then operator K-theory and Kasparov K-theory 24.1.48 are related as follows:

$$KK(\mathbb{C}, A) \cong K_0(A) \tag{39.53}$$

$$KK(C(\mathbb{R}), A) \cong K_1(A). \tag{39.54}$$

Property 39.2.47 is then simply a consequence of this isomorphism by Property 39.6.3.

[3]There exists a more general definition of operator K-theory for which this relation is a property.

Chapter 40

Synthetic Differential Geometry ♣

40.1 Neighbourhoods

Definition 40.1.1 (Neighbourhood relation). A reflexive and symmetric relation $\sim$ with the additional property that the morphisms in the category under consideration preserve this relationship.

Example 40.1.2 (Monad). Let M be a set. Given a neighbourhood relation $\sim$ on M, the (first order) monad around $x \in M$ is defined as

$$\mathfrak{M}(x) := \{y \in M \mid y \sim x\}. \tag{40.1}$$

Definition 40.1.3 (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 every $i, j \leq k$.

Definition 40.1.4 (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

$$(x \approx y) \wedge (y \approx z) \wedge (x \sim z) \implies x \approx z \tag{40.2}$$

for all $x, y, z \in M$.

Definition 40.1.5 (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 40.1.6 (Frobenius' theorem). *An involutive distribution admits maximal connected integral subsets. These subsets are called the* ***leaves*** *of the distribution.*

40.2 Affine connections

Definition 40.2.1 (Affine connection). An affine connection is a map $\lambda(x, y, z)$ 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 40.1.

Remark. By looking at these diagrams the concept of parallel transport can be made a lot more intuitive than in classic differential geometry, e.g. Diagram 40.1 shows the parallel transport of the point z along xy.

$$\begin{tikzcd}
 & w \\
z \arrow[ur, dashed] & y \arrow[u, dashed] \\
x \arrow[u] \arrow[ur] &
\end{tikzcd}$$

Figure 40.1: Connection in synthetic theories.

Definition 40.2.2 (Symmetric connection). An affine connection λ is said to be symmetric or **torsion-free** if $\lambda(x, y, z) = \lambda(x, z, y)$.

Definition 40.2.3 (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 40.2.4 (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

$$\underline{\mathfrak{M}}(x_0) \to \underline{\mathfrak{M}}(x_0) : z \mapsto \text{result of parallel transporting z around } \{x_0, x_1, x_2\}.$$

Definition 40.2.5 (Geodesic). A subset $S \subseteq M$ stable under the affine connection λ.

40.3 Euclidean geometry

40.3.1 Infinitesimal elements

Definition 40.3.1 (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:

$$\Delta := \{x \in R \mid x^2 = 0\}. \tag{40.3}$$

A neighbourhood relation on R is induced by setting $\underline{\mathfrak{M}}(0) = \Delta$.

Remark 40.3.2. If the Euclidean doctrine would be followed where $R \equiv \mathbb{R}$, this set would be $\{0\}$. However, by not requiring R to be a field, a larger set is obtained.

Axiom 40.1 (Kock-Lawvere). For every map $f : \Delta \to R$ there exists a unique element $b \in R$, called the **slope** of f, such that

$$f(d) = f(0) + d \cdot b \tag{40.4}$$

for all $d \in \Delta$.

Corollary 40.3.3. 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 40.3.4. Let $a, b \in R$. If $d \cdot a = d \cdot b$ for all $d \in \Delta$, then $a = b$.

Notation 40.3.5. Analogous to the infinitesimal line one can define the sets $\mathbb{D}^k$ in the following way:

$$\mathbb{D}^k := \{x \in R \mid x^{k+1} = 0\}. \tag{40.5}$$

With this notation one has $\Delta \equiv \mathbb{D}^1$.

[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.

40.3.2 Calculus

Formula 40.3.6 (Taylor expansion). The Kock-Lawvere axiom implies the following exact Taylor expansion:

$$f(x+d) = f(x) + d \cdot f'(x), \tag{40.6}$$

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 40.3.7. The Kock-Lawvere axiom also 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 40.1 is generalized:

Axiom 40.2. For every $g : \mathbb{D}^k \to R$ there exist unique elements $\{b_1, \ldots, b_k\}$ in R such that

$$f(d) = f(0) + \sum_j d^j \cdot b_j \tag{40.7}$$

for all $d \in \mathbb{D}^k$.

Corollary 40.3.8. Let $f : R \to R$ and $d \in D_k$.

$$f(x+d) = f(x) + d \cdot f'(x) + \cdots + \frac{d^k}{k!} \cdot f^{(k)}(x) \tag{40.8}$$

Chapter 41

Noncommutative Geometry

This chapter heavily uses the concepts introduced in Chapter 26.

41.1 Quantum geometry

Definition 41.1.1 (Quantum metric). Consider an FODC (A, Γ) as in Definition 26.4.1. A generalized inner-product is a bimodule morphism $\langle\cdot|\cdot\rangle : \Gamma \otimes \Gamma \to A$. A (generalized) metric with respect to an inner product $\langle\cdot|\cdot\rangle$ is an element $g \in \Gamma \otimes \Gamma$ such that

$$\langle\omega|\cdot\rangle \otimes \mathbb{1}(g) = \omega = \mathbb{1} \otimes \langle\cdot|\omega\rangle(g) \tag{41.1}$$

for all $\omega \in \Gamma$. This condition represents the invertibility of the metric.

Definition 41.1.2 (Connection). Let $(\Omega^\bullet, d)$ be a differential calculus on an associative algebra A. A connection on a left A-module Γ is a linear map

$$\nabla : \Gamma \to \Omega^1 \otimes_A \Gamma \tag{41.2}$$

such that

$$\nabla(a\omega) = da \otimes_A \omega + a\nabla\omega. \tag{41.3}$$

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:

$$T := d - \pi \circ \nabla, \tag{41.4}$$

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 41.1.3 (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

$$\nabla = p \circ d, \tag{41.5}$$

where $p : A \to P$ is the canonical projection.

Example 41.1.4. Let A be a C^*-algebra 24.1.2. By the Gel'fand-Naimark theorem 24.1.41, 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 26.4.7:

$$\widetilde{\rho} : \Omega^\bullet(A) \to \mathcal{B}(\mathcal{H}) : a_0 da_1 \otimes \cdots \otimes da_n \mapsto \rho(a_0)[D, \rho(a_1)] \cdots [D, \rho(a_n)]. \tag{41.6}$$

The problem here is that $\widetilde{\rho}(a) = 0$ need not imply that $\widetilde{\rho}(da) = 0$. One can however obtain a smaller differential calculus by modding out the differential ideal generated by the kernel of this action: $\widetilde{\Omega}^\bullet(A) := \Omega^\bullet(A)/\{\ker(\widetilde{\rho}) + d\ker(\widetilde{\rho})\}$.

Connes proved that if $A = C^\infty(M)$ for some Riemannian manifold M and D is 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 41.1.5. Recall Property 41.1.3 above. If P is finite-rank and projective, there exists a canonical connection $p \circ d$ 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:

$$\nabla := \nabla_0 + \alpha. \tag{41.7}$$

The morphism α is often called a (noncommutative) **gauge field**.

41.2 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 41.2.1 (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**[1]: There exists a grading γ, turning $\mathcal{H}$ into a super-Hilbert space, such that

$$[\gamma, A] = 0 \qquad \{\gamma, D\}_+ = 0. \tag{41.8}$$

Remark 41.2.2 (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]One sometimes makes a difference between **even** and **odd** spectral triples based on the validity of this axiom.

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 23.4.63 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

$$\delta : T \mapsto [|D|, T] \tag{41.9}$$

and consider the spaces

$$\mathcal{H}^\infty := \bigcap_{i=1}^{\infty} \mathrm{dom}(D^i) \qquad \mathcal{B}^\infty(\mathcal{H}) := \bigcap_{i=1}^{\infty} \mathrm{dom}(\delta^i). \tag{41.10}$$

The spectral triple is said to be strongly regular of A, $[D, A]$ and $\mathrm{End}(\mathcal{H}^\infty)$ 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 3.5.1. By regularity there exists an induced representation

$$\pi_H : HC_n(A) \to \mathcal{B}(\mathcal{H}) : a_0 \otimes \cdots \otimes a_n \mapsto a_0[D, a_1] \cdots [D, a_n]. \tag{41.11}$$

The spectral triple is said to be orientable if there exists an antisymmetric Hochschild p-cycle χ, where p 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.)

5. **Charge conjugation**: There exists an antiunitary operator $C \in \mathrm{End}(\mathcal{H})$ such that

$$C^2 = \pm 1 \qquad CD = \pm DC \qquad C\gamma = \pm \gamma C, \tag{41.12}$$

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):[2]

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 $a \in A$.

$$\langle v|w\rangle_{\mathcal{H}} = \int \langle v|w\rangle_A, \tag{41.13}$$

[2] This is related to the notion of *KO*-dimension in (real) K-theory.

where

$$\int : A \to \mathbb{C} : a \mapsto \mathrm{tr}_{\mathfrak{D}}(a|D|^{-p}) \tag{41.14}$$

is the "noncommutative integral" induced by the Dixmier trace 23.4.63 ($a|D|^{-p}$ is measurable for all $a \in A$) and

$$\langle v|w\rangle_A := \sum_{i,j} v_i^* p_{ij} w_j \tag{41.15}$$

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 41.2.3 (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 34.3.19 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 $C^\infty(M)$.
- 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 32.2.7, $\mathcal{H}^\infty$ 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 41.2.4 (Hodge theory). Another example of a spectral triple, where the underlying manigfold does not have to be spin, is given by Hodge theory. On every Riemannian manifold there exists a first-order operator

$$D := d + \delta. \tag{41.16}$$

This is the Hodge-de Rham operator 34.1.5.

Construction 41.2.5 (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 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

$$\widehat{x}_i := (\widehat{x}_i^1, \dots, \widehat{x}_i^p) : U_i \to \mathbb{R}^p, \tag{41.17}$$

where $\widehat{x}$ 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 $\widehat{x}_i \circ \chi \circ \tau : \mathbb{R}^p \to \mathbb{R}^p$ is a local diffeomorphism.
- If $\{a_1, \ldots, a_n\} \subset A$ 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 $\{a_1, \ldots, a_n\}$, under the Gel'fand isomorphism is again an element of A. It will be denoted by $f(a_1, \ldots, a_n)$.
- 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 41.2.6 (Distance). Let $(A, \mathcal{H}, D)$ be a spectral triple such that

$$\{a \in A\backslash\mathbb{C} \mid \|[D, a]\| \leq 1\} \tag{41.18}$$

is norm-bounded in $\overline{A}\backslash\mathbb{C}$. Then

$$d(\phi, \psi) := \sup\{|\phi(a) - \psi(a)| \mid a \in A : \|[D, a]\| \leq 1\} \tag{41.19}$$

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 41.2.7. For the canonical spectral triple on a compact spin manifold, the above distance coincides with the geodesic distance.

Property 41.2.8 (Fredholm modules). When $(A, \mathcal{H}, D)$ is a spectral triple, the operator $D|D|^{-1}$ satisfies the properties of a Fredholm module 24.1.45. This module is even (resp. odd) when the spectral triple is even (resp. odd).

41.3 Dunkl operators

Let G be a Coxeter group 30.4.48 with root system Φ over the space $\mathbb{R}^n$. A function $m : \Phi \to \mathbb{C}$ is called a **multiplicity (function)** if $m(\alpha) = m(\beta)$ whenever σ_α and σ_β are conjugate elements in G.

Definition 41.3.1 (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:

$$T_i f(x) := \partial_i f(x) + \sum_{\alpha\in\Phi^+} m(\alpha)\frac{f(x) - f(\sigma_\alpha x)}{\langle\alpha|x\rangle}\langle\alpha|e_i\rangle, \tag{41.20}$$

where e_i is the i^{th} basis element of $\mathbb{R}^n$.

Property 41.3.2 (Commutativity). For all $i, j \leq n$ the following property holds:

$$T_iT_jf = T_jT_if. \tag{41.21}$$

Chapter 42

Higher-dimensional Geometry ♣

The main references are [79, 80]. The section on spectral geometry is based on the thesis [110], while that on smooth spaces is inspired by [75]. For an introduction to (higher) category theory, see Chapter 4. Section 27.7.2 gives a different approach to the higher-dimensional analogues of Lie algebras.

In this chapter certain constructions and theorems introduced in the previous chapters are generalized to the setting of higher categories. As such it can been as an analogue to Chapter 27 for (differential) geometry.

?? CITE BAEZ, SCHREIBER, BARTELS, ... ??

42.1 Infinite-dimensional geometry

In many situations, when consider function spaces, the objects under consideration do not form a finite-dimensional manifold. However, with some care, one can drop this size condition. In Chapter 23 it was shown how one can extend calculus from $\mathbb{R}^n$ to infinite-dimensional vector spaces.

The first approach uses a locally convex TVS 23.3.8 as local model space:

Definition 42.1.1 (E-manifold). Let E be a locally convex TVS. A Hausdorff topological space is called an E-manifold if exists an atlas of charts (U, φ), where $\varphi : U \to \varphi(U) \subset E$ is a homeomorphism and the transition maps are smooth in the sense of Gateaux.

Definition 42.1.2 (Kinematic tangent bundle). Let M be an E-manifold with a smooth atlas $\{U_i, \varphi\}_{i\in I}$. The kinematic tangent bundle of M is defined as the quotient of

$$\bigsqcup_{i\in I} U_i \times E \tag{42.1}$$

by the equivalence relations $(x, w) \sim (x, d\psi_{ji}(\varphi_i(x); v))$.

Remark 42.1.3. 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 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 ccok 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 definition is the most widely used one in the finite-dimensional setting:

Definition 42.1.4 (Kinematical 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 $L^k_{\text{alt}}(TM, M \times \mathbb{R})$, the kinematical cotangent bundle.

42.2 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 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

$$C(X \times Y, Z) \to C(X, C(Y, Z)).$$

However, in general, this is not a homeomorphism, i.e. currying does not define an adjunction and, therefore, **Top** is not Cartesian closed 4.6.21. 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.[1]

Whilst studying varieties in algebraic geometry people experienced similar problems. For this reason *Grothendieck* invented schemes (see Chapter 11 and Section 11.3 in particular). The main takeaway of this approach was that it is often better to work with a well-behaved category containing some ‘‘nasty’’ objects, than to work with a ‘‘nasty’’ category containing only nice objects.

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. 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.

42.2.1 Concrete sites

Definition 42.2.1 (Diffeological space). Let X be a set. A diffeology $\mathcal{D}$ 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. If f is constant, then $f \in \mathcal{D}$. Equivalently, every function $f : \mathbb{R}^0 \to X$ is a plot.
2. If $\{U_i\}_{i\in I}$ is an open cover of U and if $f|_{U_i} \in \mathcal{D}$ for all $i \in I$, then $f \in \mathcal{D}$.
3. If $f \in \mathcal{D}$ and $g : W \subseteq \mathbb{R}^m \to \text{dom}(f)$ is smooth, then $f \circ g \in \mathcal{D}$.

The set X can be turned into a topological space by equipping it with the **$\mathcal{D}$-topology**, i.e. the final topology with respect to $\mathcal{D}$.

Remark 42.2.2. The domain of different plots can be subsets of different Euclidean spaces $\mathbb{R}^m$ and $\mathbb{R}^n$.

[1] This is in general not a problem since all interesting spaces, such as CW complexes, belong to this class.

Definition 42.2.3 (Smooth map). Let $(X, \mathcal{D})$ and $(Y, \mathcal{D}')$ be diffeological spaces. A map $g : X \to Y$ is said to be smooth if for every $f \in \mathcal{D}$ the composition $g \circ f \in \mathcal{D}'$.

The diffeological spaces together with their differentiable morphisms form a category **DiffSp**.

Definition 42.2.4 (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* are obtained.

Alternative Definition 42.2.5 (Manifold). A diffeological space is called an n-manifold if it 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 29.1.12.

There exist two trivial smooth structures:

Example 42.2.6 (Discrete structure). The smooth structure defined by taking the plots to be the constant functions.

Example 42.2.7 (Indiscrete structure). The smooth structure obtained by taking all functions to be plots.

Definition 42.2.8 (Smooth set). By omitting the reference to an underlying set in the definition of smooth spaces above, a more general definition can be obtained. This way the category **SmoothSet** is obtained as the sheaf category on the site of Cartesian spaces $\mathbf{Sh}(\mathbf{CartSp}_{\mathbf{diff}})$. The topology on this site is generated by the coverage of differentiably good covers 29.1.16 (in fact, this topology coincides with the usual one consisting of open covers). Diffeological spaces can be recovered by passing to the full subcategory on *concrete sheafs*. The category of smooth spaces/sets is often denoted by $\mathbf{C}^\infty$.

?? ADD INFORMATION ON CONCRETE SHEAFS ??

Example 42.2.9 (Differential forms). Consider the k^{th} de Rham functor Ω^k on the category **Diff**. This functor assigns to every smooth manifold its space of differential k-forms (Section 32.4). Locally defined forms can be glued together if they agree on intersections and, hence, they satisfy the sheaf condition. This shows that Ω^k is a smooth space, albeit one that is far from an ordinary smooth manifold.

One can go even further. Consider the subfunctor Ω^2_{cl} that assigns closed two-forms to a smooth manifold. This also defines a smooth space and, hence, one can consider the slice category $\mathbf{C}^\infty/\Omega^2_{\text{cl}}$. It is not hard to show that the category **SpMfd** of symplectic manifolds admits an embedding into this slice category.

Property 42.2.10. There exists an adjunction

$$\mathbf{Top} \underset{diff}{\overset{top}{\underset{\longrightarrow}{\overset{\longleftarrow}{\perp}}}} \mathbf{C}^\infty. \tag{42.2}$$

The functor *diff* endows a topological space X with the smooth structure for which every continuous map $U \to X$ is a plot. The adjoint functor *top* sends a smooth space to the topological space equipped with the finest topology for which all plots become continuous maps.

Definition 42.2.11 (Smooth algebra). For any smooth manifold M the algebra of smooth functions can be obtained as a hom-object:

$$C^\infty(M) \equiv C^\infty(M, \mathbb{R}) = \mathbf{Diff}(M, \mathbb{R}).$$

Since hom-functors are (finite) product-preserving, one can see that the multiplication $C^\infty(M) \times C^\infty(M) \to C^\infty(M)$ is induced by the multiplication on $\mathbb{R}$:

$$C^\infty(M, \mathbb{R} \times \mathbb{R}) \cong C^\infty(M) \times C^\infty(M).$$

Furthermore, the hom-functor is covariant in the second argument and, hence, a copresheaf on the category $\mathbf{CartSp_{diff}}$ of Euclidean (Cartesian) spaces and smooth morphisms is obtained. Generalizing this situation, smooth algebras are defined as finite product-preserving copresheaves on $\mathbf{CartSp_{diff}}$. This (functor) category is denoted by $\mathbf{C^\infty Alg}$.

Given a smooth algebra R, its **underlying algebra** $U(R)$ is defined as the set $R(\mathbb{R})$ equipped with the canonically induced ring operations.

Definition 42.2.12 (Finitely generated smooth algebra). Since ordinary R-algebras are finitely generated if and only if they are of the form $R[x_1, \ldots, x_k]/J$ for some integer $k \in \mathbb{N}$ and some ideal J, a smooth algebra is said to be finitely generated if it is of the form $C^\infty(\mathbb{R}^n)/J$ for some $n \in \mathbb{N}$ and some ideal J in the ordinary ring underlying the smooth algebra.

Definition 42.2.13 (Smooth locus). Let $\mathbf{C^\infty Alg^{fin}}$ denote the category of finitely generated smooth algebras. The category of **smooth loci** is defined as $(\mathbf{C^\infty Alg^{fin}})^{op}$. The smooth locus corresponding to a smooth algebra R is often denoted by ℓR.

42.2.2 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.

Definition 42.2.14 (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 40.1). The straightforward definition is the following one:

$$\mathbb{D} := \mathrm{Spec}(A), \tag{42.3}$$

where $A := \mathbb{R} \oplus V$ with V a finite-dimensional nilpotent ideal. 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 R-algebra homomorphism between their assciated algebras. These form the category $\mathbf{FormalCartSp_{diff}}$.

Example 42.2.15 (First-order neighbourhood). By taking $A = \mathbb{R}[\varepsilon]/\varepsilon^2$ one exactly obtains the first-order infinitesimal neighbourhood of Definition 40.3.1. The morphism dual to the mapping implied by the Kock-Lawvere axiom 40.1 gives an inclusion map $\mathbb{D}^1 \hookrightarrow \mathbb{R}^1$. (This example can easily be generalized to k^{th}-order neighbourhoods.)

Property 42.2.16 (Morphisms). First, consider the morphisms from a Euclidean space into an infinitesimal neighbourhood $\mathbb{D}^k$. Since such morphisms are dual to algebra homomorphisms, one should look at homomorphisms of the form $\mathbb{R}[\varepsilon]/\varepsilon^{k+1} \to C^\infty(\mathbb{R}^n)$. However, being an algebra homomorphism 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 $\mathbb{R}^n \times \mathbb{D}^1 \to \mathbb{R}^n$ are in bijection with vector fields on $\mathbb{R}^n$.

Definition 42.2.17 (Formal smooth set). A sheaf on the site of infinitesimally thickened Euclidean spaces (covers are of the form $\{U_i \times \mathrm{Spec}(A) \mid U_i \hookrightarrow \mathbb{R}^n\}$). The category of formal smooth sets or, equivalently, the sheaf topos on $\mathbf{FormalCartSp_{diff}}$ is also called the **Cahiers topos**. The sets 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 spaces).

Definition 42.2.18 (Reduction). Given an infinitesimally thickened space $\mathbb{R}^n \times \mathbb{D}$, its reduction $\mathfrak{R}$ is defined to be $\mathbb{R}^n$. Every reduction induces a canonical morphism $\mathbb{R}^n \hookrightarrow \mathbb{R}^n \times \mathbb{D}$. Plots get can be reduced by ''precomposing'' with a reduction morphism.

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 42.2.19 (Shape modality). The infinitesimal shape or **de Rham shape** $\Im X$ of a formal smooth set X is defined as the formal smooth set obtained by reducing the plots of X:

$$\Im X(U) := X(\mathfrak{R}(U)). \tag{42.4}$$

For its incarnation as a modal operator, see Section 6.5.

In analogy to Definition 7.2.12 one can also define local diffeomorphisms between formal smooth sets:

Definition 42.2.20 (Local diffeomorphism[2]). A morphism of formal smooth sets $f : X \to Y$ such that the thickened plots of X can be identified with those of Y whose reduction comes from a Euclidean plot of X. More elegantly (or abstractly) this means that the naturality square of the shape modality (interpreted as a monad) forms a pullback square:

$$\begin{tikzcd}
X \arrow[r, "\eta_X"] \arrow[d, "f"'] \arrow[dr, phantom, "\mathrm{pb}"] & \Im X \arrow[d, "\Im f"] \\
Y \arrow[r, "\eta_Y"'] & \Im Y
\end{tikzcd}$$

Alternative Definition 42.2.21 (Smooth manifold). A diffeological space (in its incarnation as a smooth formal 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.

Although this section started with the promise that spaces would be generalized to the fermionic setting, only bosonic spaces were constructed up to this point. However, everything introduced in this section was formulated in such a way that supergeometry can be included through a minor modification:

Definition 42.2.22 (Superpoint). A space of the form $\mathrm{Spec}(A)$ for $A := \mathbb{R} \oplus V$ with V a finite-dimensional superalgebra 27.1.6 that forms a nilpotent ideal of A. When A is taken to be the Grassmann algebra 21.4.20 on n generators, the odd point $\mathbb{R}^{0|n}$ is obtained. 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 $C^\infty(\mathbb{R}^m \times \Pi\mathbb{R}^m)$.

Definition 42.2.23 (Super smooth set). A sheaf on the category of super Euclidean spaces $\mathbf{SuperCartSp_{diff}}$.

[2] Also called a **(formally) étale morphism**.

42.2.3 Graded manifolds

In this section some of the notions from Part V 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 42.2.24 (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 $C^\infty(M) \otimes \Lambda^\bullet\mathbb{R}^n$ for some $n \in \mathbb{N}$. More generally, a **graded manifold** is a locally ringed space that is locally isomorphic to $(\mathbb{R}^m, C^\infty(\mathbb{R}^m) \otimes \mathrm{Sym}(V^*))$ for a graded vector space V. (A supermanifold can be recovered by taking $V = \Pi\mathbb{R}^n$.)

Theorem 42.2.25 (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 $\Gamma(\Lambda^\bullet E)$, i.e. $\mathcal{A}$ is locally given by $\mathrm{Sym}(\Lambda^\bullet E^*)$. If $(M, \mathcal{A})$ is a supermanifold, there exists a vector bundle $E \to N$ such that $\mathcal{A}$ is locally given by $\Lambda^\bullet E^*$.*

Definition 42.2.26 (Vector field). A graded vector field of degree k is a degree-k derivation on $C^\infty(M)$. The integer k is called the **degree**.

Definition 42.2.27 (Cohomological vector field). A graded vector field X of degree 1 that satisfies $[X, X] = 0$. Every degree-1 graded vector field satisfies

$$[X, X] = 2X \circ X, \tag{42.5}$$

which 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 42.2.28 (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 defined by

$$Q := \sum_{i=1}^n dx^i \partial_i. \tag{42.6}$$

Note that the differentials dx^i are here regarded as coordinate functions on ΠTM. The **degree** of a homogeneous element of $\Omega^\bullet(M)$ is defined as the difference of its graded degree and its form degree. The de Rham complex itself then corresponds to the algebra $C^\infty(\Pi TM)$.

Definition 42.2.29 (Poisson manifold). Consider a degree-k symplectic form ω. This form induces a Poisson structure on the algebra $C^\infty(M)$ as follows:

$$\{f, g\} := (\partial_i^R f)\omega^{ij}(\partial_j^L g). \tag{42.7}$$

It is not hard to check that this operation is graded-commutative. As in Section 35.2, a Hamiltonian vector field can be defined for any smooth function $H \in C^\infty(M)$:

$$\omega(X_H, \cdot) = -dH(\cdot). \tag{42.8}$$

Property 42.2.30 (Euler vector field). Consider the graded vector field

$$E := \sum_{i=1}^n \deg(x^i)x^i\partial_i. \tag{42.9}$$

The Lie derivative $\mathcal{L}_E$, defined through the Cartan formula

$$\mathcal{L}_E := \iota_E d + (-1)^{\deg(E)} d\iota_E, \tag{42.10}$$

acts on homogeneous forms by multiplication by their degree.

Property 42.2.31. 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 42.2.32 (dg-symplectic manifold). Consider a Hamiltonian cohomological vector field X. There exists a Hamiltonian function H such that

$$Xf = \{H, f\} \tag{42.11}$$

for all $f \in C^\infty(M)$. If the symplectic form has degree k, 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

$$\{H, H\} = 0 \tag{42.12}$$

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 and the Hamiltonian is mapped to a Poisson bivector field exactly if it satisfies the master equation.

42.2.4 Gauge theory

Recall the notions of Chapter 13, in particular the notions of stacks and higher topoi. The $(\infty, 1)$-category of smooth ∞-stacks can be described in terms of (left Bousfield) localization of a suitable presheaf category by Lurie's theorem 12.7.1.

The first possibility is the category of ∞-presheaves on **Diff** with the localization at open covers. While, the second possibility is the dense subsite $\mathbf{CartSp}_{\text{diff}}$ with localization at good open covers. Both will result in a Čech model structure 13.6.3. However, the exact properties will differ.

Example 42.2.33 (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 just presented[3] 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 $\mathbf{CartSp}_{\text{diff}}$, the presheafs 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 presheaf

$$U \mapsto N_\Delta(*/\!/C^\infty(U, G)). \tag{42.13}$$

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.

42.3 Higher geometry

In this section some notions about groups, Lie groups and groupoids (Sections 3.2, 30.1 and 4.10) are extended the setting of higher category theory.

[3]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.

42.3.1 Groups

Definition 42.3.1 (Lie groupoid[4]). A groupoid internal to **Diff**.

Note that Definition 4.5.2 requires the existence of pullbacks. In the category **Diff** this is equivalent to assuming that the source and target morphisms are (surjective) submersions.

Remark 42.3.2. In the Ehresmannian approach one gives the manifold of composable morphisms $D_1 \times_{D_0} D_1$ as part of the data. Hence, no further assumptions have to be made about the source and target morphisms.

Definition 42.3.3 (Lie algebroid). A vector bundle $\pi : E \to M$ together with a vector bundle 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:

$$[X, fY] = f[X, Y] + \rho(X)(f)Y. \tag{42.14}$$

This property also implies that ρ preserves the Lie bracket:

$$\rho([X, Y]) = [\rho(X), \rho(Y)]. \tag{42.15}$$

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:

$$\rho(s_\alpha) = R^i_\alpha \partial_i \tag{42.16}$$
$$[s_\alpha, s_\beta] = C^\gamma_{\alpha\beta}(x) s_\gamma. \tag{42.17}$$

The Lie algebroid properties then imply the following conditions on these structure functions:

$$R^j_\alpha \frac{\partial R^i_\beta}{\partial x^j} - R^j_\beta \frac{\partial R^i_\alpha}{\partial x^j} = R^i_\gamma C^\gamma_{\alpha\beta} \tag{42.18}$$

and

$$R^i_\alpha \frac{\partial C^\kappa_{\beta\gamma}}{\partial x^i} + C^\kappa_{\alpha\mu} C^\mu_{\beta\gamma} + (\alpha \leftrightarrow \beta \leftrightarrow \gamma) = 0. \tag{42.19}$$

Example 42.3.4 (Tangent Lie algebroid). The tangent bundle over a smooth manifold is a Lie algebroid with $\rho \equiv \mathbb{1}_{TM}$.

Consider the **pair groupoid $\mathbf{M} \times \mathbf{M}$**, i.e. the groupoid with the following data:

- **Objects**: M
- **Morphisms**: $M \times M$, i.e. between every two points there exists a unique morphism.

Both the fundamental groupoid $\mathbf{\Pi_1}(M)$ (Definition 8.1.17) and the pair groupoid $\mathbf{M} \times \mathbf{M}$ integrate the tangent Lie algebroid.

One can generalize the dual construction of L_∞-algebras 27.7.15 even further:

Definition 42.3.5 (L_∞-algebroid). Consider the construction of the Chevalley-Eilenberg algebra for a 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 42.3.6. L_∞-algebras can be recovered by considering the special case $M = \{*\}$.

[4] In a similar way one could define *topological groupoids, étalé groupoids*, ...

Example 42.3.7 (de Rham complex). Consider the tangent algebroid of a smooth manifold M. The associated Chevalley-Eilenberg complex is equivalent to the de Rham complex $\Omega^\bullet(M)$.

Definition 42.3.8 (Weak 2-group). Let $(\mathbf{C}, \otimes, \mathbf{1})$ be a monoidal category. This category is called a weak 2-group, **categorical group** or **gr-category** 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 4.9.8 one can equivalently define a weak 2-group as a 2-category with a single object, weakly invertible 1-morphisms and invertible 2-morphisms.

Definition 42.3.9 (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 42.3.10 (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 42.3.11 (∞-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 42.3.12 (Lie crossed modules). The 2-category of (strict) 2-groups is biequivalent to the 2-category of (Lie) crossed modules 3.3.17. Given a 2-group $\mathcal{G}$, a crossed module is obtained as follows:

- $G := \mathrm{ob}(\mathcal{G})$,
- $H := \{h \in \mathrm{hom}(\mathcal{G}) \mid \mathfrak{s}(f) = e\}$,
- $t(h) := \mathfrak{t}(h)$, 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 $\partial : \mathfrak{h} \rightarrow \mathfrak{g}$, and
- a Lie algebra morphism $\rho : \mathfrak{g} \rightarrow \mathrm{Der}(\mathfrak{h})$.

The equivariance and Peiffer conditions induce similar conditions for the above data:

- $\partial(\rho(h)g) = [h, \partial g]$, and
- $\rho(\partial h)(h') = [h, h']$,

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 42.3.13 (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 $\mathrm{AUT}(H)$ gives a 2-group of symmetries of H, i.e. it is the automorphism 2-group of H in the 2-category **Lie**.

Definition 42.3.14 (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

$$\frac{d}{dt}g(t) = f(t)g(t) \tag{42.20}$$

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 42.3.15 (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 27.7.13. As shown in [77] only $\mathfrak{g}_0$ gives rise to a Lie 2-group (or even a topological 2-group).

42.3.2 Spaces

To overcome the problem encountered in Definition 42.3.1 above, one should pass from **Diff** to $\mathbf{C}^\infty$. It can be shown that this category admits all pullbacks, quotients, path spaces, etc.

Definition 42.3.16 (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 general sense. The notions of 2-groups as introduced in the previous section are easily generalized to this more general setting.

Definition 42.3.17 (2-group action). Consider a smooth 2-group $\mathcal{G}$ and a smooth 2-space E. A strict action of $\mathcal{G}$ on E is a smooth homomorphism $\mathcal{G} \to \mathrm{AUT}(E)$, i.e. a smooth map preserving products and inverses.

Definition 42.3.18 (Thin homotopy). Let M be a smooth manifold. A smooth homotopy $H : [0,1]^2 \to M$ is said to be thin if

$$H(s,t) = F(s) \tag{42.21}$$

for some smooth F near $t = 0, 1$ and if it pulls back every two-form to 0:

$$\forall\omega \in \Omega^2(M) : H^*\omega = 0. \tag{42.22}$$

Definition 42.3.19 (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 42.3.20 (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
- **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 full 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 42.3.21. 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.

?? COMPLETE ??

42.4 2-Bundles

A first step is the generalization of the categorical definition of a general bundle 31.1.1, i.e. as an object of a slice category:

Definition 42.4.1 (Smooth 2-bundle). A triple (E, B, π) where both E and B are smooth 2-spaces and π is a smooth map.

Definition 42.4.2 (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 31.1.5, 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 $\varphi_i\varphi_j^{-1} = U_{ij} \times F \cong U_{ij} \times F$ as usual. By the commutative diagram above, these transition maps only act on the fibre F. Because $\varphi_i\varphi_j^{-1}$ is itself an (auto)equivalence, the action on F is given by a functor $g_{ij} : U_{ij} \to \mathrm{AUT}(F)$, where the 2-space $\mathrm{AUT}(F)$ is the *coherent 2-group*[5] of autoequivalences of F together with invertible 2-maps between them.

The interesting (and important) part is how the cocycle conditions (31.1) and 31.1.4 for the maps g_{ij} are modified. Since the equivalences g_{ij} are only invertible up to 2-maps, one cannot expect these conditions to hold as equations. Instead, two higher transition maps (i.e. natural isomoprhisms) $h_{ijk} : g_{ij} \circ g_{jk} \Rightarrow g_{ik}$ 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 27.8.1).

Definition 42.4.3 ($\mathcal{G}$-bundle). A locally trivial 2-bundle with typical fibre F is said to have the 2-group $\mathcal{G}$ as its structure (2-)group if the transition data factor through an action $\mathcal{G} \to \mathrm{AUT}(F)$. If $F = \mathcal{G}$, the 2-bundle is called a **principal $\mathcal{G}$-2-bundle**.

Remark 42.4.4 (Gerbes). If the transition maps k_i are chosen to be trivial and $\mathcal{G}$ is chosen to be respectively the trivial Lie 2-group associated to an Abelian Lie group G or the automorphism

[5]Instead of the strict invertibility of maps in the definition of 2-groups above, one should allow for invertibility up to 2-isomorphisms that themselves satisfy certain coherence conditions.

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 33.3.36 of principal connections, one can define connections for principal n-bundles:

Definition 42.4.5 (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 $U_i \subset M$ a local holonomy n-functor

$$\mathrm{hol}_i : \mathcal{P}_n(U_i) \to G; \tag{42.23}$$

- for every double intersection U_{ij} a 1-transfor (i.e. an n-natural transformation)

$$g_{ij} : \mathrm{hol}_i \Rightarrow \mathrm{hol}_j; \tag{42.24}$$

- for every triple intersection U_{ijk} a 2-transfor

$$f_{ijk} : g_{ij} \circ g_{jk} \Rrightarrow g_{ik}; \tag{42.25}$$

- and so on...

This is equivalently given by a global n-functor

$$\mathrm{hol} : \mathcal{P}_n(M) \to \mathbf{Trans}_n(P). \tag{42.26}$$

?? ADD GERBES (e.g. BRYLINSKI) ??

42.5 Space and quantity

In this section the general notions of spaces and observables are again considered. From the start everyhting will be formulated in an enriched setting where $\mathcal{V}$ is a cosmos 4.7.1. The categories $\mathbf{C}$ of interest will be assumed to be small.

In the previous sections spaces modelled on a base space X, or more generally, on a category of spaces $\mathbf{S}$ were modelled as (concrete) sheaves on a suitable site. Here this notion is relaxed as much as possible:

Definition 42.5.1 (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 42.5.2 (Quantity). A (generalized[6]) quantity on a category $\mathbf{C}$ is a copresheaf on $\mathbf{C}$.

[6]It is generalized because it is "measures" a category instead of a single object.

Property 42.5.3 (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:

$$\mathcal{O} : \mathbf{Psh}(\mathbf{C}) \to \mathbf{coPsh}^{op}(\mathbf{C}) : X \mapsto \mathrm{Hom}_{\mathbf{Psh}(\mathbf{C})}(X, \mathcal{Y}-). \tag{42.27}$$

Similarly, given a quantity Q one can ask on which space it behaves as the algebra of functions. This also defines as functor:

$$\mathrm{Spec} : \mathbf{coPsh}^{op}(\mathbf{C}) \to \mathbf{Psh}(\mathbf{C}) : Q \mapsto \mathrm{Hom}_{\mathbf{coPsh}(\mathbf{C})}(\mathcal{Y}^{op}-, Q), \tag{42.28}$$

where $\mathcal{Y}^{op}$ denotes the co-Yoneda embedding $\mathbf{C} \to [\mathbf{C}, \mathcal{V}]^{op} : c \mapsto \mathbf{C}(c, -)$.

The incredible result is now that $(\mathcal{O} \dashv \mathrm{Spec})$ is an adjunction, called the **Isbell adjunction**. Objects that are preserved (up to isomorphism) under the associated (co)monad are said to be **Isbell self-dual**.

Example 42.5.4 (Cartesian spaces). When working over the site **CartSp** (with its usual topology) and restricting to coherent sheaves and product-preserving presheaves, the Isbell adjunction maps spaces to smooth algebras.

Part VI

Probability Theory & Statistics

Chapter 43

Probability

The majority of this chapter uses the language of measure theory. For an introduction see Chapter 16. The section on *imprecise probabilities* is mainly based on [13].

43.1 Probability

The Kolmogorov axioms of probability state when a set admits the definition of a probability theory:

Definition 43.1.1 (Kolmogorov axioms). A probability space (Ω, Σ, P) is a measure space 16.1.3 with finite measure $P(X) = 1$. The set Ω is called the **sample space**.

Definition 43.1.2 (Random variable). Let (Ω, Σ, P) be a probability space. A function $X : \Omega \to \mathbb{R}$ is called a random variable if $\forall a \in \mathbb{R} : X^{-1}\big([a, \infty[\big) = \{\omega \in \Omega \mid X(\omega) \geq a\} \in \Sigma$.

Definition 43.1.3 (σ-algebra of a random variable). Let X be a random variable defined on a probability space (Ω, Σ, P) and denote the Borel σ-algebra of $\mathbb{R}$ by $\mathcal{B}$. The following family of sets is a σ-algebra:

$$X^{-1}(\mathcal{B}) := \{S \in \Sigma \mid \exists B \in \mathcal{B} : S = X^{-1}(B)\}. \tag{43.1}$$

Notation 43.1.4. The σ-algebra generated by the random variable X is often denoted by $\mathcal{F}_X$, analogous to 2.5.7.

Definition 43.1.5 (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 a 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 2.5.2 of events instead of the power set 2.4.1 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.

Formula 43.1.6 (Union). Let A, B be two events. The probability that at least one of them occurs is given by the following formula:

$$P(A \cup B) = P(A) + P(B) + P(A \cap B). \tag{43.2}$$

Definition 43.1.7 (Disjoint events). Two events A and B are said to be disjoint if they cannot happen at the same time:

$$P(A \cap B) = 0. \tag{43.3}$$

Corollary 43.1.8. If A and B are disjoint, the probability that both A and B occur is just the sum of their individual probabilities.

Formula 43.1.9 (Complement). Let A be an event. The probability of A being false is denoted as $P(\overline{A})$ and is given by

$$P(\overline{A}) = 1 - P(A). \tag{43.4}$$

Corollary 43.1.10. From the previous equation and de Morgan's laws (2.8) and (2.9), one can derive the following formula:

$$P(\overline{A} \cap \overline{B}) = 1 - P(A \cup B). \tag{43.5}$$

43.2 Conditional probability

Definition 43.2.1 (Conditional probability). Let A, B be two events. The probability of A given that B is true is denoted as $P(A|B)$:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{43.6}$$

By interchanging A and B in 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 43.2.2 (Bayes). *Let A, B be two events.*

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \tag{43.7}$$

Formula 43.2.3. Let $(B_n)_{n\in\mathbb{N}}$ be a sequence of pairwise disjoint events. If $\bigsqcup_{n=1}^{\infty} B_n = \Omega$, the total probability of a given event A can be calculated as follows:

$$P(A) = \sum_{n=1}^{\infty} P(A|B_n)P(B_n). \tag{43.8}$$

Definition 43.2.4 (Independent events). Let A, B be two events. A and B are said to be independent if they satisfy the following relation:

$$P(A \cap B) = P(A)P(B). \tag{43.9}$$

Corollary 43.2.5. If A and B are two independent events, Bayes's theorem simplifies to

$$P(A|B) = P(A). \tag{43.10}$$

The above definition can be generalized to multiple events:

Definition 43.2.6. The events $A_1, \ldots, A_n$ are said to be independent if for each choice of k 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 43.2.7 (Independence). The σ-algebras $\mathcal{F}_1, \ldots, \mathcal{F}_n$ defined on a probability space $(\Omega, \mathcal{F}, P)$ are said to be independent if for all choices of distinct indices $i_1, \ldots, i_k$ and for all choices of sets $F_{i_n} \in \mathcal{F}_{i_n}$ the following equation holds:

$$P(F_{i_1} \cap \cdots \cap F_{i_k}) = P(F_{i_1}) \cdots P(F_{i_k}). \tag{43.11}$$

Corollary 43.2.8. Let X, Y be two random variables. X and Y are independent if the σ-algebras generated by them are independent.

43.3 Probability distribution

Definition 43.3.1 (Probability distribution). Let X be a random variable defined on a probability space (Ω, Σ, P). The following function is a measure on the Borel σ-algebra of $\mathbb{R}$:

$$P_X(B) = P(X^{-1}(B)). \tag{43.12}$$

This measure is called the probability distribution of X.

Example 43.3.2 (Rademacher variable). A random variable on $\Omega = \{-1, 1\}$ with probability distribution $P_X(-1) = P_X(1) = \frac{1}{2}$.

Definition 43.3.3 (Density). Let $f \geq 0$ be an integrable function and recall Property 16.2.18. 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. is normalized to 1, f is called a **probability density function**.

More generally, by the Radon-Nikodym theorem 16.5.8, every absolutely continuous probability distribution P is of the form

$$P(A) = \int_A f\,d\lambda \tag{43.13}$$

for some integrable function f.

In the case where P is discrete, i.e. 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 43.3.4 (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:

$$F_X(a) := P_X(\{x \in \mathbb{R} \mid x \leq a\}). \tag{43.14}$$

Theorem 43.3.5 (Skorokhod's representation theorem). *Let $F : \mathbb{R} \to [0, 1]$ be a function that satisfies the following three properties:*

- *F is nondecreasing.*
- $\lim_{x\to-\infty} F(x) = 0$ *and* $\lim_{x\to\infty} F(x) = 1$.
- *F is right-continuous, i.e.* $\lim_{y\nearrow y_0} F(y) = F(y_0)$.

There exists a random variable $X : [0, 1] \to \mathbb{R}$ defined 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.

The following theorem is a specific instance of the more general change-of-variables formula:

Theorem 43.3.6 (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 $g \in L^1(\mathbb{R})$:*

$$\int_\Omega g \circ X\,dP = \int_{\mathbb{R}} g\,dP_X. \tag{43.15}$$

Remark 43.3.7. The name of this theorem stems from the fact that many scientists take this equality to be a definition of the expectation value $\mathrm{E}[g(X)]$. 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 43.3.8. Consider an absolutely continuous probability function P defined 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.

$$\int_{\mathbb{R}^n} g\, dP = \int_{\mathbb{R}^n} f(x)g(x)\, dx \tag{43.16}$$

Corollary 43.3.9. The previous formula together with Theorem 43.3.6 gives rise to

$$\int_{\Omega} g \circ X\, dP = \int_{\mathbb{R}^n} f_X(x)g(x)\, dx. \tag{43.17}$$

Formula 43.3.10. 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

$$f_g(y) = f(g^{-1}(y))\left|\frac{dg^{-1}}{dy}(y)\right|. \tag{43.18}$$

Weak convergence of measures 16.1.51 induces a notion for convergence of random variables:

Definition 43.3.11 (Convergence in distribution). A sequence $(X_n)_{n\in\mathbb{N}}$ of random variables is said to converge in distribution to a random variable Y if the associated cumulative distribution functions F_{X_n} converge pointwise to F_Y, i.e. $\lim_{n\to\infty} F_{X_n}(x) = F_Y(x)$ for all $x \in \mathbb{R}$, where F is continuous. This is equivalent to requiring that the associated probability measures P_{X_n} converge weakly to P_X (Definition 16.1.51).

Notation 43.3.12. If a sequence $(X_n)_{n\in\mathbb{N}}$ converges in distribution to a random variable Y, this is often denoted by $X_n \xrightarrow{d} Y$. Sometimes the d (for ‘‘distribution’’) is replaced by the $\mathcal{L}$ (for ‘‘law’’).

Theorem 43.3.13 (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:*

- $X_n + Y_n \xrightarrow{d} X + c$,
- $X_nY_n \xrightarrow{d} cX$, *and*
- $X_n/Y_n \xrightarrow{d} X/c$.

Definition 43.3.14 (Convergence in probability). A sequence $(X_n)_{n\in\mathbb{N}}$ of random variables on a metric space (Ω, d) is said to converge in probability to a random variable Y if for all $\varepsilon > 0$ the following statement holds:

$$\lim_{n\to\infty} \Pr(d(X_n, X) > \varepsilon) = 0. \tag{43.19}$$

Convergence in probability implies convergence in distribution.

Definition 43.3.15 (Giry monad ♣). Consider the category **Meas** of measurable spaces. On this space one can define a monad 4.3.16 that sends a set 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.

The unit of the Giry monad $\mathbb{P}$ is defined by assigning Dirac measures:

$$\eta_X(x) := \delta_x. \tag{43.20}$$

The multiplication map is defined as follows:

$$\mu_X(Q)(U) := \int_{P\in\mathbb{P}X} \mathrm{ev}_U(P)\, dQ. \tag{43.21}$$

43.4 Moments

43.4.1 Expectation value

Definition 43.4.1 (Expectation value). Let X be random variable defined on a probability space (Ω, Σ, P).

$$\mathrm{E}[X] := \int_\Omega X\, dP \tag{43.22}$$

Notation 43.4.2. 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 43.4.3 (Markov's inequality). Let X be a random variable. For every constant $a > 0$ the following inequality holds:

$$\Pr(X \geq a) \leq \frac{\mathrm{E}[X]}{a}. \tag{43.23}$$

Definition 43.4.4 (Moment of order r). The moment of order r is defined as the expectation value of the r^{th} power of X. By Equation (43.17) this becomes

$$\mathrm{E}[X^r] = \int_\mathbb{R} x^r f_X(x)\, dx. \tag{43.24}$$

Definition 43.4.5 (Central moment of order r).

$$\mathrm{E}[(X-\mu)^r] = \int_\mathbb{R} (x-\mu)^r f_X(x)\, dx \tag{43.25}$$

Remark 43.4.6. 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 43.4.7 (Variance). The central moment of order 2 is called the variance:

$$\mathrm{Var}[X] := \mathrm{E}\big[(X-\mu)^2\big]. \tag{43.26}$$

Definition 43.4.8 (Standard deviation).

$$\sigma_X := \sqrt{V[X]} \tag{43.27}$$

Property 43.4.9. If $\mathrm{E}[|X|^n]$ is finite for some $n > 0$, then $\mathrm{E}\big[X^k\big]$ exists and is finite for all $k \leq n$.

Property 43.4.10 (Chebyshev's inequality). Let X be a nonnegative random variable. For every constant $a > 0$ the following inequality holds:

$$\Pr(|X - \mathrm{E}[X]| \geq a) \leq \frac{\mathrm{Var}[X]}{a^2}. \tag{43.28}$$

Definition 43.4.11 (Moment generating function).

$$M_X(t) := \mathrm{E}\big[e^{tX}\big] = \int_\mathbb{R} e^{tx} f_X(x)\, dx \tag{43.29}$$

Property 43.4.12. If the moment generating function exists, the moments $\mathrm{E}[X^n]$ can be expressed in terms of M_X:

$$\mathrm{E}[X^n] = \left.\frac{d^n M_X(t)}{dt^n}\right|_{t=0}. \tag{43.30}$$

Method 43.4.13 (Chernoff bound). The Chernoff bound for a random variable gives a bound on the tail probabilities. For all constants $\lambda > 0$, the Markov inequality implies the following statement:

$$\Pr(X \geq a) = \Pr(e^{\lambda X} \geq e^{\lambda a}) \leq \frac{\mathrm{E}\left[e^{\lambda X}\right]}{e^{\lambda a}}. \tag{43.31}$$

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 43.4.14 (Hoeffding's inequalities). Consider a collection of bounded, independent random variables $X_1, \ldots, X_n$. Without loss of generality one can assume that they are bounded by the unit intevral, i.e. $0 \leq X_i \leq 1$. For every constant $\lambda \geq 0$ the following inequality holds:

$$\Pr\left(\overline{X} - \mathrm{E}\left[\overline{X}\right] \geq \lambda\right) \leq \exp(-2n\lambda^2). \tag{43.32}$$

If one can sharpen the bounds for the variables such that $X_i \in [a_i, b_i]$, then

$$\Pr\left(\overline{X} - \mathrm{E}\left[\overline{X}\right] \geq \lambda\right) \leq \exp\left(-\frac{2n^2\lambda^2}{\sum_{i=1}^n (b_i - a_i)^2}\right). \tag{43.33}$$

Definition 43.4.15 (Characteristic function).

$$\varphi_X(t) := \mathrm{E}\left[e^{itX}\right] \tag{43.34}$$

Property 43.4.16. 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 43.4.17. If $\varphi_X(t)$ is k times continuously differentiable, then X has a finite k^{th} moment and

$$\mathrm{E}\left[X^k\right] = \frac{1}{i^k}\frac{d^k}{dt^k}\varphi_X(0). \tag{43.35}$$

Conversely, if X has a finite k^{th} moment, then $\varphi_X(t)$ is k times continuously differentiable and the above formula holds.

Formula 43.4.18 (Inversion formula). Let X be a random variable. If the CDF of X is continuous at $a, b \in \mathbb{R}$, then

$$F_X(b) - F_X(a) = \lim_{c\to\infty} \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it}\varphi_X(t)\, dt. \tag{43.36}$$

Formula 43.4.19. If $\varphi_X(t)$ is integrable, the CDF is given by:

$$f_X(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\varphi_X(t)\, dt. \tag{43.37}$$

Remark 43.4.20. This formula implies that the density function and the characteristic function form a Fourier transform pair.

43.4.2 Correlation

Property 43.4.21. Two random variables X, Y are independent if and only if $\mathrm{E}[f(X)g(Y)] = \mathrm{E}[f(X)]\mathrm{E}[g(Y)]$ holds for all measurable bounded functions f, g.

The value $\mathrm{E}[XY]$ is equal to the inner product $\langle X|Y\rangle$ as defined in (16.36). It follows that independence of random variables implies orthogonality. To generalize this concept, the following notions are introduced:

Definition 43.4.22 (Centred random variable). Let X be a random variable with finite expectation value $\mathrm{E}[X]$. The centred random variable X_c is defined as $X_c = X - \mathrm{E}[X]$.

Definition 43.4.23 (Covariance). The covariance of two random variables X, Y is defined as follows:

$$\mathrm{cov}(X, Y) := \langle X_c \mid Y_c\rangle = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])]. \tag{43.38}$$

Some basic math gives

$$\mathrm{cov}(X, Y) = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]. \tag{43.39}$$

Definition 43.4.24 (Correlation). The correlation of two random variables X, Y is defined as the cosine of the angle between X_c and Y_c:

$$\rho_{XY} := \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y}. \tag{43.40}$$

Corollary 43.4.25. From Theorem 43.4.21 it follows that independent random variables are uncorrelated.

Corollary 43.4.26. If the random variables X and Y are uncorrelated, they satisfy $\mathrm{E}[XY] = \mathrm{E}[X]\mathrm{E}[Y]$.

Formula 43.4.27 (Bienaymé formula). Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent (or uncorrelated) random variables. Their variances satisfy the following equation:

$$\mathrm{Var}\left[\sum_{i=1}^{\infty} X_i\right] = \sum_{i=1}^{\infty} \mathrm{Var}[X_i]. \tag{43.41}$$

43.4.3 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 16.3.3 implies that the spaces $L^2(\Sigma)$ and $L^2(\mathcal{G})$ are complete and, hence, the projection theorem 23.2.19 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:

$$\forall Z \in L^2(\mathcal{G}) : \langle X - Y \mid Z\rangle \equiv \int_\Omega (X - Y)Z\, dP = 0. \tag{43.42}$$

Since $\mathbb{1}_G \in L^2(\mathcal{G})$ for every $G \in \mathcal{G}$, Equation (16.23) can be rewritten as

$$\int_G X\, dP = \int_G Y\, dP \tag{43.43}$$

for all $G \in \mathcal{G}$. This leads to the following definition:

Definition 43.4.28 (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 Equation (43.43) for every $G \in \mathcal{G}$. This variable Y is called the conditional expectation of X given $\mathcal{G}$ and it is denoted by $\mathrm{E}[X \mid \mathcal{G}]$:

$$\int_G \mathrm{E}[X \mid \mathcal{G}]\, dP = \int_G X\, dP. \tag{43.44}$$

Remark 43.4.29. Although this construction was based on orthogonal projections, one could as well have used the (signed) Radon-Nikodym theorem 16.7.4 since $G \mapsto \int_G X\, dP$ is absolutely continuous with respect to $P|_{\mathcal{G}}$.

Property 43.4.30. 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

$$\mathrm{E}[X \mid \mathcal{G}] = X \text{ a.s.} \tag{43.45}$$

On the other hand, if X is independent of $\mathcal{G}$, then

$$\mathrm{E}[X \mid \mathcal{G}] = \mathrm{E}[X] \text{ a.s.} \tag{43.46}$$

43.5 Joint distributions

Definition 43.5.1 (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) isa probability measure defined on the Borel algebra of $\mathbb{R}^2$ defined by

$$P_{(X,Y)}(B) = P((X, Y)^{-1}(B)). \tag{43.47}$$

Definition 43.5.2 (Joint density). If the probability measure from the previous definition can be written as

$$P_{(X,Y)}(B) = \int_B f_{(X,Y)}(x, y)\, dxdy \tag{43.48}$$

for some integrable $f_{(X,Y)}$, it is said that X and Y have a joint density.

Definition 43.5.3 (Marginal distribution). The distributions of the one-dimensional random variables is determined by the joint distribution:

$$P_X(A) = P_{(X,Y)}(A \times \mathbb{R}) \tag{43.49}$$

$$P_Y(A) = P_{(X,Y)}(\mathbb{R} \times A). \tag{43.50}$$

Corollary 43.5.4. If the joint density exists, the marginal distributions are absolutely continuous and the associated density functions are given by

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x, y)\, dy \tag{43.51}$$

$$f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x, y)\, dx. \tag{43.52}$$

The converse, however, is not always true. The one-dimensional distributions can be absolutely continuous without the existence of a joint density.

Property 43.5.5 (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:

$$P_{(X,Y)} = P_X \otimes P_Y. \tag{43.53}$$

If X and Y are absolutely continuous, the previous properties also applies to the densities instead of the distributions.

Formula 43.5.6 (Sum of random variables). Consider two independent random variables X, Y and let $Z = X + Y$ denote their sum. The density f_Z is given by the following convolution:

$$f_Z(z) := f * g(z) = \int_{\mathbb{R}} g(x)h(z-x)\,dx = \int_{\mathbb{R}} g(z-y)h(y)\,dy, \tag{43.54}$$

where g, h denote the densities of X, Y respectively.

Formula 43.5.7 (Product of random variables). Consider two independent random variables X, Y and let $Z = XY$ denote their product. The density f_Z is given by

$$f_Z(z) = \int_{\mathbb{R}} g(x)h(z/x)\,\frac{dx}{|x|} = \int_{\mathbb{R}} g(z/y)h(y)\,\frac{dy}{|y|}, \tag{43.55}$$

where g, h denote the densities of X, Y respectively.

Corollary 43.5.8. Taking the Mellin transform 17.3.25 of both the positive and negative part of the above integrand (to be able to handle the absolute value) gives the following relation:

$$\mathcal{M}\{f\} = \mathcal{M}\{g\}\mathcal{M}\{h\}. \tag{43.56}$$

Formula 43.5.9 (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

$$h(y \mid X \in A) = \frac{\int_A f_{(X,Y)}(x,y)\,dx}{\int_A f_X(x)\,dx}. \tag{43.57}$$

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

$$h(y \mid A = a) := \frac{f_{(X,Y)}(a,y)}{f_X(a)}, \tag{43.58}$$

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:

$$P(Y \in B \mid X = a) := \int_B h(y \mid X = a)\,dy. \tag{43.59}$$

Formula 43.5.10 (Conditional expectation).

$$\mathrm{E}[Y \mid X](\omega) = \int_{\mathbb{R}} yh(y \mid X(\omega))\,dy \tag{43.60}$$

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:

$$\int_A \mathrm{E}[Y \mid X]\,dP = \int_A Y\,dP. \tag{43.61}$$

This implies that the conditional expectation $\mathrm{E}[Y \mid X]$ on $\mathcal{F}_X$ coincides with Definition 43.4.28.

Applying Property 43.4.30 to the case $\mathcal{G} = \mathcal{F}_X$ gives the law of total expectation:

Property 43.5.11 (Law of total expectation[1]).

$$\mathrm{E}[\mathrm{E}[Y \mid X]] = \mathrm{E}[Y] \tag{43.62}$$

Theorem 43.5.12 (Bayes's theorem). *The conditional density can be computed without prior knowledge of the joint density:*

$$g(x \mid y) = \frac{h(y \mid x) f_X(x)}{f_Y(y)}. \tag{43.63}$$

43.6 Stochastic calculus

Definition 43.6.1 (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 time series. This will be assumed from here on.

Definition 43.6.2 (Filtered probability space). Consider a probability space (Ω, Σ, P) together with a filtration 2.4.11 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 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).

Definition 43.6.3 (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 $t \in T$.

Definition 43.6.4 (Predictable process). A stochastic process $(X_t)_{t\in T}$ on a filtered probability space $(\Omega, \Sigma, \mathbb{F}, P)$ is said to be predictable if X_{t+1} is $\mathbb{F}_t$-measurable for all $t \in T$.

Definition 43.6.5 (Stopping time). Consider a random variable τ on 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

$$\{\tau \leq t\} \in \mathbb{F}_t \tag{43.64}$$

for all t. The stopping time is a "time indicator" that only depends on the knowledge of the process up to time $t \in T$.

43.6.1 Martingales

From here on the index set T will be $\mathbb{R}_+ \equiv [0, \infty[$ so that the index t can be interpreted as a true time parameter. The discrete case $T = \mathbb{N}$ can be obtained as the restriction of most definitions or properties and, if necessary, this will be made explicit.

Definition 43.6.6 (Martingale). Consider a filtered probability space $(\Omega, \Sigma, \mathbb{F}, P)$. A stochastic process $(X_t)_{t\in T}$ is called a martingale relative to $\mathbb{F}$ if it satisfies the following conditions:

1. $(X_t)_{t\in T}$ is adapted to $\mathbb{F}$.
2. Each random variable X_t is integrable, i.e. $X_t \in L^1(P)$ for all $t \geq 0$.
3. For all $t > s \geq 0 : \mathrm{E}[X_t \mid \mathbb{F}_s] = X_s$.

If the equality in the last condition is replaced by the inequality $\leq$ (resp. $\geq$), the stochastic process is called a **supermartingale** (resp. **submartingale**).

[1] Also called the **tower property**.

Example 43.6.7 (Doob martingale). Consider an integrable random variable X and a filtration $\mathbb{F}$. The associated Doob martingale (a martingale with respect to $\mathbb{F}$) is given by

$$Y_t := \mathrm{E}[X \mid \mathbb{F}_t]. \tag{43.65}$$

Property 43.6.8 (Doob-Ville inequality). Consider a càdlàg submartinagle $(X_t)_{t\in T}$.

$$\Pr\left(\sup_{t\leq\tau} X_t \geq C\right) \leq \frac{\mathrm{E}[\max(0, X_\tau)]}{C} \tag{43.66}$$

for all $C \geq 1$ and $\tau \in T$.

The following property generalizes the Hoeffding inequalities 43.4.14:

Property 43.6.9 (Hoeffding-Azuma inequality). Let $(X_n)_{n\in\mathbb{N}}$ be a (super)martingale with bounded differences, i.e. there exist constants $c_k > 0$ such that

$$|X_k - X_{k-1}| \leq c_k. \tag{43.67}$$

The following inequality holds for all $\lambda \geq 0$:

$$\Pr(X_N - X_0 \geq \lambda) \leq \exp\left(-\frac{\lambda^2}{2\sum_{i=1}^N c_i^2}\right). \tag{43.68}$$

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

$$A_k \leq X_k - X_{k-1} \leq B_k \tag{43.69}$$

and

$$B_k - A_k \leq c_k \tag{43.70}$$

for all $k \in \mathbb{N}$, the inequality can be sharpened:

$$\Pr(X_N - X_0 \geq \lambda) \leq \exp\left(-\frac{2\lambda^2}{\sum_{i=1}^N c_i^2}\right). \tag{43.71}$$

Now, consider a function $f : \Omega^n \to \mathbb{R}$ such that

$$\sup_{x_1,\ldots,x_n,x'_k} |f(x_1, \ldots, x_k, \ldots, x_n) - f(x_1, \ldots, x'_k, \ldots, x_n)| \leq c_k \tag{43.72}$$

for all $k \in \mathbb{N}$. By applying the above inequalities to the Doob martingale

$$Z_m := \mathrm{E}[f(X_1, \ldots, X_n) \mid X_1, \ldots, X_m], \tag{43.73}$$

one obtains the following inequality:

$$\Pr(f(X_1, \ldots, X_n) - \mathrm{E}[f] \geq \lambda) \leq \exp\left(-\frac{2\lambda^2}{\sum_{i=1}^n c_i^2}\right). \tag{43.74}$$

This inequality is sometimes called the **McDiarmid inequality**.

Theorem 43.6.10 (Doob decomposition). *Any integrable adapted process $(X_t)_{t\in T}$ can be decomposed as $X_t = X_0 + M_t + A_t$, where $(M_t)_{t\in T}$ is a martingale and $(A_t)_{t\in T}$ is a predictable process. These two processes are constructed iteratively as follows:*

$$A_0 = 0 \qquad\qquad M_0 = 0 \tag{43.75}$$

$$\Delta A_t = \mathrm{E}[\Delta X_t \mid \mathbb{F}_{t-1}] \qquad\qquad \Delta M_t = \Delta X_t - \Delta A_t. \tag{43.76}$$

Furthermore, $(X_t)_{t\in T}$ is a submartingale if and only if $(A_t)_{t\in T}$ is (almost surely) increasing.

Corollary 43.6.11. Consider the special case $X = Y^2$ for some martingale Y. One can show the following property:

$$\Delta A_t = \mathrm{E}\big[(\Delta Y_t)^2 \mid \mathbb{F}_{t-1}\big] \qquad \forall t \in \mathbb{R}_+. \tag{43.77}$$

The process $(A_t)_{t\in T}$ is often called the **quadratic variation process** of $(X_t)_{t\in T}$ and is denoted by $([X]_t)_{t\in T}$.

Definition 43.6.12 (Discrete stochastic integral[2]). Let $(M_n)_{n\in\mathbb{N}}$ be a martingale on a filtered probability space $(\Omega, \Sigma, \mathbb{F}, P)$ and let $(X_n)_{n\in\mathbb{N}}$ be a predictable stochastic process with respect to $\mathbb{F}$. The (discrete) stochastic integral of X with respect to M is defined as follows:

$$(X \cdot M)_t(\omega) := \sum_{i=1}^{t} X(\omega)_i \Delta M_i(\omega), \tag{43.78}$$

where $\omega \in \Omega$. For $t = 0$ the convention $(X \cdot M)_0 = 0$ is used.

Property 43.6.13. If the process $(X_n)_{n\in\mathbb{N}}$ is bounded, the stochastic integral defines a martingale.

Property 43.6.14 (Itô isometry). Consider a martingale $(M_n)_{n\in\mathbb{N}}$ and a predictable process $(X_n)_{n\in\mathbb{N}}$. Using the Doob decomposition theorem one can show the following equality for all $n \geq 0$:

$$\mathrm{E}\Big[(X \cdot M)_n^2\Big] = \mathrm{E}\big[(X^2 \cdot [M])_n\big]. \tag{43.79}$$

It is this property that allows for the definition of integrals with respect to continuous martingales, since although the martingales are not in general of bounded variation (and hence do not induce a well-defined Lebesgue-Stieltjes integral), their quadratic variations are (e.g. the Wiener process).

43.6.2 Markov processes

Definition 43.6.15 (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

$$P(X_t \mid \mathbb{F}_s) = P(X_t \mid X_s) \tag{43.80}$$

for all $t, s \in T$. For discrete processes, the first-order Markov chains are the most common. These satisfy

$$P(X_t \mid X_{t-1}, \ldots, X_{t-r}) = P(X_t \mid X_{t-1}) \tag{43.81}$$

for all $t, r \in \mathbb{N}$.

[2]Sometimes called the **martingale transform**.

43.7 Information theory

Definition 43.7.1 (Self-information). The self-information of an event x described by a distribution P is defined as follows:

$$I(x) := -\ln P(x). \tag{43.82}$$

This definition is modeled on the following (reasonable) requirements:

- 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 43.7.2 (Shannon entropy). The amount of uncertainty in a discrete distribution P is characterized by its (Shannon) entropy

$$H(P) := \mathrm{E}[I(X)] = -\sum_i P_i \ln(P_i). \tag{43.83}$$

Definition 43.7.3 (Kullback-Leibler divergence). Let P, Q be two probability distributions. The Kullback-Leibler divergence (or **relative entropy**) of P with respect to Q is defined as follows:

$$D_{\mathrm{KL}}(P\|Q) := \int_\Omega \log\left(\frac{P}{Q}\right) dP. \tag{43.84}$$

This quantity can be interpreted as the information gained when using the distribution P instead of Q. Instead of a base-10 logarithm, any other logarithm can be used since this simply changes the result by a (positive) scaling constant.

Property 43.7.4 (Gibbs's inequality). By noting that the logarithm is a concave function and applying Jensen's equality 14.8.5, one can prove that the Kullback-Leibler divergence is nonnegative:

$$D_{\mathrm{KL}}(P\|Q) \geq 0. \tag{43.85}$$

Furthermore, the Kullback-Leibler divergence is zero if and only if P and Q are equal almost everywhere.

43.8 Extreme value theory

Definition 43.8.1 (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:

$$F_u(y) = \Pr(X - u \leq y \mid X > u) = \frac{P(u+y) - P(u)}{1 - P(u)}. \tag{43.86}$$

Definition 43.8.2 (Extreme value distribution). The extreme value distribution is given by the following formula:

$$F(x;\xi) = \exp\Big(-(1 + x\xi)^{-1/\xi}\Big). \tag{43.87}$$

In the case that $\xi = 0$, one can use the definition of the Euler number to rewrite the definition as

$$F(x;0) = \exp(-e^{-x}). \tag{43.88}$$

The number ξ is called the **extreme value index**.

[3] And by extension $P(x) \approx 1 \implies I(x) \approx 0$.

Definition 43.8.3 (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 $F^n(a_nx + b_n) \longrightarrow H(x)$.

Theorem 43.8.4 (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 43.8.5 (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 $u \longrightarrow \infty$.*

43.9 Copulas

Property 43.9.1 (Uniformization transform). 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 43.9.2 (Copula). The joint cumulative distribution function of a random variable with uniform marginal distributions.

The following alternative definition is more analytic in nature:

Alternative Definition 43.9.3 (Copula). A function $C : [0,1]^d \to [0,1]$ satisfying the following properties:

1. **Normalization** $C(x_1, \ldots, x_d) = 0$ if any of the x_i is zero.
2. **Uniformity:** $C(1, 1, \ldots, x_i, 1, \ldots) = x_i$ for all $1 \leq i \leq d$.
3. **d-nondecreasing:** For every box $B = \prod_{1\leq i\leq d}[a_i, b_i] \subseteq [0,1]^d$ the C-volume is nonnegative:

$$\int_B dC := \sum_{\mathbf{z}\in\prod_i\{a_i,b_i\}} (-1)^{N_b(\mathbf{z})}C(\mathbf{z}) \geq 0, \tag{43.89}$$

where $N_B(\mathbf{z}) = \mathrm{Card}(\{i \mid a_i = z_i\})$.

Theorem 43.9.4 (Sklar). *For every joint distribution function H with marginals F_i there exists a unique copula C such that*

$$H(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)). \tag{43.90}$$

Property 43.9.5 (Fréchet-Hoeffding bounds). Every copula $C : [0,1]^d \to [0,1]$ is bounded in the following way:

$$\max\left(\sum_{i=1}^d u_i - d + 1, 0\right) \leq C(u_1, \ldots, u_d) \leq \min_i u_i \tag{43.91}$$

for all $(u_1, \ldots, u_d) \in [0,1]^d$. Furthermore, the upper bound is sharp, i.e. $\min_i u_i$ is itself a copula.[4]

[4] The lower bound is only a copula for $d = 2$. In general this bound is only pointwise sharp.

Definition 43.9.6 (Extreme value copula). A copula C for which there exists a copula $\widetilde{C}$ such that

$$\left[\widetilde{C}(u_1^{1/n},\ldots,u_d^{1/n})\right]^n \longrightarrow C(u_1,\ldots,u_d) \tag{43.92}$$

for all $(u_1,\ldots,u_d)\in[0,1]^d$.

Property 43.9.7. A copula C is an extreme value copula if and only if it is stable in the following sense:

$$C(u_1,\ldots,u_d) = \left[C(u_1^{1/n},\ldots,u_d^{1/n})\right]^n \tag{43.93}$$

for all $n\geq 1$.

43.10 Randomness ♣

This section is strongly related to Section 6.6 on computability theory.

Definition 43.10.1 (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

$$C(\kappa) := \min\{|\sigma| \mid \sigma \text{ is finite} \wedge U(\sigma)=\kappa\}. \tag{43.94}$$

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 43.10.2. 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 43.10.3. 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 pleasing if this notion of randomness could easily be extended to infinite bit strings, for example by giving such a string the label random if there exists a uniform choice of constant k 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.

43.11 Optimal transport

In this section a new notion of atomicity of measures will be used:

Definition 43.11.1. 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.

43.11.1 Kantorovich duality

The problem of optimal transport constitutes the search of 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 43.11.2 (Cost function). A measurable function $X\times Y\to\overline{\mathbb{R}}$.

Definition 43.11.3 (Transportation scheme). A transportation scheme or **transference plan** is a joint distribution $\pi\in\mathbb{P}(X\times Y)$ whose marginals coincide with μ_X and μ_Y.

Definition 43.11.4 (Monge-Kantorovich problem). The optimal transportation scheme for a given cost function according to *Kantorovich* is the solution of the following optimization problem:

$$\inf_{\pi\in\mathbb{P}(X\times Y)} \mathrm{E}_\pi[c] = \inf_{\pi\in\mathbb{P}(X\times Y)} \int_{X\times Y} c(x,y)\,d\pi(x,y). \tag{43.95}$$

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, namely

$$\int_{X\times Y} c(x,y)\,d\pi(x,y) = \int_X c(x,T(x))\,d\mu_X(x) \tag{43.96}$$

for some measurable function $T : X \to Y$ such that $T_*\mu_X = \mu_Y$.

Example 43.11.5 (Finite state spaces). Consider the case where both X and Y are finite of the same size 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 sums 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 23.3.18 to restrict the attention to the extremal points, which in this case are given by permutation matrices. So, the optimal solution is given by the optimal one-to-one pairing of producers and consumers.

Property 43.11.6 (Kantorovich duality). Let X,Y be Polish spaces 10.4.4 and consider a lower semicontinuous function $c : X\times Y \to \overline{\mathbb{R}^+}$ (Definition 7.2.9). Denote by $\mathbb{P}_{\text{Borel}}(\mu,\nu)$ the space of Borel measures on $X\times Y$ whose marginals are given by μ_X and μ_Y. Moreover, denote by $\Phi_c \subseteq \mathcal{L}^1(X)\times\mathcal{L}^1(Y)$ the space of pairs of integrable functions satisfying

$$c_X(x) + c_Y(y) \leq c(x,y) \tag{43.97}$$

for μ_X-almost all $x \in X$ and μ_Y-almost all $y \in Y$. Then

$$\inf_{\pi\in\mathbb{P}_{\text{Borel}}(\mu,\nu)} \int_{X\times Y} c\,d\pi = \sup_{(c_X,c_Y)\in\Phi_c} \int_X c_X\,d\mu_X + \int_Y c_Y\,d\mu_Y \tag{43.98}$$

and the this problem admits a solution. Moreover, one can restrict the space of would-be solutions on the right-hand side to those that are also bounded and continuous without changing the solution.

Definition 43.11.7 (Kantorovich distance). Let X be a Polish space and consider a lower semicontinuous 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:

$$\mathcal{T}_d(\mu,\nu) := \inf_{\pi\in\mathbb{P}_{\text{Borel}}(X\times X)} \int_{X\times X} d(x,x')\,d\pi(x,x'). \tag{43.99}$$

If the metric d is the one inducing the topology on X, one obtains the definition of the **Wasserstein 1-metric**.

Theorem 43.11.8 (Kantorovich-Rubinstein). *If $X = Y$ and c is equal to some metric d on X, the Kantorovich distance is given by*

$$\mathcal{T}_d(\mu,\nu) = \sup\left\{\int_X \varphi\,d\mu - \int_X \varphi\,d\nu \,\middle|\, \varphi\in\mathrm{Lip}(X,d)\cap\mathcal{L}^1(\mu)\cap\mathcal{L}^1(\nu)\wedge\|\varphi\|_{\text{Lip}}\leq 1\right\}, \tag{43.100}$$

where

$$\|\varphi\|_{\text{Lip}} := \sup_{x\neq x'\in X} \frac{|\varphi(x)-\varphi(x')|}{d(x,x')} \tag{43.101}$$

is the ***Lipschitz norm****.*

Property 43.11.9 (Translation invariance). The Kantorovich distance is invariant under translations by finite measures.

Property 43.11.10. 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

$$\mathcal{T}_d(\mu,\nu) = \inf\{\|\sigma\|_1 \mid \nabla\cdot\sigma = \mu-\nu\}, \tag{43.102}$$

where the condition on σ makes sense by the Riesz-Markov theorem 17.1.6.

43.11.2 Convex costs

In this section cost functions of the form

$$c(x,y) = h(x-y) \tag{43.103}$$

for some convex function $h : \mathbb{R}^n \to \mathbb{R}$ are considered. Moreover, the function h will be assumed to be at least differentiable with locally Lipschitz gradient.

Definition 43.11.11 (c-concave function). 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

$$f(x) = \inf_{(x',\lambda)\in A} c(x,x') + \lambda. \tag{43.104}$$

Theorem 43.11.12 (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*

$$T(x) = x - (\nabla h)^{-1}(\nabla\psi(x)) \tag{43.105}$$

for some h-concave function $\psi : \mathbb{R}^n \to \overline{\mathbb{R}}$.

Remark 43.11.13. 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 14.8.7.

43.11.3 Concave costs

In this section cost functions of the form

$$c(x,y) = g(\|x-y\|) \tag{43.106}$$

for some concave function $g : \mathbb{R} \to \mathbb{R}$ are considered.

Property 43.11.14. 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
$$T(x) = x - (\nabla g)^{-1}\nabla\varphi(x) \tag{43.107}$$
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.

43.11.4 Densities

Property 43.11.15 (Continuity equation). Let X be a complete smooth manifold and consider a family $(T_t)_{0\leq t\leq 1}$ 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 $(\mu_t := T_{t,*}\mu)_{0\leq t\leq 1}$ uniquely satisfies the **continuity equation**:

$$\frac{\partial \mu_t}{\partial t} + \nabla\cdot(\mu_t v_t) = 0, \tag{43.108}$$

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:

$$T_t(x) := x - tv(x). \tag{43.109}$$

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:

$$\frac{\partial v_t}{\partial t} + (v_t\cdot\nabla)v_t = 0. \tag{43.110}$$

Formula 43.11.16. Given a solution of the continuity equation, the associated flow determines an optimal transport scheme for a cost function c if and only if

$$v_0 = -(\nabla c)^{-1}\nabla\psi \tag{43.111}$$

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^*:

$$\frac{\partial u}{\partial t} + c^*(\nabla u) = 0. \tag{43.112}$$

In this section one considers absolutely continuous measures with respect to the Lebesgue measure on $\mathbb{R}^n$:

$$d\mu_X = \rho_0 dx \qquad\qquad d\mu_Y = \rho_1 dx. \tag{43.113}$$

The transport cost in the Monge problem can then be rewritten as

$$\int_{\mathbb{R}^n} c(x,T(x))\rho_0(x)\,dx \tag{43.114}$$

with

$$\int_{T^{-1}(A)}\rho_0(x)\,dx = \int_A \rho_1(x)\,dx \tag{43.115}$$

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

$$\det(DT(x))\rho_1(T(x)) = \rho_0(x). \tag{43.116}$$

Example 43.11.17 (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:

$$T(x) = \nabla\varphi(x), \tag{43.117}$$

and the optimal cost is equal to the square of the **Wasserstein 2-metric**:

$$\mathcal{T}_{\|\cdot\|_2^2}(\rho_0, \rho_1) = \inf_{\pi \in \mathbb{P}_{\text{Borel}}(\rho_0, \rho_1)} \int_{\mathbb{R}^n} \|x - x'\| \, d\pi(x, x') = W_2^2(\rho_0, \rho_1). \tag{43.118}$$

Moreover, this minimum is unique a.e.

It can also be shown that the flow acts affinely:

$$\sigma_t(x) = t\nabla\Phi(x) + (1 - t)x. \tag{43.119}$$

In fact, the affinity of the flow can be shown more generally:

Property 43.11.18. 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:

$$x_t = x + t(x' - x). \tag{43.120}$$

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 48.6) in the following way:

Method 43.11.19 (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* 48.87:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0. \tag{43.121}$$

The optimization problem now becomes minimizing the *action* or *kinetic energy*:

$$K(\rho, v) := \frac{1}{2} \int_{\mathbb{R}^n} \int_0^T \rho(t, x) \|v(t, x)\|^2 \, dt dx. \tag{43.122}$$

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 43.11.20. 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.

Chapter 44

Information Geometry

The main reference for this chapter is [19]. For more information on differential geometry, see Chapter 29 and onwards.

44.1 Statistical manifolds

In this section an important subclass of Riemannian manifolds that admit two related flat connections will be introduced. These manifolds will formalize the geometric backbone of many statistical concepts and methods.

Definition 44.1.1 (Conjugate connections). Consider a Riemannian manifold (M, g) with an affine connection ∇. The conjugate (or dual) connection $\widetilde{\nabla}$ is uniquely defined by the following equation:

$$X(g(Y,Z)) = g(\nabla_X Y, Z) + g(Y, \widetilde{\nabla}_X Z), \tag{44.1}$$

where $X, Y, Z \in TM$. Moreover, this construction is involutive:

$$\widetilde{\widetilde{\nabla}} = \nabla. \tag{44.2}$$

Property 44.1.2. 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:

$$g(v,w) = g\Big(\mathcal{P}_\gamma v, \widetilde{\mathcal{P}}_\gamma w\Big) \tag{44.3}$$

for every smooth path γ.

Property 44.1.3. Consider two conjugate connections $\nabla, \widetilde{\nabla}$ on a Riemannian manifold (M, g). The connection

$$\overline{\nabla} := \frac{\nabla + \widetilde{\nabla}}{2} \tag{44.4}$$

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 34.1.19.

The above properties lead to the following definition:

Definition 44.1.4 (Statistical manifold). A Riemannian manifold (M, g) equipped with an affine connection that satisfies the **Codazzi condition**

$$\nabla_X g(Y, Z) = \nabla_Y g(X, Z), \tag{44.5}$$

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 $\widetilde{\nabla}$:

$$T_{ijk} = \widetilde{\Gamma}_{ijk} - \Gamma_{ijk}. \tag{44.6}$$

In the case where ∇ has nonvanishing torsion, one can generalize this definition by also relaxing the Codazzi equation:

$$\nabla_X g(Y, Z) - \nabla_Y g(X, Z) = -g(T^\nabla(X, Y), Z). \tag{44.7}$$

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] The existence of a torsion-free connection is sufficient to turn a (pseudo-)Riemannian manifold into a statistical manifold admitting torsion.

Remark 44.1.5. One can show that the above definition is equivalent to that of a Riemannian manifold admitting a totally symmetric rank-3 tensor.

Definition 44.1.6 (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.

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

$$\nabla_{\partial_i} \partial_j = 0 \tag{44.8}$$

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:

$$g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial y_j}\right) = \delta_i^j. \tag{44.9}$$

This coordinate system is called the **dual (coordinate) system**.

44.1.1 Divergences

Definition 44.1.7 (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

$$D^*(p\|q) := D(q\|p) \tag{44.10}$$

for all $p, q \in M$.

[1] This situation arises in the study of quantum systems and density operators.

Property 44.1.8 (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:

$$g_{ij}(\theta) := \frac{\partial^2 D}{\partial p^i \partial p^j}(p||q)\bigg|_{p=q=\theta} = \frac{\partial^2 D}{\partial q^i \partial q^j}(p||q)\bigg|_{p=q=\theta} = -\frac{\partial^2 D}{\partial p^i \partial q^j}(p||q)\bigg|_{p=q=\theta}. \tag{44.11}$$

Example 44.1.9 (f-divergences and α-divergences). Let f 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:

$$D_f(p\|q) := \int_\Omega f\left(\frac{dp}{dq}\right) dq. \tag{44.12}$$

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

$$D_f(p\|q) = \int_\Omega h(x) f\left(\frac{g(x)}{h(x)}\right) d\mu(x). \tag{44.13}$$

It is not hard to see that f-divergences remain invariant under affine transformations of the form

$$f(x) \longrightarrow f(x) + c(x-1), \tag{44.14}$$

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 scale transformations 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

$$f_\alpha(x) = \frac{1 - x^\alpha}{\alpha(1-\alpha)}. \tag{44.15}$$

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 44.1.10 (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:

$$D_F(p\|q) := F(p) - F(q) - \vec{\nabla} F(q) \cdot (p - q), \tag{44.16}$$

where the gradient is denoted by $\vec{\nabla}$ to avoid confusion with further occurences 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.)

[2] *Tsallis* and *Rényi* introduced different divergences/entropies with the same name.

Example 44.1.11 (Kullback-Leibler divergence). The Kullback-Leibler divergence 43.7.3 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 $f = x\ln(x)$.

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 14.8.6 to obtain a new function:

$$\widetilde{F}(x^*) := x^* \cdot y - F(y), \tag{44.17}$$

where $x^* = \vec{\nabla}F(y)$.[3] It can be shown that $\widetilde{F}$ is again (strictly) convex and, hence, also defines a Bregman divergence. This second Bregman divergence coincides with the dual divergence D_F^*:

$$D_F(p\|q) = D_{\widetilde{F}}(q^*\|p^*). \tag{44.18}$$

Using this relation one can also rewrite the original expression for the Bregman divergence:

$$D_F(p\|q) = F(p) + \widetilde{F}(q) - x^i(p)y_i(q). \tag{44.19}$$

Now, the two convex functions $F, \widetilde{F}$ define two coordinate systems that are related as follows:

$$y = \vec{\nabla}F \qquad \text{and} \qquad x = \vec{\nabla}\widetilde{F}(y). \tag{44.20}$$

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:

$$\gamma_{q\to p}(t) \equiv tx(p) + (1-t)x(q). \tag{44.21}$$

Geodesics for the conjugate connection are called **dual geodesics**. It is important to note that the Legendre transform that 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 $\widetilde{\nabla}$ does preserve the metric. The metric structures induced by the Hessians are also intertwined:

$$g_{ij} = \frac{\partial^2 F}{\partial x^i \partial x^j} \qquad \text{and} \qquad \widetilde{g}^{ij} = \frac{\partial^2 \widetilde{F}}{\partial y_i \partial y_j} \tag{44.22}$$

are mutual inverses. It can be concluded that a Bregman divergence endows a set with the structure of a dually flat manifold 44.1.6.

Example 44.1.12 (Euclidean distance). On $\mathbb{R}^n$ the most common choice of divergence measure is the Euclidean distance:

$$D_{\mathrm{eucl}}(p\|q) := \frac{1}{2}\|p-q\|. \tag{44.23}$$

It is not hard to show that this function is in fact self-dual 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.

[3] For general convex functions this relation is not necessarily invertible.

Property 44.1.13 (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

$$g_{ij} = \partial_i \partial_j \psi. \tag{44.24}$$

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 $(M, g, \nabla, \widetilde{\nabla})$.

44.1.2 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 44.1.14 (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 $\{h_i : \mathbb{R}^k \to \mathbb{R}\}_{1 \leq i \leq n}$ and any smooth function $\lambda : \mathbb{R}^k \to \mathbb{R}$ one can define the following family of distributions indexed by some parameter $\theta \in \mathbb{R}^n$:

$$p(X; \theta) := \exp\Big(h_i(X)\theta^i + \lambda(X) - \psi(\theta)\Big). \tag{44.25}$$

The function $\psi(\theta)$ is introduced as a normalization function:

$$\psi(\theta) := \ln \int \exp(h_i(X)\theta^i) e^{\lambda(X)} d\mu(X). \tag{44.26}$$

The function λ can be removed through a redefinition of the measure:

$$d\mu(X) \longrightarrow d\nu(X) := \exp^{\lambda(X)} d\mu(X).$$

Remark 44.1.15. 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 $\mathrm{E}[X]$ and the associated dual convex function ϕ is the Shannon entropy. Accordingly, the Bregman divergence associated to ψ is given by the <u>dual</u> Kullback-Leibler divergence:

$$D_\psi(\theta \| \theta') = D_{\mathrm{KL}}(\theta' \| \theta). \tag{44.27}$$

The metric induced by this structure is the Fisher information:

$$g_{ij} = \mathcal{I}_{ij}[X; \theta] := \mathrm{E}\left[\left(\frac{\partial}{\partial\theta^i} \ln p(X; \theta)\right)\left(\frac{\partial}{\partial\theta^j} \ln p(X; \theta)\right)\right]. \tag{44.28}$$

Now, consider an affine combination of natural parameters, hence an affine geodesic in the manifold of an exponential family:

$$\theta(t) := t\theta_2 + (1 - t)\theta_1.$$

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).

44.1.3 Mixtures

Another important class of probability distributions is given by the mixture families:

Definition 44.1.16 (Mixture). Consider a collection of probability distributions $\{p_i(X)\}_{i\leq n}$. For every point $\eta \equiv (\eta_0, \ldots, \eta_n)$ in the probability simplex Δ^n, one can define the distribution

$$p(X;\eta) := \sum_{i=0}^{n} \eta_i p_i(X). \tag{44.29}$$

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:

$$D_\varphi(\eta\|\eta') = D_{\mathrm{KL}}(\eta\|\eta'). \tag{44.30}$$

Example 44.1.17 (Discrete distribution). An interesting example of mixtures is given by the class of discrete (or categorical) distributions:

$$p(X;\eta) = \sum_{i=0}^{n} \eta_i \delta_i(X), \tag{44.31}$$

where $\eta \in \Delta^n$. At the same time these models can be considered as distributions in an exponential family with affine coordinates $\theta^i := \ln \frac{p_i}{p_0}$. 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

$$\eta(t) = t\eta_2 + (1-y)\eta_1.$$

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 44.1.18. The reader should be aware of an important source of confusion. The above sections would point to the following naming convention:

$$\begin{array}{lcl} \text{e-geodesic} & \leftrightarrow & \text{primary geodesic} \\ \text{m-geodesic} & \leftrightarrow & \text{dual geodesic} \end{array}$$

However, because the Kullback-Leibler divergence is the most widely used divergence measure, Equation (44.27) where the KL-divergence is the dual divergence, made it possible that the above convention got reversed in the bulk of the literature. 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 that one pays 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 one knows the divergence.

[4]For arbitrary families the dual geodesic does not necessarily induce a mixture of distributions.

44.1.4 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 44.1.19 (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) if the distribution of X conditioned on $\xi(X)$ is independent of θ. The **Fisher-Neyman factorization theorem** states that this is equivalent to the existence of the following decomposition

$$p(X;\theta) = f(X)g_\theta(\xi(X)) \tag{44.32}$$

for some nonnegative functions f, g_θ.

The invariance criterion states that these transformations are the only transformations that leave the divergence invariant:

Axiom 44.1 (Invariance criterion). Consider a dually flat manifold M. Compatible divergences should satisfy the following inequality for all transformations $\overline{x} := \xi(x)$:

$$\overline{D}(\theta\|\theta') \leq D(\theta\|\theta'). \tag{44.33}$$

The equality holds if and only if the transformed variable $\overline{x}$ is a sufficient statistic.

Example 44.1.20 (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**

$$D_f(p\|q) = \sum_{i=0}^{n} d(p_i, q_i) \tag{44.34}$$

for some nonnegative function d.

The following result gives a partial characterization of invariant divergences on the manifold of discrete distributions Δ^n:

Property 44.1.21. 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 $D = D_{\text{KL}}$. When extended to $\mathbb{R}^n_+$ (the discrete positive measures), the collection of all α-divergences is recovered.

Corollary 44.1.22. 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 44.1.23 (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.

44.1.5 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:

$$\theta^i \equiv h_\alpha(m_i) := m_i^{\frac{1-\alpha}{2}}. \tag{44.35}$$

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:

$$\psi_\alpha(\theta) := \frac{1-\alpha}{2}\sum_{i=0}^n m_i = \frac{1-\alpha}{2}\sum_{i=0}^n (\theta^i)^{\frac{2}{1-\alpha}}. \tag{44.36}$$

The dual coordinates are given by

$$\eta_i \equiv h_{-\alpha(m_i)} = m_i^{\frac{1+\alpha}{2}}. \tag{44.37}$$

It should be noted that the normalization constraint $\sum_{i=0}^n m_i = 1$ 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:

$$m_h(x,y) := h^{-1}\left(\frac{h(x)+h(y)}{2}\right). \tag{44.38}$$

The α-representations are exactly the ones inducing linearly scaling h-means:

$$m_\alpha(\lambda x, \lambda y) = \lambda m_\alpha(x,y). \tag{44.39}$$

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.

44.2 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 44.2.1 (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:*

$$D(P\|R) = D(P\|Q) + D(Q\|R). \tag{44.40}$$

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 23.2.19, 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 44.2.2 (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 simlar way.

Because in general there exist multiple projections, a strict minimality theorem cannot be formulated:

Theorem 44.2.3 (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 44.2.4. 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 44.2.5 (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: $\pi_m(p) := \arg\min_{q\in S} D_{\mathrm{KL}}(p\|q)$.

Chapter 45

Statistics

In this chapter, most definitions and formulas will be based on either a standard calculus approach or a data-driven approach. For a measure-theoretic approach, see Chapter 43. For some sections the language of information geometry will be used as introduced in the previous chapter.

45.1 Data samples

45.1.1 Moment estimators

Formula 45.1.1 (r^{th} sample moment).

$$\overline{x^r} := \frac{1}{N}\sum_{i=1}^{N} x_i^r \tag{45.1}$$

Example 45.1.2 (Arithmetic mean). The arithmetic mean is used to average out differences between measurements. It is defined as the first sample moment:

$$\overline{x} := \frac{1}{N}\sum_{i=1}^{N} x_i. \tag{45.2}$$

Theorem 45.1.3 (Weak law of large numbers[1]). *Assume that the sequence $(X_n)_{n\in\mathbb{N}}$ of random variables is i.i.d. The sample average converges in probability 43.3.14 to the expectation value:*

$$\lim_{n\to\infty} \Pr\big(|\overline{X}_n - \mu| > \varepsilon\big) = 0 \tag{45.3}$$

for all $\varepsilon > 0$.

Theorem 45.1.4 (Strong law of large numbers[2]). *Assume that the sequence $(X_n)_{n\in\mathbb{N}}$ of random variables is i.i.d. The sample average converges almost surely 16.1.5 to the expectation value:*

$$\Pr\Big(\lim_{n\to\infty} \overline{X}_n = \mu\Big) = 0. \tag{45.4}$$

In fact, the i.i.d. assumption can be weakened. The convergence

$$\overline{X}_n \xrightarrow{a.s.} \mathrm{E}\big[\overline{X}_n\big] \tag{45.5}$$

[1] Also called **Khinchin's law**.
[2] Also called **Kolmogorov's law**.

holds as long as the random variables are independent, have finite second moment and satisfy

$$\sum_{k=1}^{\infty}\frac{1}{k^2}\mathrm{Var}[X_k] < \infty. \tag{45.6}$$

Formula 45.1.5 (r^{th} central sample moment).

$$m_r := \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^r \tag{45.7}$$

Definition 45.1.6 (Weighted mean). Let $f : \mathbb{R} \to \mathbb{R}^+$ be a weight function. The weighted mean is given by:

$$\overline{x} := \frac{\sum_i f(x_i)x_i}{\sum_i f(x_i)}. \tag{45.8}$$

Example 45.1.7 (Binned mean). If the data has been grouped in bins, the weight function is given by the number of elements in each bin.

$$\overline{x} = \frac{1}{N}\sum_{i=1} n_i x_i. \tag{45.9}$$

Remark 45.1.8. In the above definitions, the measurements x_i can be replaced by function values $f(x_i)$ to calculate the mean of the function $\overline{f(x)}$. This follows from Theorem 43.3.6. 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 45.1.9 (Geometric mean). Let $\{x_i\}$ be a data set taking values in either $\mathbb{R}_+$ or $\mathbb{R}_-$. The geometric mean is used to average out *normalized* measurements, i.e. ratios with respect to a reference value.

$$g := \left(\prod_{i=1}^{N} x_i\right)^{1/N} \tag{45.10}$$

The following relation exists between the arithmetic and geometic mean:

$$\ln g = \overline{\ln x}. \tag{45.11}$$

Definition 45.1.10 (Harmonic mean).

$$h := \left(\frac{1}{N}\sum_{i=1}^{N} x_i^{-1}\right)^{-1} \tag{45.12}$$

The following relation exists between the arithmetic and harmonic mean:

$$\frac{1}{h} = \overline{x^{-1}}. \tag{45.13}$$

Property 45.1.11. Let $\{x_i\}$ be a data set taking values in $\mathbb{R}_+$.

$$h \leq g \leq \overline{x} \tag{45.14}$$

The equalities only hold when all x_i are equal.

Definition 45.1.12 (Mode). The most occurring value in a data set.

Definition 45.1.13 (Median). The element x_i in a data set such that half of the values is greater than x_i and half of the values is smaller than x_i.

45.1.2 Dispersion

Definition 45.1.14 (Range). The simplest indicator for statistical dispersion:

$$R := x_{\text{max}} - x_{\text{min}}. \tag{45.15}$$

However, it is very sensitive for outliers.

Definition 45.1.15 (Mean absolute difference).

$$\text{MD} := \frac{1}{N}\sum_{i=1}^{N}|x_i - \overline{x}| \tag{45.16}$$

Definition 45.1.16 (Sample variance).

$$\text{Var}(x) := \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2 \tag{45.17}$$

Formula 45.1.17. The variance can also be rewritten in the following way:

$$\text{Var}(x) = \overline{x^2} - \overline{x}^2. \tag{45.18}$$

Remark 45.1.18 (Bessel corection). A better estimator for the variance of a sample is given by the following formula:

$$\hat{s}^2 := \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2. \tag{45.19}$$

See Remark 45.4.9 for more information.

Definition 45.1.19 (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:

$$m_3 = \gamma\sigma^3. \tag{45.20}$$

A positive skewness indicates a tail to the right or alternatively a median smaller than $\overline{x}$. A negative skewness indicates a median larger than $\overline{x}$.

Definition 45.1.20 (Pearson's mode skewness).

$$\gamma_P := \frac{\overline{x} - \text{mode}}{\sigma} \tag{45.21}$$

Definition 45.1.21 (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:

$$m_4 = c\sigma^4. \tag{45.22}$$

Definition 45.1.22 (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, a negative excess indicates short "thin" tails.

Definition 45.1.23 (Percentile). The p-percentile c_p is defined as the value that is larger than $p\%$ of the measurements. The median is the 50-percentile.

Definition 45.1.24 (Interquartile range). The difference between the upper and lower quartile (75- and 25-percentiles respectively).

Definition 45.1.25 (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**.

Property 45.1.26. For Gaussian distributions the following relation exists between the FWHM and the standard deviation:

$$\mathrm{FWHM} = 2.35\sigma. \tag{45.23}$$

45.1.3 Multivariate data sets

When working with bivariate (or even multivariate) distributions it is useful to describe the relationship between the different random variables.

Definition 45.1.27 (Covariance). The covariance of two data sequences is defined as follows:

$$\mathrm{cov}(x,y) := \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y}) = \overline{xy} - \overline{x}\,\overline{y}. \tag{45.24}$$

The covariance is also often denoted by σ_{xy} because of the next property:

Property 45.1.28. The covariance and standard deviation are related by the following equality:

$$\sigma_x^2 = \sigma_{xx}. \tag{45.25}$$

Definition 45.1.29 (Correlation coefficent).

$$\rho_{xy} := \frac{\mathrm{cov}(x,y)}{\sigma_x\sigma_y} \tag{45.26}$$

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.

Remark 45.1.30. For multivariate distributions the above definitions can be generalized using matrices:

$$V_{ij} = \mathrm{cov}(x_{(i)}, x_{(j)}) \tag{45.27}$$

$$\rho_{ij} = \rho_{(i)(j)}. \tag{45.28}$$

45.2 Probability distributions

In the following sections and subsections, all distributions will be taken to be continuous. The formulas can be generalized to discrete distributions by replacing the integral with a summation.

Definition 45.2.1 (Percentile). The p-percentile c_p of a distribution F is defined as:

$$c_p = F^{-1}(p). \tag{45.29}$$

Definition 45.2.2 (Parametric family). A family of probability densities indexed by one or more parameters θ.

Example 45.2.3 (Mixture family). Consider a collection of distributions $\mathcal{P} = \{P_i\}_{i\leq n}$. The mixture family generated by $\mathcal{P}$ consist of all convex combintations of elements in $\mathcal{P}$:

$$\left\{\sum_{i=1}^{n} w_iP_i \,\middle|\, w_i \geq 0, \sum_{i=1}^{n} w_i = 1\right\}. \tag{45.30}$$

Every element of this family is called a **mixture distribution**.

45.2.1 Empirical distribution

Definition 45.2.4 (Empirical distribution function). The (discrete) empirical probability distribution function is defined as the uniform mixture distribution with Dirac measures at the observations:

$$F_n := \frac{1}{n}\sum_{i=1}^n \delta_{x_i}. \tag{45.31}$$

Theorem 45.2.5 (Borel's law of large numbers). *If the sample size approaches infinity, the observed frequencies approach the theoretical propabilities.*

Corollary 45.2.6 (Frequentist probability[3]).

$$\Pr(x) := \lim_{n\to\infty} \frac{f_n(x)}{n} \tag{45.32}$$

The law of large numbers can also be phrased in terms of the empirical distribution function:

Theorem 45.2.7 (Glivenko-Cantelli). *Consider a cumulative distribution function F on a probability space Ω. Denote the empirical distribution function of n random variables on Ω by F_n. If the random variables are i.i.d. according to F, then*

$$\sup_{x\in\Omega} |F(x) - F_n(x)| \xrightarrow{a.s.} 0. \tag{45.33}$$

Remark 45.2.8. 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 45.2.9 (Kolmogorov-Smirnov statistic). Let F be a given cumulative distribution function. The n^{th} Kolmogorov-Smirnov statistic is defined as follows:

$$D_n := \sup_{x\in\Omega} |F_n(x) - F(x)|. \tag{45.34}$$

Definition 45.2.10 (Kolmogorov distribution).

$$F_{\text{Kol}}(x) := 1 - 2\sum_{i=1}^\infty (-1)^{i-1} e^{-2i^2x^2} = \frac{\sqrt{2\pi}}{x}\sum_{i=1}^\infty e^{-(2i-1)^2\pi^2/(8x^2)} \tag{45.35}$$

Property 45.2.11 (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

$$\sqrt{n}D_n > K_\alpha, \tag{45.36}$$

where K_α is defined by the Kolmogorov distribution: $F_{\text{Kol}}(K_\alpha) = 1 - \alpha$.

Definition 45.2.12 (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:

$$\|P\|_{\mathcal{F}} := \sup\{\mathrm{E}_P[f] \mid f \in \mathcal{F}\}. \tag{45.37}$$

[3] Also called the **empirical probability**.

A class $\mathcal{F}$ of measurable functions is said to be Glivenko-Cantelli with respect to a probability measure P if it satisfies

$$\|P_n - P\|_{\mathcal{F}} \xrightarrow{\text{a.s.}} 0, \tag{45.38}$$

where P_n is the empirical measure.[4] The Glivenko-Cantelli theorem 45.2.7 says that the indicator functions of the sets $]-\infty, x]$ form a Glivenko-Cantelli 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 45.2.13. 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 45.2.14 (Bracketing number). Consider a collection of measurable functions $\mathcal{F}$ on a meaurable space (Ω, Σ) and recall Definition 23.1.17 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 45.2.15 (Metric entropy). Consider a collection of measurable functions $\mathcal{F}$ on a meaurable space (Ω, Σ) and recall Definition 10.5.3 of the metric entropy. Moreover, assume that $\mathcal{F}$ admits an integrable envelope F. Let $\mathcal{F}_M$ denote the collection of functions $f\mathbb{1}_{F\leq M}$, where $f \in \mathcal{F}$. If

$$\frac{1}{n} \ln N_C(\varepsilon, \mathcal{F}_M, \|\cdot\|_1) \xrightarrow{d} 0, \tag{45.39}$$

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 45.2.16 (Symmetrized empirical measure). Let $\{\sigma_1, \ldots, \sigma_n\}$ be a set of i.i.d. Rademacher variables 43.3.2. The symmetrized empirical measure is defined as follows:

$$P_n^\sigma : f \mapsto \frac{1}{n} \sum_{i=1}^n \sigma_i f(x_i). \tag{45.40}$$

Given a collection $\mathcal{F}$ of measurable functions on (Ω, Σ, P), the following inequality holds:

$$\mathrm{E}[\|P_n - P\|_{\mathcal{F}}] \leq 2\mathrm{E}[\|P_n^\sigma\|_{\mathcal{F}}]. \tag{45.41}$$

Theorem 45.2.17 (Donsker). *Consider an empirical distribution function F_n and define its normalized empirical process as*

$$G_n := \sqrt{n}(F_n - F), \tag{45.42}$$

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:

$$G_n \xrightarrow{d} U \tag{45.43}$$

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 10.3.3.

[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.

45.2.2 Common distributions

Definition 45.2.18 (Uniform distribution).

$$f(x;a,b) := \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{elsewhere} \end{cases} \tag{45.44}$$

$$\mathrm{E}[x] = \frac{a+b}{2} \tag{45.45}$$

$$\mathrm{Var}[x] = \frac{(b-a)^2}{12} \tag{45.46}$$

Definition 45.2.19 (Gaussian distribution). Let σ be a positive number.

$$\mathcal{N}(x;\mu,\sigma) := \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \tag{45.47}$$

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, the **covariance** matrix. The associated multivariate normal distribution is given by

$$\mathcal{N}(\vec{\boldsymbol{x}};\vec{\boldsymbol{\mu}},\Sigma) = \frac{1}{\sqrt{2\pi}\det(\Sigma)}\exp\left(-\frac{(\vec{\boldsymbol{x}}-\vec{\boldsymbol{\mu}})^T\Sigma^{-1}(\vec{\boldsymbol{x}}-\vec{\boldsymbol{\mu}})}{2}\right). \tag{45.48}$$

Definition 45.2.20 (Standard normal distribution).

$$\mathcal{N}(z) := \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}} \tag{45.49}$$

The cumulative distribution of $\mathcal{N}$ is called the **error function** $\mathrm{erf}(x)$.

Remark 45.2.21. 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 45.2.22 (Central limit theorem). *A sum of n i.i.d. random variables X_i distributed according to a distribution with mean μ and variance σ^2 satisfies the following property:*

$$\sqrt{n}\left(\sum_{i=1}^n X_i - \mu\right) \xrightarrow{d} \mathcal{N}(0,\sigma^2). \tag{45.50}$$

Remark 45.2.23. 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 45.2.24. The sum of any number of (independent) Gaussian random variables is again Gaussian with the sum of the means and variances as parameters:

$$\forall i \in I : X_i \sim \mathcal{N}(\mu_i,\sigma_i^2) \implies \sum_{i\in I} X_i \sim \mathcal{N}\left(\sum_{i\in I}\mu_i, \sum_{i\in I}\sigma_i^2\right). \tag{45.51}$$

Definition 45.2.25 (Exponential distribution).

$$f(x;\tau) := \frac{1}{\tau}e^{-\frac{x}{\tau}} \tag{45.52}$$

$$\mathrm{E}[x] = \tau \tag{45.53}$$

$$\mathrm{Var}[x] = \tau^2. \tag{45.54}$$

Property 45.2.26. The exponential distribution is **memoryless**:

$$\Pr(X > x_1 + x_2 \mid X > x_2) = \Pr(X > x_1). \tag{45.55}$$

Definition 45.2.27 (Bernoulli distribution). A random variable that can only take 2 possible values is described by a Bernoulli distribution. When the possible values are 0 and 1, with respective chances ρ and $1-\rho$, the distribution is given by

$$p(k;\rho) := \rho^k(1-\rho)^{1-k} \tag{45.56}$$

$$\mathrm{E}[k] = \rho \tag{45.57}$$

$$\mathrm{Var}[k] = \rho(1-\rho). \tag{45.58}$$

Definition 45.2.28 (Binomial distribution). A process with n i.i.d. Bernoulli trials with probability ρ, is described by a binomial distribution:

$$\mathrm{Binom}(k;\rho,n) := \binom{n}{k}\rho^k(1-\rho)^{n-k} \tag{45.59}$$

$$\mathrm{E}[k] = n\rho \tag{45.60}$$

$$\mathrm{Var}[k] = n\rho(1-\rho). \tag{45.61}$$

Definition 45.2.29 (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 λ.

$$\mathrm{Poisson}(r;\lambda) := \frac{e^{-\lambda}\lambda^r}{r!} \tag{45.62}$$

$$\mathrm{E}[r] = \mathrm{Var}[r] = \lambda. \tag{45.63}$$

Formula 45.2.30. If multiple independent Poisson processes occur simultaneously, the probability of r events is also described by a Poisson distribution:

$$\forall i \in I : X_i \sim \mathrm{Poisson}(\lambda_i) \implies \sum_{i\in I} X_i \sim \mathrm{Poisson}\left(\sum_{i\in I}\lambda_i\right). \tag{45.64}$$

The number of events coming from the process described by $\lambda_i | r = \sum_{i\in I}\lambda_i$ is given by a binomial distribution $\mathrm{Binom}(r_i;\Lambda_i,r)$ with $\Lambda_i = \frac{\lambda_i}{\sum_{i\in I}\lambda_i}$.

Remark 45.2.31. For $\lambda \longrightarrow \infty$, the Poisson distribution $\mathrm{Poisson}(r;\lambda)$ can be approximated by a Gaussian distribution $\mathcal{N}(x;\lambda,\sqrt{\lambda})$.

Theorem 45.2.32 (Raikov). *If the sum of two independent random variables is Poisson, the individual random variables are also Poisson.*

Definition 45.2.33 (χ^2-distribution). The sum of k squared independent (standard) normally distributed random variables Y_i defines the random variable:

$$\chi_k^2 := \sum_{i=1}^{k} Y_i^2, \tag{45.65}$$

where k is said to be the number of **degrees of freedom**. The associated density is

$$f(\chi^2; n) := \frac{\chi^{n-2} e^{-\frac{\chi^2}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}. \tag{45.66}$$

Property 45.2.34. Due to the CLT 45.2.22, the χ^2-distribution approximates a Gaussian distribution for large n:

$$f(\chi^2; n) \overset{n>30}{\longrightarrow} \mathcal{N}(\sqrt{2\chi^2}; \sqrt{2n-1}, 1). \tag{45.67}$$

Definition 45.2.35 (Student-t distribution). The Student-t distribution describes the difference between the true mean and a sample average with estimated standard deviation $\hat{\sigma}$:

$$f(t; n) := \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\, \Gamma\left(\frac{n}{2}\right) \left(1 + \frac{t^2}{n}\right)^{\frac{n+1}{2}}}, \tag{45.68}$$

where

$$t := \frac{(x-\mu)/\sigma}{\hat{\sigma}/\sigma} = \frac{z}{\sqrt{\chi^2/n}}. \tag{45.69}$$

Definition 45.2.36 (Cauchy distribution[6]). The general density $f(x; x_0, \gamma)$ is given by

$$f(x; x_0, \gamma) := \frac{1}{\pi} \frac{\gamma}{(x-x_0)^2 + \gamma^2}. \tag{45.70}$$

The associated characteristic function is given by

$$\mathrm{E}\left[e^{itx}\right] = e^{ix_0 t - \gamma|t|}. \tag{45.71}$$

Remark 45.2.37. Both the mean and variance of the Cauchy distribution are undefined.

45.3 Errors

Definition 45.3.1 (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 45.3.2 (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 45.1.2 in a meaningful way because the measurements are not of the same type. Therefore, it is also impossible to apply the CLT 45.2.22.

[6] Also known, especially in particle physics, as the **Breit-Wigner** distribution.

These problems can be resolved by the using the weighted mean 45.1.6:

$$\overline{x} := \frac{\sum_i \frac{x_i}{\sigma_i^2}}{\sum_i \frac{1}{\sigma_i^2}}. \tag{45.72}$$

The variation of the weighted mean is given by

$$\mathrm{Var}(\overline{x}) := \frac{1}{\sum_i \sigma_i^{-2}}. \tag{45.73}$$

Formula 45.3.3 (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

$$\mathrm{Var}[f(X)] \approx \sum_{i=1}^n \left(\frac{\partial f}{\partial X_i}\right)^2 \mathrm{Var}[X_i] + \sum_{i\neq j}\left(\frac{\partial f}{\partial X_i}\right)\left(\frac{\partial f}{\partial X_j}\right)\mathrm{cov}[X_i, X_j]. \tag{45.74}$$

Corollary 45.3.4. The correlation coefficient 45.1.29 of a random variable X and a **linear** function of X is independent of σ_x and is always equal to ± 1.

Definition 45.3.5 (Fractional error). Let X, Y be two independent random variables. The standard deviation of $f(X,Y) = XY$ is given by the fractional error:

$$\left(\frac{\sigma_f}{f}\right)^2 = \left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2. \tag{45.75}$$

The fractional error of a variable is equal to the fractional error of the reciprocal of that variable.

Property 45.3.6 (Logarithm). Let X be a random variable. The error of the logarithm of X is equal to the fractional error of X.

Formula 45.3.7 (Covariance of functions).

$$\mathrm{cov}[f, g] \approx \sum_{i,j}\left(\frac{\partial f}{\partial X_i}\right)\left(\frac{\partial g}{\partial X_j}\right)\mathrm{cov}[X_i, X_j] \tag{45.76}$$

Corollary 45.3.8. Let $f \equiv (f_1, \ldots, f_k)$ be a vector-valued function. The covariance matrix is to first order given by

$$\mathrm{Var}[f(X)] \approx J\mathrm{Var}[X]J^T, \tag{45.77}$$

where J is the Jacobian matrix of f.

45.4 Parameter estimation

45.4.1 General properties

Definition 45.4.1 (Consistency). An estimator $\hat{a}$ is said to be consistent if it is asymptotically equal to the true parameter:

$$\lim_{N\to\infty} \hat{a} = a. \tag{45.78}$$

Definition 45.4.2 (Unbiased estimator). An estimator $\hat{a}$ is said to be unbiased if its expectation value is equal to the true parameter:

$$\langle \hat{a} \rangle = a. \tag{45.79}$$

Note that neither consistency, nor unbiasedness implies the other.

Definition 45.4.3 (Bias).

$$B(\hat{a}) := |\langle\hat{a}\rangle - a|. \tag{45.80}$$

Definition 45.4.4 (Mean squared error).

$$\mathrm{MSE}(\hat{a}) := B(\hat{a})^2 + \mathrm{Var}(\hat{a}). \tag{45.81}$$

Remark 45.4.5. If an estimator is unbiased, the MSE is equal to the variance of the estimator.

45.4.2 Common estimators

Property 45.4.6 (Unbiased mean). The CLT 45.2.22 implies that the sample mean 45.1.2 is a consistent and unbiased estimator of the population mean.

Formula 45.4.7 (Standard error of the mean). Using the Bienaymé formula 43.4.27 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:

$$\mathrm{Var}[\overline{x}] = \frac{\sigma^2}{N}. \tag{45.82}$$

Formula 45.4.8 (Variance estimator for known mean). If the true mean μ is known, a consistent and unbiased estimator for the variance is given by

$$\widehat{\mathrm{Var}[X]} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2. \tag{45.83}$$

Formula 45.4.9 (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

$$\hat{s}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2. \tag{45.84}$$

The modified factor $\frac{1}{N-1}$ is called the **Bessel correction**. It corrects the bias of the estimator given by the sample variance 45.1.16. The consistency is guaranteed by the CLT.

Property 45.4.10 (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.

45.4.3 Estimation error

Formula 45.4.11 (Variance of the estimator of the variance).

$$\mathrm{Var}\left(\widehat{\mathrm{Var}[X]}\right) = \frac{(N-1)^2}{N^3}\langle(x - \langle x\rangle)^4\rangle - \frac{(N-1)(N-3)}{N^3}\langle(x - \langle x\rangle)^2\rangle^2 \tag{45.85}$$

Formula 45.4.12 (Variance of the estimator of the standard deviation).

$$\mathrm{Var}(\widehat{\sigma}) = \frac{1}{4\sigma^2}\mathrm{Var}\left(\widehat{\mathrm{Var}[X]}\right) \tag{45.86}$$

Remark 45.4.13. 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.

45.4.4 Likelihood function

Definition 45.4.14 (Likelihood). The likelihood $\mathcal{L}(a;\mathbf{x})$ is the joint density of a set of measurements $\mathbf{x} := \{x_1,\ldots,x_N\}$:

$$\mathcal{L}(a;\mathbf{x}) = \prod_{i=1}^{N} f(x_i;a). \tag{45.87}$$

Theorem 45.4.15 (Cramer-Rao bound). *The variance of an* ***unbiased*** *estimator has a lower bound called the Cramer-Rao bound or* ***minimum variance bound (MVB)****:*

$$\mathrm{Var}(\hat{a}) \geq \frac{1}{\left\langle\left(\frac{d\ln\mathcal{L}}{da}\right)^2\right\rangle}. \tag{45.88}$$

For a biased estimator with bias b, the MVB takes on the following form:

$$\mathrm{Var}(\hat{a}) \geq \frac{\left(1+\frac{db}{da}\right)^2}{\left\langle\left(\frac{d\ln\mathcal{L}}{da}\right)^2\right\rangle}. \tag{45.89}$$

Remark 45.4.16.

$$\left\langle\left(\frac{d\ln\mathcal{L}}{da}\right)^2\right\rangle = -\left\langle\frac{d^2\ln\mathcal{L}}{da^2}\right\rangle \tag{45.90}$$

Definition 45.4.17 (Fisher information).

$$I_X(a) := \left\langle\left(\frac{d\ln\mathcal{L}}{da}\right)^2\right\rangle = N\int\left(\frac{d\ln f}{da}\right)^2 f\,d\mu \tag{45.91}$$

Using this definition one can rewrite the Cramer-Rao inequality as follows:

$$\mathrm{Var}(\hat{a}) \geq I_X(a). \tag{45.92}$$

Definition 45.4.18 (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:

$$e(\hat{a}) := \frac{I_X(a)^{-1}}{\mathrm{Var}(\hat{a})}. \tag{45.93}$$

45.4.5 Maximum likelihood estimation

From Definition 45.4.14 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's 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, does 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 45.4.19 (Maximum likelihood estimator). The maximum likelihood estimator $\hat{a}$ is obtained by solving the following equation:

$$\left.\frac{d\ln\mathcal{L}}{da}\right|_{a=\hat{a}} = 0. \tag{45.94}$$

Remark 45.4.20. MLE estimators are mostly consistent but often biased.

Property 45.4.21. MLE estimators are invariant under parameter transformations.

Corollary 45.4.22. The invariance implies that the two estimators $\hat{a}$ and $\widehat{f(a)}$ cannot both be unbiased at the same time.

Property 45.4.23. Every consistent estimator asymptotically becomes unbiased and efficient.

Property 45.4.24 (Minimizing KL-divergence). It can be shown that maximizing the log-likelihood is equivalent to minimizing the Kullback-Leibler divergence 43.7.3 between the would-be density $f(x;\theta)$ and the true density $q(x)$:

$$\begin{aligned}\arg\max_\theta \ln\mathcal{L} &= \arg\max_\theta \sum_{i\in I}\ln f(x_i;\theta)\\ &= \arg\max_\theta \sum_{i\in I}\ln f(x_i;\theta) - \ln q(x_i)\\ &= \arg\min_\theta \frac{1}{n}\sum_{i\in I}\ln\frac{q(x_i)}{f(x_i;\theta)}\\ &\longrightarrow \arg\min_\theta \int q(x;\theta)\ln\frac{q(x)}{f(x;\theta)}\,dx = \arg\min_\theta D_{\mathrm{KL}}(q\|f_\theta),\end{aligned}$$

where the law of large numbers was used in the last line.

45.4.6 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 45.4.25 (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):

$$\chi^2 := \sum_i \frac{d_i^2}{f_i}, \tag{45.95}$$

   where $f_i = f(x_i;a)$.

3. Find the most probable value of $\hat{a}$ by solving the equation

$$\frac{d\chi^2}{da} = 0. \tag{45.96}$$

Property 45.4.26. The optimal χ^2-value is asymptotically distributed according to a χ^2-distribution with n degrees of freedom. The parameter n is equal to the number of events N minus the number of fitted parameters k. (See more in Section 45.8.1.)

Formula 45.4.27 (Linear fit). When all uncertainties σ_i are equal, the slope $\hat{m}$ and intercept $\hat{c}$ are given by the following formulas:

$$\hat{m} := \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{\mathrm{cov}(x,y)}{\mathrm{Var}[x]} \tag{45.97}$$

$$\hat{c} := \overline{y} - \hat{m}\overline{x} = \frac{\overline{x^2} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2}. \tag{45.98}$$

Remark 45.4.28. The equation $\overline{y} = \hat{c} + \hat{m}\overline{x}$ says that the linear fit passes through the center of mass $(\overline{x}, \overline{y})$.

Formula 45.4.29 (Errors of linear fit).

$$\mathrm{Var}[\hat{m}] = \frac{1}{N(\overline{x^2} - \overline{x}^2)}\sigma^2 \tag{45.99}$$

$$\mathrm{Var}[\hat{c}] = \frac{\overline{x^2}}{N(\overline{x^2} - \overline{x}^2)}\sigma^2 \tag{45.100}$$

$$\mathrm{cov}(\hat{m}, \hat{c}) = \frac{-\overline{x}}{N(\overline{x^2} - \overline{x}^2)}\sigma^2 \tag{45.101}$$

The least squares method is very useful to fit data that has been grouped in bins (histograms):

Method 45.4.30 (Binned least squares).

1. N i.i.d. events with density $f(X; a)$ divided in N_B intervals, where the interval j is centered on the value x_j, has a width W_j and contains n_j events.
2. The ideally expected number of events in the j^{th} interval: $f_j = NW_jf(x_j; a)$.
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

$$\chi^2 := \sum_i^{N_B} \frac{(n_i - f_i)^2}{f_i^2}. \tag{45.102}$$

45.4.7 Geometric approach

Consider a sample $\mathbf{x} := \{x_1, \ldots, x_n\}$ drawn from a distribution $f(x; \theta)$ in an exponential family. The likelihood 45.4.14 is given by

$$\mathcal{L}(\theta; \mathbf{x}) := \prod_{i=1}^n f(x_i; \theta).$$

The m-coordinates of the observed point are

$$\eta = \frac{1}{n}\sum_{i=1}^n x_i = \overline{x}. \tag{45.103}$$

The optimal value for θ can be found by maximizing the log-likelihood as before. In Property 45.4.24 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 45.4.24. Minimization then corresponds to m-projecting the observed point η on the submanifold S of "admissible" distributions.

Theorem 45.4.31 (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* (45.2.4). *Further,* *let Γ be a collection of probability distributions such that $P_n \in \Gamma$. The joint distribution q^n satisfies the following inequality:*

$$q^n(P_n \in \Gamma) \leq (n+1)^{|S|} 2^{-nD_{\mathrm{KL}}(p^*\|q)}, \tag{45.104}$$

where p^* is the information projection of q on Γ. If $\Gamma = \overline{\Gamma^\circ}$, this can be restated as

$$\lim_{n\to\infty} \frac{1}{n} q^n(P_n \in \Gamma) = -D_{\mathrm{KL}}(p^*\|q). \tag{45.105}$$

45.5 Bayesian modelling

Definition 45.5.1 (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 45.5.2. The simplest 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*.

45.6 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 45.6.1 (Prediction interval). Let X be a normally distributed random variable. A measurement 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.

45.6.1 Interval types

Definition 45.6.2 (Two-sided confidence interval).

$$\Pr(x_- \leq X \leq x_+) = \int_{x_-}^{x_+} f(x)\,dx = C \tag{45.106}$$

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 45.6.3. For Gaussian distributions these three definitions are equivalent.

Definition 45.6.4 (One-sided confidence interval).

$$\Pr(x \geq x_-) = \int_{x_-}^{\infty} f(x)\,dx = C \tag{45.107}$$

$$\Pr(x \leq x_+) = \int_{-\infty}^{x_+} f(x)\,dx = C \tag{45.108}$$

Definition 45.6.5 (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 $[x_-, x_+]$.

$$x_- = \arg\min_\theta \left[\frac{1-C}{2} - \sum_{x=0}^{\theta-1} p(x)\right] \tag{45.109}$$

$$x_+ = \arg\min_\theta \left[\frac{1-C}{2} - \sum_{x=\theta+1}^{\infty} p(x)\right] \tag{45.110}$$

45.6.2 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 45.6.6. Let x_0 be a point estimate of the parameter X. From the confidence region it is possible to infere 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.

45.6.3 Interval for a sample mean

Formula 45.6.7 (Interval with known variance). If the sample size is large enough, the real distribution is unimportant, because the CLT ensures a Gaussian distribution of the sample mean $\overline{X}$. The α-level confidence interval such that $\Pr(-z_{\alpha/2} < Z < z_{\alpha/2})$ with $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{N}}$ is given by

$$\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{N}}, \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{N}}\right]. \tag{45.111}$$

Remark 45.6.8. If the sample size is not sufficiently large, the measured quantity must follow a normal distribution.

Formula 45.6.9 (Interval with unknown variance). To account for the uncertainty of the estimated standard deviation $\hat{\sigma}$, the student-t distribution 45.2.35 is used instead of a Gaussian distribution to describe the sample mean $\overline{X}$. The α-level confidence interval is given by

$$\left[\overline{X} - t_{\alpha/2;(n-1)}\frac{s}{\sqrt{N}}, \overline{X} + t_{\alpha/2;(n-1)}\frac{s}{\sqrt{N}}\right], \tag{45.112}$$

where s is the estimated standard deviation 45.4.9.

Formula 45.6.10 (Wilson score interval). For a sufficiently large sample, a sample proportion $\hat{P}$ is approximately Gaussian distributed with expectation value π and variance $\frac{\pi(\pi-1)}{N}$. The α-level confidence interval is given by

$$\left[\frac{(2N\hat{P}+z_{\alpha/2}^2)-z_{\alpha/2}\sqrt{z_{\alpha/2}^2+4N\hat{P}(1-\hat{P})}}{2(N+z_{\alpha/2}^2)}, \frac{(2N\hat{P}+z_{\alpha/2}^2)+z_{\alpha/2}\sqrt{z_{\alpha/2}^2+4N\hat{P}(1-\hat{P})}}{2(N+z_{\alpha/2}^2)}\right]. \tag{45.113}$$

Remark. The expectation value and variance are these of a binomial distribution 45.2.28 with $r = X/N$.

45.7 Hypothesis testing

Definition 45.7.1 (Simple hypothesis). A hypothesis where the distribution is fully specified.

Definition 45.7.2 (Composite hypothesis). A hypothesis where the distribution is given relative to some parameter values.

45.7.1 Testing

Definition 45.7.3 (Type I error). Rejection of a true null hypothesis.

Definition 45.7.4 (Type II error). Acceptance of a false null hypothesis.

Definition 45.7.5 (Significance). The probability of making a type I error:

$$\alpha := \int P_{\mathrm{I}}(x)\,dx. \tag{45.114}$$

Property 45.7.6. Let $\alpha_1 > \alpha_2$. An α_2-level test is also significant at the α_1-level.

Remark 45.7.7. 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 45.7.8 (Power). The probability of not making a type II error:

$$\beta := \int P_{\mathrm{II}}(x)\,dx \quad\longrightarrow\quad \text{power: } 1-\beta. \tag{45.115}$$

Remark 45.7.9. 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 45.7.10 (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:

$$W := \frac{\sum_{i=1}^n (a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \overline{x})^2}, \tag{45.116}$$

where

- $x_{(i)}$ are the order statistics,

- $(a_i, \ldots, a_n) := \frac{m^T V}{\|V^{-1}m\|}$,
- $m \equiv (m_1, \ldots, m_n)$ 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 45.7.11 (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:

$$\Lambda(x) = \frac{\mathcal{L}(\theta_0 \mid x)}{\mathcal{L}(\theta_1 \mid x)} \leq \eta, \tag{45.117}$$

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 45.7.12 (Neyman-Pearson lemma). *The likelihood ratio test is the most powerful test at significance level α.*

Definition 45.7.13 (Family-wise error). Given a collection of hypothesis tests, the family-wise error is defined as the probability of making at least one type-I error.

Construction 45.7.14 (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:

$$\alpha \longrightarrow \frac{\alpha}{n}. \tag{45.118}$$

45.7.2 Comparison tests

Definition 45.7.15 (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

(45.119)

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:

$$\begin{aligned} H_0 &: b = c \\ H_1 &: b \neq c. \end{aligned}$$

The test statistic is the McNemar chi-squared statistic:

$$\chi^2 = \frac{(b-c)^2}{b+c}. \tag{45.120}$$

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 45.7.16 (Edwards correction). It is common to apply a continuity correction (similar to the *Yates-correction* for the ordinary chi-squared test):

$$\chi^2 := \frac{(|b-c|-1)^2}{b+c}. \tag{45.121}$$

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):

$$p = 2\sum_{i=b}^{b+c}\binom{b+c}{i}0.5^i(1-0.5)^{b+c-i}. \tag{45.122}$$

Definition 45.7.17 (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. The test statistic is defined as follows:

> 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:
>
> $$T := \min(R_+, R_-). \tag{45.123}$$
>
> 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
>
> $$z := \frac{T - \frac{1}{4}n(n+1)}{\sqrt{\frac{1}{24}n(n+1)(2n+1)}}.$$

Remark 45.7.18. 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 45.7.19 (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 45.7.20 (Friedman test). Consider k 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 i^{th} model on the j^{th} data set. Under the null hypothesis "all models perform equally well", the average ranks should be the same for all models.

The Friedman statistic

$$\chi_F^2 := \frac{12N}{k(k+1)}\left(\sum_{i\leq k} R_i^2 - \frac{k(k+1)^2}{4}\right) \tag{45.124}$$

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 45.7.21. It was shown that the original Friedman test is rather conservative and that a better statistic is

$$F := \frac{(N-1)\chi_F^2}{N(k+1)-\chi_F^2}. \tag{45.125}$$

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).

45.7.3 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 45.7.22 (Nemenyi test). Consider the average ranks R_i from the Friedman test. As a test statistic one uses

$$z := \frac{R_i - R_j}{\sqrt{\frac{k(k+1)}{6N}}}, \tag{45.126}$$

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 45.7.23 (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 45.7.24 (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 45.7.25 (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.

45.8 Goodness of fit

Definition 45.8.1 (Akaike information criterion). Consider a model $f(x;\theta)$ with k 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:

$$\mathrm{AIC} := 2k - 2\ln(\mathcal{L}_0). \tag{45.127}$$

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.

45.8.1 χ^2-test

Property 45.8.2. If there are $N-n$ fitted parameters one has:

$$\int_{\chi^2}^{\infty} f_{\chi^2}(x \mid n)\,dx \approx 1 \implies \begin{cases} \bullet \text{ good fit} \\ \bullet \text{ errors were overestimated} \\ \bullet \text{ selected measurements} \\ \bullet \text{ lucky shot} \end{cases} \tag{45.128}$$

Property 45.8.3 (Reduced χ^2). The reduced chi-squared statistic is defined as follows:

$$\chi^2_{\text{red}} := \chi^2/n, \tag{45.129}$$

where 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.

45.8.2 Runs test

A good χ^2-test does not mean that the fit is good. As mentioned in Property 45.8.2, 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 45.8.5.

Remark 45.8.4. 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 45.8.5 (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):

$$\begin{aligned}
P(r_{\text{even}}) &= 2\frac{\binom{N_+-1}{\frac{r}{2}-1}\binom{N_--1}{\frac{r}{2}-1}}{\binom{N}{N_+}} \\
P(r_{\text{odd}}) &= \frac{\binom{N_+-1}{\frac{r-3}{2}}\binom{N_--1}{\frac{r-1}{2}} + \binom{N_--1}{\frac{r-3}{2}}\binom{N_+-1}{\frac{r-1}{2}}}{\binom{N}{N_+}},
\end{aligned} \tag{45.130}$$

where C^n_k is the binomial coefficient $\binom{n}{k}$. The first two moments of this distribution are given by the following formulas:

$$\mathrm{E}[r] = 1 + 2\frac{N_+N_-}{N} \tag{45.131}$$

$$\mathrm{Var}[r] = 2\frac{N_+N_-}{N}\frac{2N_+N_- - N}{N(N-1)}. \tag{45.132}$$

Remark 45.8.6. For $r > 15$, the runs distribution approximates a Gaussian distribution.

Chapter 46

Data Analysis

The main reference for the sections on optimization problems is [14]. For the geometry of clustering methods, see [15]. The main references for the section on *conformal prediction* are [16, 17]. Although a part of this chapter is a continuation of the previous one, 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 20).

46.1 Data sampling

46.1.1 Inverse CDF sampling

Although one of the most straightforward sampling algorithms, this approach makes the strong assumption that the cumulative distribution function 43.3.4 is invertible.

Method 46.1.1. Sample a point λ uniformly from the unit interval $[0, 1]$. This number gives the cumulative distribution of the point to be sampled. (The CDF $F_X(X)$ is itself uniformly distributed.) The new point x' is simply given by $F_X^{-1}(\lambda)$.

In the case where 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 $x \in \mathbb{R}$ and its associated cumulative probability $F_X(x)$ one can sample a new point x' as follows. One increases (or decreases) x until the unique point x' is found such that $F_X(x' - 1) < \lambda \leq F_X(x')$.

46.1.2 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 46.1.2. Consider an interval $[a, b]$ such that f_X vanishes outside this interval and let q_0 be an upper bound for f_X. Now, sample a point x' uniformly on $[a, b]$ and sample a point q uniformly on $[0, q_0]$. If $f_X(x') \geq q$, then x' is a good sample. If not, repeat this procedure.

The proof of that this algorithm works is quite easy and mainly depends on the fact that

$$\Pr(Q \leq f_X(X)) = \frac{1}{q_0(b-a)}.$$

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.

46.1.3 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 46.1.3 (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' from the normal distribution.
- Calculate the **acceptance ratio**

$$\lambda := \frac{f_X(x')}{f_X(x_{i-1})}.$$

- If $\lambda \geq 1$, $x_i = x'$.
- If not, sample a point q uniformly on $[0, 1]$. If $\lambda \geq q$, $x_i = x'$, else $x_i = x_{i-1}$.[1]

To obtain an efficient algorithm, it is helpful to choose $\mu = \mathrm{E}[X]$ and $\sigma^2 = \mathrm{Var}[X]$. This ensures that the points are sampled in a region that resembles the form of f_X.

In fact, this method can be drastically generalized. First of all it is possible to replace the normal distribution by any symmetric transition probability:

$$g(x' \mid x) = g(x \mid x'). \tag{46.1}$$

In case the transition probability is not symmetric, the acceptance ratio needs to be modified:

$$\lambda := \frac{f_X(x')g(x_{i-1} \mid x')}{f_X(x_{i-1})g(x' \mid x_{i-1})}. \tag{46.2}$$

Remark 46.1.4. 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.

46.2 Optimization

46.2.1 Linear equations

Method 46.2.1 (Normal equation). Given the equation

$$Ax = b$$

as in Section 20.4.1, one can try to numerically solve for x by minimizing the ℓ^2-norm $\|Ax - b\|^2$:

$$\hat{x} := \arg\min_x (Ax - b)^T(Ax - b). \tag{46.3}$$

This leads to the so-called normal equation[2]

$$A^TAx = A^Tb. \tag{46.4}$$

This can be 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 46.2.2. 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 46.2.3 (Tikhonov regularization). Consider a linear (regression) problem

$$Ax = b.$$

The most straightforward way to solve for x is the least squares method introduced in Chapter 45, 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:

$$\|Ax - b\|^2 + \|\Gamma x\|^2, \tag{46.5}$$

where Γ is called the **Tikhonov matrix**. In the case that $\Gamma = \lambda\mathbb{1}$, one speaks of **$\ell^2$-regularization**. This regularization technique benefits solutions with smaller norms.

Remark 46.2.4. 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. 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.

A benefit of ℓ^2-regularization is that it can be derived from a Bayesian approach. By choosing a Gaussian prior $\mathcal{N}(0, \lambda^{-1})$, Bayesian inference immediately gives the ℓ^2-regularized cost function as the posterior distribution. 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 generalized normal distributions (such as the Laplace distribution for $p = 1$).

Definition 46.2.5 (Multicollinearity). Consider a finite set of random variables $\{X_i\}_{1\leq i\leq n}$. 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

$$\lambda_0 + \lambda_1X_1 + \cdots + \lambda_nX_n = 0. \tag{46.6}$$

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 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 46.2.6 (Variance inflation factor). The VIF is an estimate for how much the variance of a coefficient is inflated by multicollinearity. The VIF of a coefficient β_i is defined as follows:

$$\mathrm{VIF}_i := \frac{1}{1 - R_i^2}, \tag{46.7}$$

where R_i^2 is the R^2-value obtained after regressing the predictor $\hat{X}_i$ on all other predictors. The rule of thumb is that $\mathrm{VIF} \geq 10$ implies that a significant amount of multicollinearity is present in the model.

46.2.2 Gradient descent

The gradient descent algorithm is first introduced in the case of quadratic forms:

Method 46.2.7 (Steepest descent). Consider the quadratic form

$$f(x) = \frac{1}{2}x^T Ax - b^T x + c.$$

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 $\nabla f(x_0) = Ax_0 - b =: -r_0$:

$$x_{i+1} := x_i + \alpha r_i. \tag{46.8}$$

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:

$$\alpha_i = \frac{r_i^T r_i}{r_i^T Ar_i}. \tag{46.9}$$

This choice forces the descent direction to be orthogonal to the previous one since

$$\frac{d}{d\alpha}f(x_i) = -\nabla f(x_i) \cdot \nabla f(x_{i-1}).$$

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.

46.2.3 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 occur after a long time.

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 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 46.2.8 (Conjugate vectors). Consider a symmetric positive-definite matrix A. Any such matrix induces an inner product as follows:

$$\langle v|w\rangle_A := v^T Aw. \tag{46.10}$$

Two vectors v, w are said to be (A-)conjugate if they are orthogonal with respect to $\langle\cdot|\cdot\rangle_A$. The general approach to obtain a basis of A-conjugate vectors is a modified version of the Gram-Schmidt procedure 20.3.15 where the ordinary Euclidean inner product is replaced by (46.10). 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:

$$r_{i+1} = r_i - \alpha_i A d_i, \tag{46.11}$$

where the step size α_i is defined similar to the step size for ordinary steepest descent:

$$\alpha_i = \frac{d_i^T r_i}{d_i^T A d_i}. \tag{46.12}$$

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 $r_{i+1} \in \mathcal{D}_i + Ad_i$ leads to the following expression because of the above recursion relation:

$$\mathcal{D}_i = \text{span}\big\{r_0, Ar_0, \ldots, A^{i-1}r_0\big\}. \tag{46.13}$$

Because of their prominence in the literature on numeric optimization techniques, these subspaces have earned their own name:

Definition 46.2.9 (Krylov subspace). A vector space $\mathcal{K}$ of the form

$$\mathcal{K} := \text{span}\{v, Av, \ldots, A^n v\} \tag{46.14}$$

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 time-wise and memory-wise) per iteration from $O(n^2)$ to $O(n)$.

The steps in the CG algorithm are summarized below:

Method 46.2.10 (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 being the dimension of the coefficient matrix A) algorithm to obtain the solution to $Ax = b$:

$$\alpha_i := \frac{r_i^T r_i}{d_i^T A d_i} \tag{46.15}$$

$$x_{i+1} := x_i + \alpha_i d_i \tag{46.16}$$

$$r_{i+1} := r_i - \alpha_i A d_i \tag{46.17}$$

$$d_{i+1} := r_{i+1} + \frac{r_{i+1}^T r_{i+1}}{r_i^T r_i} d_i. \tag{46.18}$$

Remark 46.2.11. In exact arithmetic the above optimization scheme would result in an exact solution after 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 one has to take into account the occurrence of floating-point errors. These not only ruin the accuracy of the residual recursion relation (46.17), but more importantly[3] it might result in the search directions not being A-conjugate.

Now, what about general coefficient matrices A, for example those resulting in under- or overdetermined systems? For nonsymmetric or nondefinite square matrices one can still solve the normal equation (46.4) using the same methods, since A^TA 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^TA will be nonsingular and the 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 are used A, A^T directly. This way not one Krylov space is generated, but two dual ‘‘copies'' are constructed:

$$\begin{aligned}\mathcal{D}_i &:= \mathrm{span}\{r_0, Ar_0, \ldots, A^{i-1}r_0\}\\ \widetilde{\mathcal{D}}_i &:= \mathrm{span}\{\widetilde{r}_0, A^T\widetilde{r}_0, \ldots, (A^T)^{i-1}\widetilde{r}_0\},\end{aligned}$$

where $\widetilde{r}_0$ does not have to be related to r_0. In this case there are two Galerkin conditions $r_i \perp \mathcal{D}_i$ and $\widetilde{r}_i \perp \widetilde{\mathcal{D}}_i$ (only the first one is relevant). The residuals form biorthogonal bases of the Krylov subspaces:

$$\langle r_i|r_j\rangle = \|r_i\|^2\delta_{ij}. \tag{46.19}$$

As a consequence the search directions also form biconjugate bases:

$$\langle d_i|d_j\rangle_A = \|d_i\|_A^2\delta_{ij}. \tag{46.20}$$

46.2.4 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.

On 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 Equation (46.18). 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 46.2.12 (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:

$$\alpha_i := \arg\min_\alpha f(x_i + \alpha d_i) \tag{46.21}$$

$$x_{i+1} := x_i + \alpha_i d_i \tag{46.22}$$

$$r_{i+1} := -\nabla f(x_i) \tag{46.23}$$

$$d_{i+1} := r_{i+1} + \beta_{i+1}d_i, \tag{46.24}$$

[3]The residual problem can be solved by computing the residual ‘‘exactly'', i.e. by the formula $r_i = b - Ax_i$, every k iterations.

where β_{i+1} is computed by one of the following formulas:

- **Fletcher-Reeves formula**:

$$\beta_{i+1} := \frac{r_{i+1}^T r_{i+1}}{r_i^T r_i}. \tag{46.25}$$

- **Polak-Ribière formula**:

$$\beta_{i+1} := \max\left\{\frac{r_{i+1}^T(r_{i+1} - r_i)}{r_i^T r_i}, 0\right\}. \tag{46.26}$$

Some general remarks have to be made concerning the nonlinear CG algorithm:

Remark 46.2.13. 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 quicker 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 multiple local minima. The CG method does not care about local vs. global 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 Equation (46.21) as easily. The main idea, i.e. that f' should be orthogonal to the previous search direction remains, is still valid. Here, only the **Newton-Raphson approach** is considered:[5]

$$\alpha_i = \frac{\nabla f(x_i)^T d_i}{d_i^T \mathrm{Hess} f(x_i) d_i}. \tag{46.27}$$

To obtain the optimal α-value, one should iteratively apply the Newton-Raphson method in every CG iteration. If the action of the Hessian 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 descent 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 it is negative.

46.2.5 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

$$x_i = x_0 + \sum_{k=0}^{i-1} r_k = x_0 + \sum_{k=0}^{i-1} (\mathbb{1} - A)^k r_0. \tag{46.28}$$

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 $x_i - x_0 = \sum_{k=0}^{i-1} P_k(A) r_0$.

[4] The max operation in Equation (46.26) 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*.

Method 46.2.14 (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 zero diagonal elements. If one assumes that D is invertible, the following recursive scheme is obtained:

$$x_{i+1} := D^{-1}(b - Ex_i). \tag{46.29}$$

A sufficient condition for convergence is strict diagonal dominance, i.e. $|D_{ii}| > \sum_{j\neq i} |E_{ij}|$.

?? COMPLETE (e.g. Lanczos)??

46.3 Constrained optimization

46.3.1 Lagrange multipliers

A common generalization of the above optimization problems is the addition of constraints involving equalities:

$$\arg\min_x f(x) \qquad \text{such that} \qquad g_i(x) = 0 \qquad \forall 1 \leq i \leq n. \tag{46.30}$$

The general approach to solving such constrained problems is by extending the optimization loss:

Method 46.3.1 (Lagrange multipliers). Given a constrained optimization problem of the form (46.30), one can construct the enhanced loss function

$$\mathcal{L}(x, \lambda_1, \ldots, \lambda_n) := f(x) + \sum_{i=1}^{n} \lambda_i g_i(x). \tag{46.31}$$

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):

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial x} = 0 \\ \\ \dfrac{\partial \mathcal{L}}{\partial \lambda_i} = 0 \quad \forall 1 \leq i \leq n. \end{cases} \tag{46.32}$$

The situation becomes even more interesting when one also allows constraints involving inequalities:

$$\arg\min_x f(x) \qquad \text{such that} \qquad \begin{cases} g_i(x) = 0 & \forall 1 \leq i \leq m \\ h_j(x) \leq 0 & \forall 1 \leq j \leq n. \end{cases} \tag{46.33}$$

Problems of this form are called **primal optimization problems**. By defining an enhanced loss using Lagrange multipliers as before

$$\mathcal{L}(x, \alpha, \beta) := f(x) + \sum_{i=1}^{m} \alpha_i g_i(x) + \sum_{i=1}^{n} \beta_i h_i(x), \tag{46.34}$$

it is not hard to see that

$$\max_{\alpha,\beta;\beta_j\geq 0} \mathcal{L}(x, \alpha, \beta) = \begin{cases} \infty & \text{if a constraint is violated} \\ f(x) & \text{if all constraints are satisfied.} \end{cases} \tag{46.35}$$

Definition 46.3.2 (Primal optimization problem). Denote the maximum of $\mathcal{L}(x, \alpha, \beta)$ by $\theta_P(x)$.

$$p^* := \min_x \theta_P(x) = \min_x \max_{\alpha,\beta;\beta_i \geq 0} \mathcal{L}(x, \alpha, \beta). \tag{46.36}$$

By interchanging the max and min operators in the primal formulation, another problem is obtained:

Definition 46.3.3 (Dual optimization problem).

$$d^* := \max_{\alpha,\beta;\beta_i \geq 0} \theta_D(\alpha, \beta) = \max_{\alpha,\beta;\beta_i \geq 0} \min_x \mathcal{L}(x, \alpha, \beta). \tag{46.37}$$

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 46.3.4 (Slater conditions). Consider a convex optimization problem, i.e. a problem of the form (46.33) 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 46.3.5 (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 optima and dual optima respectively.

The following property gives a set of sufficient conditions:

Property 46.3.6 (Karush-Kuhn-Tucker conditions). If there exist x, α and β such that strong duality holds, the following conditions are satisfied:

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial x} = 0 \\[2ex] \dfrac{\partial \mathcal{L}}{\partial \alpha_i} = 0 \end{cases} \quad \forall 1 \leq i \leq m \qquad \text{and} \qquad \begin{cases} \beta_j h_j(x) = 0 \\ h_j(x) \leq 0 \\ \beta_j \geq 0. \end{cases} \quad \forall 1 \leq j \leq n \tag{46.38}$$

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 46.3.7 (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 46.3.8. 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**.

46.3.2 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 we have to treat is the occurence 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 is a covariant vector. However, a short proof shows that the Riemannian gradient from Remark 34.1.9 actually gives the direction of steepest ascent. So even on Riemannian manifolds the gradient is the correct direction to work with. However, as mentioned above, the form of the update will be a problem in general.

?? COMPLETE ??

46.4 Approximation theory

46.4.1 Bayes optimality

Definition 46.4.1 (Bayes risk). The minimal risk over all models:

$$R^* := \inf_{f:\mathcal{X}\rightarrow\mathcal{Y}} R(f). \tag{46.39}$$

Definition 46.4.2 (Bayes classifier). Given a joint probability distribution P over the instance space $\mathcal{X}\times\mathcal{Y}$ and a loss function $l:\mathcal{Y}\times\mathcal{Y}\rightarrow\mathbb{R}$, the pointwise Bayes predictor is defined as follows:

$$f^* : x \mapsto \arg\min_{y\in\mathcal{Y}} \int_{\mathcal{Y}} l(y,y')dP(y' \mid x). \tag{46.40}$$

Property 46.4.3 (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 46.4.4 (Approximation error). Given a data set $\mathcal{D}\subset\mathcal{X}\times\mathcal{Y}$, the empirical risk R_{emp} is defined as follows:

$$R_{\text{emp}} : \hat{y} \mapsto \frac{1}{|\mathcal{D}|}\sum_{(x,y)\in\mathcal{D}} l(\hat{y}(x),y). \tag{46.41}$$

Consider a hypothesis space $\mathcal{H}\subseteq\mathcal{Y}^{\mathcal{X}}$ (e.g. selected 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.

46.4.2 PAC theory and empirical risk minimization

Property 46.4.5. Note that by the (strong) law of large numbers 45.1.4, the empirical risk converges to the true risk almost surely whenever the data poitns are sampled i.i.d. However, in practice, the true question becomes how fast this convergence happens.

Definition 46.4.6 (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

$$\Pr\left(R(\hat{f}) - \inf_{f\in\mathcal{H}} R(f) > \varepsilon\right) < \delta. \tag{46.42}$$

Corollary 46.4.7. If one can prove a PAC bound, one gets the following bound for free:

$$R(\hat{f}) \leq \inf_{f\in\mathcal{H}} R(f) + 2\varepsilon \tag{46.43}$$

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 reasonbaly good estimate.

Property 46.4.8 (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:

$$\begin{aligned}\Pr(\exists f \in \mathcal{H} : |R_{\text{emp}}(f) - R(f)| \geq \varepsilon) &= \Pr(\sup_{f\in\mathcal{H}} |R_{\text{emp}}(f) - R(f)| \geq \varepsilon)\\ &\leq \sum_{f\in\mathcal{H}} \Pr(|R_{\text{emp}}(f) - R(f)| \geq \varepsilon)\\ &\leq \sum_{f\in\mathcal{H}} 2e^{-2n\varepsilon^2}\\ &= 2|\mathcal{H}|e^{-2n\varepsilon^2},\end{aligned} \tag{46.44}$$

where the second inequality comes from Hoeffding's inequality 43.4.14 (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

$$R(f) \leq R_{\text{emp}}(f) + \sqrt{\frac{\ln(|\mathcal{H}|) + \ln(1/\delta)}{2n}}, \tag{46.45}$$

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 $\hat{f}$:

$$\mathrm{E}\Big[R(\hat{f})\Big] \leq \inf_{f\in\mathcal{H}} R(f) + \sqrt{\frac{\ln(|\mathcal{H}|) + \ln(1/\delta)}{2n}} + \delta, \tag{46.46}$$

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 46.4.9 (Complexity regularizer). A function $c : X \to \mathbb{R}^+$ such that

$$\sum_{x\in X} e^{-c(x)} \leq 1. \tag{46.47}$$

A simple example would be the log-probabilities $c(x) := \ln(P(x))$ when a probability distribution P on X is given.

Using a complexity regularizer on $\mathcal{H}$, the following risk bound can be obtained with probability $1-\delta$:

$$R(f) \leq R_{\text{emp}}(f) + \sqrt{\frac{c(f) + \ln(1/\delta)}{2n}}. \tag{46.48}$$

46.5 Classification problems

46.5.1 Clustering

Probably the most well-known and simplest algorithm for clustering in the unsupervised setting is the k-means algorithm:

Method 46.5.1 (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:
$$i = \arg\min_{j\leq k} \|d - c_j\|^2. \tag{46.49}$$
2. Update the centroids c_i to represent the center of mass of the associated cluster C_i:
$$c_i \longleftarrow \frac{1}{|C_i|}\sum_{d\in C_i} d. \tag{46.50}$$

This algorithm optimizes the following global cost function with respect to the centroids c_i:

$$\mathcal{L}_{k\text{-means}}(c_1,\ldots,c_k) = \sum_{i=1}^k \sum_{d\in C_i} \|d - c_i\|^2. \tag{46.51}$$

Given the above idea, one could ask for a more general algorithm where clustering is performed with respect to a divergence function 44.1.7. In the case of Bregman divergences 44.1.10 it can be shown that all one needs to do is replace the Euclidean distance by the divergence D_f:

Property 46.5.2 (Centroid position). Let D_f be a Bregman divergence. The minimizer

$$\arg\min_\kappa \sum_{i=1}^k D_f(x_i\|\kappa) \tag{46.52}$$

is given by the arithmetic average

$$\kappa = \frac{1}{k}\sum_{i=1}^k x_i. \tag{46.53}$$

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. It can be furthermore 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 46.5.3 (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]

[6]See [18] for more information. This is also introduced in [19], but there the author has confusingly interchanged the affine and dual coordinates.

46.5.2 Nearest neighbour search

?? COMPLETE ??

46.6 Garden

?? ADD (e.g. trees, forests) ??

46.7 Support-vector machines

46.7.1 Kernel methods

This section will introduce the mathematics of kernel methods. This mainly involves the language of Hilbert spaces (see Chapter 23 for a refresher).

Definition 46.7.1 (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 46.7.2 (Reproducing kernel Hilbert space). A Hilbert space $\mathcal{H} \subset \mathrm{Map}(X, \mathbb{C})$ of functions over a set X for which all evaluation functionals $\delta_x : f \mapsto f(x)$ are bounded (or continuous by Property 23.4.8). Reproducing kernel Hilbert spaces are often abbreviated as **RKHS**s.

Using the Riesz representation theorem 23.2.7 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 X:

Definition 46.7.3 (Reproducing kernel). Let $\mathcal{H}$ be an RKHS on a set X. The (reproducing) kernel k on X is defined as follows:

$$k(x, y) := \delta_x(K_y) \overset{\text{Riesz}}{=} \langle K_x | K_y \rangle_{\mathcal{H}}. \tag{46.54}$$

Because k is given by an inner product, it is not hard to see that the reproducing kernel is a kernel 46.7.1.

Starting from a kernel one can also characterize an RKHS as follows:

Alternative Definition 46.7.4 (RKHS). A Hilbert space $\mathcal{H} \subset \mathrm{Map}(X, \mathbb{C})$ of functions over a set X such that there exists a kernel k on X with the following properties:

1. **Reproducing property**: For all $x \in X, f \in \mathcal{H}$ the evaluation functional δ_x satisfies $\delta_x(f) = \langle k(\cdot, x) | f \rangle_{\mathcal{H}}$.
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 46.7.5 (Convergence). In an RKHS, convergence in norm implies pointwise convergence.

Theorem 46.7.6 (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 $\mathrm{span}\{K_x \mid x \in X\}$, where the inner product is

[7] Also called a **Mercer kernel**. See Mercer's theorem below for more information.

defined as follows

$$\left\langle \sum_{x\in X} a_x K_x \middle| \sum_{y\in X} b_y K_y \right\rangle := \sum_{x,y\in X} \overline{a_x} b_y k(x,y). \tag{46.55}$$

Formula 46.7.7. 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

$$k(x,y) = \sum_{i=1}^{\dim(\mathcal{H})} e_i(x)\overline{e_i(y)}. \tag{46.56}$$

Remark 46.7.8. Note that one can use different conventions in the above definitions, e.g. the definition $k(x,y) = \langle K_y|K_x\rangle_{\mathcal{H}}$ is also valid.

Theorem 46.7.9 (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***

$$\iint_{X\times X} k(x,y)\overline{f(x)}f(y)dxdy \geq 0 \tag{46.57}$$

for all $f \in L^2(X,\mathbb{C})$, the Hilbert-Schmidt operator

$$T_k : L^2(X,\mathbb{C}) \to L^2(X,\mathbb{C}) : f \mapsto \int_X k(\cdot,x)f(x)dx \tag{46.58}$$

admits a countable orthonormal basis $\{e_i\}_{i\in\mathbb{N}}$ with nonnegative eigenvalues $\{\lambda_i\}_{i\in\mathbb{N}}$ such that

$$k(x,y) = \sum_{i=1}^{\infty} \lambda_i e_i(x)\overline{e_i(y)}. \tag{46.59}$$

Theorem 46.7.10 (Bochner). *A continuous function satisfies the Mercer condition if and only if it is a kernel.*

Alternative Definition 46.7.11 (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

$$k(x,y) = \langle\phi(x)|\phi(y)\rangle_{\mathcal{H}}. \tag{46.60}$$

When using Mercer's theorem, the feature maps are given by

$$\phi_i : x \mapsto \sqrt{\lambda_i}e_i(x). \tag{46.61}$$

Remark 46.7.12. 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

$$\mathcal{H} = \left\{ f \in L^2(X,\mathbb{C}) \middle| \sum_{i=1}^{\infty} \frac{\langle f,e_i\rangle_{L^2}^2}{\lambda_i} < +\infty \right\}. \tag{46.62}$$

Remark 46.7.13 (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 $k : X\times X \to \mathbb{C}^{d\times d}$.

46.7.2 Decision boundaries

Consider a linear model for a classification problem $y = w^T x + b$. 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

$$\mathrm{sgn}(y) = \begin{cases} 1 & y > 0 \\ -1 & y < 0 \end{cases} \tag{46.63}$$

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 $d(x)$ from a data point x to the decision boundary by projecting onto this unit vector. The point $x - d(x)\frac{w}{\|w\|}$ is an element of the decision boundary and, hence, satisfies the hyperplane equation. Rewriting this gives an expression for the distance

$$d(x) = \frac{w^T x + b}{\|w\|}. \tag{46.64}$$

To account for the direction of the arrow, this number should be multiplied by the class $\mathrm{sgn}(y) = \pm 1$. This result is called the **geometric margin** $\gamma(x) := \mathrm{sgn}(y)d(x)$. 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

$$\max_w \frac{\gamma}{\|w\|} \qquad \text{such that} \qquad y_i(w^T x_i + b) \geq \gamma\|w\| \qquad \forall 1 \leq i \leq n, \tag{46.65}$$

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:

$$\min_w \|w\|^2 \qquad \text{such that} \qquad y_i(w^T x_i + b) \geq 1 \qquad \forall 1 \leq i \leq n. \tag{46.66}$$

The KKT conditions for this problem give the following results:

$$w = \sum_{i=1}^n \beta_i y_i x_i \tag{46.67}$$

and

$$\sum_{i=1}^n \beta_i y_i = 0, \tag{46.68}$$

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:

$$y \equiv w^T x + b = \sum_{i=1}^n \beta_i y_i \langle x_i | x\rangle + b. \tag{46.69}$$

Two observations can be made at this point. First of all, complementary slackness 46.3.7 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 solve 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:

$$\min_w \|w\|^2 + C\sum_{i=1}^n \xi \qquad \text{such that} \qquad \begin{cases} \xi_i \geq 0 & \forall 1 \leq i \leq n \\ y_i(w^Tx_i + b) \geq 1 - \xi_i & \forall 1 \leq i \leq n. \end{cases} \tag{46.70}$$

The resulting KKT conditions are as follows:

$$0 \leq \beta_i \leq C \tag{46.71}$$

and

$$\beta_i = 0 \implies y_i(w^Tx_i + b) \geq 1 \tag{46.72}$$

$$\beta_i = C \implies y_i(w^Tx_i + b) \leq 1 \tag{46.73}$$

$$\beta_i \in]0, C[\implies y_i(w^Tx_i + b) = 1. \tag{46.74}$$

?? COMPLETE (e.g. geometry)??

46.8 Vapnik-Chervonenkis theory

46.8.1 VC dimension

Definition 46.8.1 (Shatter coefficient). Let Ω denote the universe of discours and consider a set $C \subset P(\Omega)$. C **shatters** a set $A \subset \Omega$ if for every subset $a \subseteq A$ there exists a subset $c \in C$ such that

$$a = c \cap A. \tag{46.75}$$

The shatter(ing) coefficients of C are defined as follows:

$$S_n(C) := \max_{x_1,\ldots,x_n \in \Omega} \bigl|\bigl\{\{x_1, \ldots, x_n\} \cap c \bigm| c \in C\bigr\}\bigr|. \tag{46.76}$$

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 $\mathcal{F} \subseteq \{0,1\}^X$, the shatter(ing) coefficients (or **growth functions**) of $\mathcal{F}$ are given by:

$$S_n(\mathcal{F}) := \max_{x_1,\ldots,x_n \in X} \bigl|\bigl\{\{f(x_1), \ldots, f(x_n)\} \bigm| f \in \mathcal{F}\bigr\}\bigr|. \tag{46.77}$$

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 46.8.2 (Vapnik-Chervonenkis dimension). The Vapnik-Chervonenkis dimension of a collection of functions $\mathcal{F}$ is defined as follows:

$$\mathrm{VC}(\mathcal{F}) := \max\{k \in \mathbb{N} \mid S_k(\mathcal{F}) = 2^k\}. \tag{46.78}$$

Note that if a collection shatters n points, it also necessarily shatters a subset of these points, therefore, one also has

$$S_n(\mathcal{F}) = 2^n \tag{46.79}$$

for all $n < \mathrm{VC}(\mathcal{F})$.

A collection is called a **Vapnik-Chervonenkis class** if its VC dimension is finite.

Property 46.8.3 (Sauer's lemma[8]). The VC dimension bounds shatter coefficients in the following way:

$$S_n(\mathcal{F}) \leq \sum_{k=0}^{\mathrm{VC}(\mathcal{F})} \binom{n}{k} \tag{46.80}$$

for all $n \in \mathbb{N}$. For $n \leq \mathrm{VC}(\mathcal{F})$, 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:

$$\forall n \geq \mathrm{VC}(\mathcal{F}) : S_n(\mathcal{F}) \leq \left(\frac{e\,n}{\mathrm{VC}(\mathcal{F})}\right)^{\mathrm{VC}(\mathcal{F})}, \tag{46.81}$$

where e is the Euler number.

The generalization bounds of Section 46.4.2 on empirical risk minimization can be extended to uncountable hypothesis spaces as follows:

Property 46.8.4 (Generalization bound). The expected risk of the empirical risk minimizer $\hat{f}$ is bounded as follows:

$$R(\hat{f}) \leq \inf_{f\in\mathcal{H}} R(f) + 4\sqrt{2\frac{\mathrm{VC}(\mathcal{H})\ln(e\,n/\mathrm{VC}(\mathcal{H})) + \ln(2/\delta)}{n}} \tag{46.82}$$

with probability at least $1 - \delta$.

?? CHECK ALL THESE BOUNDS ??

46.8.2 Rademacher complexity

Remark 46.8.5 (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):

$$\mathrm{VC}(\mathcal{F}) := \mathrm{VC}\big(\{\theta(f(x) - \lambda) \,\big|\, f \in \mathcal{F}, \lambda \in \mathbb{R}\}\big). \tag{46.83}$$

This definition is equivalent to the following one based on subgraphs:

$$\mathrm{VC}(\mathcal{F}) = \mathrm{VC}\big(\{C_f := \{(x,\lambda) \in X \times \mathbb{R} \mid \lambda < f(x)\} \,\big|\, f \in \mathcal{F}\}\big). \tag{46.84}$$

[8]Sometimes called the **Sauer-Shelah lemma**.

Example 46.8.6 (Linear spaces). Every vector space V of real-valued functions has VC dimension at most $\dim(V) + 1$.

Example 46.8.7 (Translations). The set of translations 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 above make sense in this setting. To obtain useful bounds, it is important to introduce a new notion of effective size:

Definition 46.8.8 (Rademacher complexity). Consider a collection of functions $\mathcal{F} := \{f : X \to \mathbb{R}\}$. The Rademacher complexity is defined as follows:

$$\mathfrak{R}_n(\mathcal{F}) := \mathrm{E}\left[\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^n \sigma_i f(X_i)\right], \tag{46.85}$$

where the σ_i are Rademacher variables 43.3.2 and the expectation is taken over both the Rademacher variables and the sample $(X_1, \ldots, X_n)$.

Property 46.8.9 (Risk bound). Consider a collection of bounded functions $\mathcal{F} \subseteq [a, b]^{\mathcal{X}}$.

$$\mathrm{E}[R(f)] \leq R_{\text{emp}}(f) + 2\mathfrak{R}_n(\mathcal{F}) + (b-a)\sqrt{\frac{\log(1/\delta)}{2n}} \tag{46.86}$$

with probability at least $1 - \delta$.

Property 46.8.10 (VC dimension). The shatter coefficient bounds the Rademacher complexity in the following way:

$$\mathfrak{R}_n(\mathcal{F}) \leq \sqrt{\frac{2\ln(S_n(\mathcal{F}))}{n}}. \tag{46.87}$$

46.8.3 Relation to Glivenko-Cantelli classes

Property 46.8.11. Recall Property 45.2.15. 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 $\{(f(x_1), \ldots, f(x_n)) \mid f \in \mathcal{F}\}$. The latter is itself bounded above by the shatter coefficient $S_n(\mathcal{F})$. If $\mathcal{F}$ has finite VC dimension, Sauer's lemma 46.8.3 implies that this coefficient grows polynomial in n, so

$$\frac{1}{n}\ln N_C(\varepsilon, \mathcal{F}_M, \|\cdot\|_1) \xrightarrow{d} 0 \tag{46.88}$$

and, thus, $\mathcal{F}$ is Glivenko-Cantelli.

Theorem 46.8.12. *A class of sets is Vapnik-Chervonenkis if and only if it is Glivenko-Cantelli 45.2.12.*

46.9 Time series analysis

Definition 46.9.1 (Time series). A $\mathbb{N}$- or $\mathbb{Z}$-indexed stochastic process. Since $\mathbb{N}$ and $\mathbb{Z}$ are isomorphic in a very simple way, the two conventions for time series will be used interchangeably.

46.9.1 Stationarity

Definition 46.9.2 (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:

$$P(X_{t_1}, \dots, X_{t_r}) = P(X_{t_1+s}, \dots, X_{t_r+s}). \tag{46.89}$$

Definition 46.9.3 (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**: $\mathrm{E}[X_n] = \mathrm{E}[X_0]$ for all $n \in \mathbb{N}$.
2. **Finite covariance**: $\mathrm{cov}(X_i, X_j) < \infty$ for all $i, j \in \mathbb{N}$.
3. **Covariance-stationary**: $\mathrm{cov}(X_i, X_{i+j}) = \mathrm{cov}(X_0, X_j)$ for all $i, j \in \mathbb{N}$.

The following definition is a reformulation of Birkhoff ergodicity 16.2.26:

Definition 46.9.4 (Ergodicity). A time series $\{X_t\}_{t\in\mathbb{Z}}$ is ergodic if for every measurable function f the following equation holds for all $t \in \mathbb{Z}$:

$$\lim_{T\to\infty} \frac{1}{2T+1} \sum_{k=-T}^{T} f(X_k) = \mathrm{E}[f(X_t)]. \tag{46.90}$$

Intuitively this means that state space averages can be evaluated as time averages.

46.9.2 Correlation

Definition 46.9.5 (Autocorrelation function). Consider a time series $(X_n)_{n\in\mathbb{N}}$. The autocovariance (resp. autocorrelation) function of this time series is defined as the covariance (resp. autocorrelation) function of the random variables $(X_n)_{n\in\mathbb{N}}$.

Definition 46.9.6 (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:

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{i\omega k}, \tag{46.91}$$

where $\gamma(k)$ is the autocovariance function at lag k.

Under the assumption that the spectral density exists, the time series is said to have **short memory** if $f(0)$ is finite. Otherwise the series is said to have **long memory**.

Definition 46.9.7 (Lag operator[9]). The lag operator sends a variable in a time series to the preceding value:

$$BX_t = X_{t-1}. \tag{46.92}$$

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):

$$\theta(B) = 1 + \sum_{i=1}^{k} \theta_i B^i \tag{46.93}$$

$$\varphi(B) = 1 - \sum_{i=1}^{k} \varphi_i B^i. \tag{46.94}$$

[9] Also called the **backshift operator**.

Notation 46.9.8 (Difference operator). The difference operator Δ is defined as follows:

$$\Delta = 1 - B. \tag{46.95}$$

In a similar way one can define the **seasonal** difference operator:

$$\Delta_s = 1 - B^s. \tag{46.96}$$

Method 46.9.9 (Ljung-Box test). A test to see if a given set of autocorrelations of a time series is different from zero. Consider a time series of n elements and let $\{\rho_i\}_{1\leq i\leq k}$ be the first k lagged autocorrelation functions. The test statistic is defined as

$$Q = n(n+2)\sum_{i=1}^{k}\frac{\rho_k}{n-k}. \tag{46.97}$$

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 46.9.10 (Augmented Dickey-Fuller test). Consider a time series $(X_t)_{t\in T}$. The (augmented) Dickey-Fuller test checks if the time series is (trend) stationary. For this test one considers the following regression model (similar to the ARIMA-*models* discussed in the next section):

$$\Delta X_t = \alpha + \beta t + \gamma X_{t-1} + \sum_{i=1}^{p-1}\theta_i\Delta X_{t-i} + \varepsilon_t. \tag{46.98}$$

The test statistic is

$$\mathrm{DF} = \frac{\gamma}{\mathrm{SE}(\gamma)}, \tag{46.99}$$

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. 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).

46.9.3 Autoregressive models

Definition 46.9.11 (AR(p)-model). Consider a time series $(X_t)_{t\in T}$. The autogressive model of order p is defined as the multiple linear regression model of X_t with respect to the first p lagged values $X_{t-1},\ldots,X_{t-p}$ of the time series:

$$X_t = \beta_0 + \beta_1 X_{t-1} + \cdots + \beta_p X_{t-p} + \varepsilon_t. \tag{46.100}$$

Definition 46.9.12 (Partial autocorrelation function). The p^{th} autocorrelation function is defined as the p^{th} coefficient in the AR(p)-model.

Remark 46.9.13. The optimal order p of an autoregressive model is the one for which all higher partial autocorrelation functions (almost) vanish.

Definition 46.9.14 (MA(p)-model). Consider a time series $(X_t)_{t\in T}$ where every X_t contains a white noise contribution $\varepsilon_t \sim \mathcal{N}(0,\sigma^2)$. The moving average model of order p is defined as the multiple linear regression model of X_t with respect to the first p lagged values $\varepsilon_{t-1},\ldots,\varepsilon_{t-p}$ of the error term:

$$X_t = \beta_0 + \beta_1\varepsilon_{t-1} + \cdots + \beta_p\varepsilon_{t-p} + \varepsilon_t. \tag{46.101}$$

Since the error terms are assumed to have mean zero, one can see that the intercept term β_0 gives the mean of the time series.

Remark 46.9.15. The optimal order p of an autoregressive model is the one for which all higher autocorrelation functions (almost) vanish.

Definition 46.9.16 (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 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 46.9.17 (ARMA(p, q)-model).

$$X_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i} + \sum_{j=1}^{q} \beta_j \varepsilon_{t-j} + \varepsilon_t \tag{46.102}$$

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:

$$\varphi(B) X_t = \alpha_0 + \theta(B)\varepsilon_t. \tag{46.103}$$

By considering the special case where the polynomial $\mathcal{B}_\alpha^-$ has a unit root $1 - B$ with multiplicity d, one can obtain a generalization of the model:

$$\varphi(B)(1 - B)^d X_t = \alpha_0 + \theta(B)\varepsilon_t. \tag{46.104}$$

The interpretation of this additional factor $(1 - B)^d$ is related to the stationarity of the time series. The operator $1 - B$ is a finite difference operator:

$$\begin{aligned}(1 - B)X_t &= X_t - X_{t-1}\\ (1 - B)^2 X_t &= (X_t - X_{t-1}) - (X_{t-1} - X_{t-2})\\ &\cdots\end{aligned}$$

By successive applications, one can obtain a stationary time series from a nonstationary time series. This combination of differencing, autoregression and moving averages is called the **ARIMA**-model[11].

Remark 46.9.18. Including so-called *exogenous* variables, i.e. external predictors, leads to an **ARIMAX**-model.

Remark 46.9.19 (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 a MA-model, people have introduced multiple techniques (see for example [20]). One of the most famous ones is the method by *Durbin*:

---

[10] Sometimes this is used as a definition of invertibility.
[11] The 'I' stands for "integrated".

Method 46.9.20 (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 time series 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 is the period of the seasonal variation):

Definition 46.9.21 (ARIMA$(p,q,d)(P,Q,D)_s$-model).

$$\Phi(B^s)\varphi(B)\Delta_s^D\Delta^d X_t = \theta(B)\Theta(B^s)\varepsilon_t \tag{46.105}$$

46.9.4 Causality

Definition 46.9.22 (Granger causality). Consider two time series $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$. The time series 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:

$$P[Y_{t+k} \in A \mid \Omega(t)] \neq P[Y_{t+k} \in A \mid \Omega\backslash X(t)] \tag{46.106}$$

for some k, 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 46.9.23. 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 46.9.24. 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 time series).

46.10 Uncertainty modelling

46.10.1 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.

[12] In fact this was the original definition by *Granger*.

Definition 46.10.1 (Validity). Consider a measurable region predictor Γ and let P be the joint distribution on the instance space $Z \equiv X \times Y$. Γ is said to be valid (at **significance level** $\alpha \in [0,1]$ or **confidence level** $1-\alpha$) if it satisfies

$$P(y \in \Gamma^{\alpha}(x)) \geq 1 - \alpha. \tag{46.107}$$

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. Equation (46.107) characterizes pointwise validity in the sense that the probability of having an error is given by a Bernoulli process with parameter α. Asymptotic validity is a frequentist notion in the following sense:

$$\lim_{n\to\infty} \frac{\mathrm{Err}_n(\Gamma)}{n} \leq \alpha, \tag{46.108}$$

where $\mathrm{Err}_n(\Gamma)$ is the number of errors made by Γ after n trials. It should be clear that pointwise validity (both exact and conservative) implies asymptotic validity.

Remark 46.10.2 (Confidence regions). The definition of valid confidence predictors above is similar to the definition of confidence regions (Section 45.6). However, in contrast to confidence regions of population parameters, the size of confidence regions for predictive distributions does not go towards to 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]

46.10.2 Conformal prediction

A very general framework for the construction of valid prediction intervals in a model-independent manner is given by the conformal prediction framework by *Vovk et al.* The main ingredients for the construction are randomization and *conformity measures*.

The first step will be studying the behaviour under randomization of the existing data (be it measurements or past predictions). To ensure that the procedure satisfies the required confidence (or probability) bounds, one has to make some assumptions. One of the main benefits of this framework is that one can relax the condition of the data being i.i.d. to it being exchangeable:

Definition 46.10.3 (Exchangeability). Consider an ordered data sample $\{z_i\}_{1\leq i\leq N}$. The joint distribution $P(z_1,\ldots,z_N)$ is said to be exchangeable if it is invariant under any permutation of the data points. A distribution Q is said to be exchangeable if Q^n is exchangeable for all $n \in \mathbb{N}$.

This definition can be restated in a purely combinatorial way. First, define the notion of a **bag** obtained from a (possibly ordered) data sample $\{z_i\}_{1\leq i\leq N}$ as the (unordered) multiset $\mathcal{B}$ containing these elements. The joint distribution P is then said to be exchangeable if the probability of finding any sequence of data points is equal to the probability of drawing this same sequence from the bag of these elements. Since this probability is purely combinatorial and, hence, completely independent of the ordering, it should be clear that this coincides with the first definition. The set of bags in a space X is sometimes denoted by X^{∞}.

Definition 46.10.4 (Nonconformity measure). Consider a bag $\mathcal{B}$ together with a new element z^* in a probability space (Z,Σ,P). A noncomformity measure $f : Z^{\infty} \times Z \to \mathbb{R}$ is a measurable function that gives a number indicating how different z^* is from the content of $\mathcal{B}$.

[13]Note that when observations are not sampled according to a distribution, but are perfectly predictable, this is of course not true.

Remark. One could restate all statements in this section in terms of "conformity measures" and, hence, look at similarities instead of dissimilarities. It will become clear that the procedure is invariant under monotone transformations and, hence, everything can be multiplied by -1.

Example 46.10.5 (Point predictors). A general class of nonconformity measures is obtained from point predictors for a metric space (Y, d) . Given a point predictor $\rho : X \to Y$ trained on a bag $\mathcal{B}$, one can define a nonconformity measure as follows:

$$A_\rho(\mathcal{B}, (x, y)) := d(\rho(x), y). \tag{46.109}$$

Example 46.10.6 (Interval predictors). For every model $C \equiv (l, u) : X \to \mathbb{R}^2$ that predicts intervals, i.e. $\forall x \in X : l(x) \leq u(x)$, that is trained on a bag $\mathcal{B}$, one can define a nonconformity measure as follows:

$$A_C(\mathcal{B}, (x, y)) := \max\Big(l(x) - y, y - u(x)\Big). \tag{46.110}$$

It should be noted that although common, nonconformity measures and, by extension, conformal prediction are also applicable to nonmetric spaces:

Example 46.10.7 (Nested region predictors). Let T be a totally ordered set 2.6.5. For every model $(C_t)_{t\in T} : X \to P(Y)$ that predicts a sequence of nested regions, i.e.

$$s \leq t \implies C_s(x) \subseteq C_t(x), \tag{46.111}$$

trained on a bag $\mathcal{B}$, one can define a nonconformity measure as follows:

$$A_T(\mathcal{B}, (x, y)) := \inf\{t \in T \mid y \in C_t(x)\}. \tag{46.112}$$

Construction 46.10.8 (Conformal predictor). Consider a data sample given as a bag $\mathcal{B} \in Z^\infty$ together with a nonconformity measure A and let α denote the confidence level of the prediction region to be constructed. For any new element $z^* \in Z$ the algorithm proceeds as follows:

1. Denote the nonconformity score $A(\mathcal{B}, z^*)$ by μ_{z^*}.
2. For every element z in $\mathcal{B}$, define μ_z by replacing z by z^* in the bag and calculating the nonconformity score as in the previous step.
3. Calculate the conformal p-value as the fraction of elements $z \in \mathcal{B} \cup \{z^*\}$ for which $\mu_z \geq \mu_{z^*}$:
$$p^* := \frac{|\{z \in \mathcal{B} \cup \{z^*\} \mid \mu_z \geq \mu_{z^*}\}|}{|\mathcal{B}| + 1}. \tag{46.113}$$
4. Include an element z^* in the prediction region C^α if and only if $p^* > \alpha$.

It should be noted that, in general, the construction of these regions can be quite time-consuming because if the nonconformity measure A depends on model that has to be trained on $\mathcal{B}$, this training has to be reperformed for every element $z \in \mathcal{B}$ in step 2. For low-dimensional regions it can often be achieved by solving inequalities derived from the specific form of the nonconformity measure.

Property 46.10.9 (Optimality). A conformal predictor satisfies the following conditions:

- Regions are (conservatively) valid, i.e.
$$P(p^* \leq \alpha) \leq \alpha, \tag{46.114}$$
and thus also
$$P(y^* \in \Gamma^\alpha(x^*, \mathcal{B})) \geq 1 - \alpha, \tag{46.115}$$
where the probability is taken over both the bag $\mathcal{B}$ and the point (x^*, y^*).

- Regions are nested, i.e. $\alpha \leq \beta \implies \forall x \in X : \Gamma^\alpha(x) \subseteq \Gamma^\beta(x)$.

Given any region predictor satisfying these three properties below, there exists a conformal predictor that is more efficient, i.e. produces smaller prediction regions.

Property 46.10.10 (Smooth conformal predictors). One can modify the above construction in such a way that the resulting conformal predictors are not only conservatively valid but are also exactly valid. To this end one replaces the conformal p-value (46.113) by

$$p^*(\theta) := \frac{|\{z \in \mathcal{B} \cup \{z^*\} \mid \mu_z > \mu_{z^*}\}| + \theta\,|\{z \in \mathcal{B} \cup \{z^*\} \mid \mu_z = \mu_{z^*}\}|}{|\mathcal{B}| + 1}, \tag{46.116}$$

where θ is independently and uniformly sampled from the unit interval $[0, 1]$. Exact validity is then obtained by also marginalizing over the random variable θ.

Now, one could wonder if the assumption of exchangeability is a realistic assumption. Obviously if one applies the procedure to independent observations, everything is fine since i.i.d. sequences are clearly exchangeable. However, some important classes of data sequences are clearly not exchangeable, e.g. time series. This kind of data often contains intrinsic correlation and, hence, the exchangeability assumption is almost always violated. However, a solution exists. One can restate the construction above using an explicit randomization as is done in [21]. There, one replaces the nonconformity measure by a function that acts on ordinary sequences instead of unordered bags. The fraction p^* can then be expressed as follows:

$$p^* = \frac{1}{|S_{N+1}|} \sum_{\sigma \in S_{N+1}} \mathbb{1}(A(\sigma \cdot \vec{z}) \geq A(\vec{z})), \tag{46.117}$$

where $\vec{z} \equiv (z_1, \ldots, z_N, z^*)$. Using these explicit permutations, one can generalize the construction of conformal predictors to arbitrary randomization schemes, i.e. to subgroups of S_{N+1}. However, in general this will ruin the validity of the procedure.

?? FINISH (is this even relevant for this compendium) ??

At last a computationally efficient modification of the original CP algorithm is introduced. For most applications, especially those in machine learning and big data, the computational inefficiency of conformal predictors would make them hard to use. To overcome this issue *Papadopoulos et al.* introduced the following modification:

Construction 46.10.11 (Inductive CP). Consider a data set $\mathcal{D} \subset Z$ and a nonconformity measure A based on an underlying predictor. First, split $\mathcal{D}$ into a training set $\mathcal{T}$ and a calibration set $\mathcal{C}$. Using $\mathcal{T}$, train the underlying predictor of A. Then, for every point $z \equiv (x, y) \in \mathcal{C}$, construct the nonconformity score $\mu_z := A(z)$. As before, for every new element $z^* \in Z$ the conformal p-value is defined as the fraction of elements in $\mathcal{C}$ for which the nonconformity measure is larger than the one for z^*:

$$p^* := \frac{|\{z \in \mathcal{C} \cap \{z^*\} \mid \mu_z \geq \mu_{z^*}\}|}{|\mathcal{C}| + 1}. \tag{46.118}$$

As in the original CP algorithm, a new observation z^* is included in the prediction region if and only if $p^* > \alpha$:

$$\Gamma^\alpha(x) := \{y \in Y \mid p(x, y) > \alpha\}. \tag{46.119}$$

It should be clear that the underlying predictor only needs to be trained once using this scheme.

Remark 46.10.12 (Terminology). The name "inductive CP" stems from the fact that the general behaviour is deduced from a small subset of all observations. For this reason one sometimes calls the original algorithm a "transductive" method.

Property 46.10.13 (Validity). Although the above ICP algorithm is already computationally much more efficient than its transductive counterpart, one can go even further. However, in this case one needs to pay attention in order not to ruin the validity. To use Equation (46.118) one does not have to retrain the model every time, but one still needs to reevaluate the nonconformity score for possible $y^* \in Y$. It would be even better if one could extract the boundaries of the prediction region straight from the calibration data.

If the data is exchangeable (and the resulting nonconformitys score are too) it is not hard to see that the rank of any new nonconformity score among the calibration scores $A(\mathcal{C}) := \{\mu_z \in \mathbb{R} \mid z \in \mathcal{C}\}$ is uniformly sampled from $\{1, \dots, |\mathcal{C}| + 1\}$. So, given a quantile level β, the probability of finding a new nonconformity score smaller than or equal to the β-quantile of $A(\mathcal{C})$ is

$$P\Big(\mu_{z^*} \leq q_\beta(A(\mathcal{C}))\Big) = \frac{\lceil \beta|\mathcal{C}|\rceil}{|\mathcal{C}| + 1}. \tag{46.120}$$

If one constructs a prediction region by including all points $(x^*, y^*) \in Z$ such that the associated nonconformity score is smaller than the $(1 - \alpha)$-quantile, one immediately obtains

$$P(y^* \in \Gamma^\alpha(x^*, \mathcal{C})) = \frac{\lceil (1-\alpha)|\mathcal{C}|\rceil}{|\mathcal{C}| + 1}. \tag{46.121}$$

However, this is where the TCP and ICP algorithms differ. When using Equation (46.113) for the transductive algorithm, the above probabilities would contain a factor $|\mathcal{C}| + 1$ in both the numerator and the denominator, so the coverage condition is satisfied. However, for this quantile-based reformulation of the inductive algorithm, a minimal modification also leads to validity:

$$\Gamma^\alpha(x) \longrightarrow \Gamma^\alpha(x) := \big\{y \in Y \mid A(x, y) \leq q_{(1+1/|\mathcal{C}|)(1-\alpha)}(A(\mathcal{C}))\big\}, \tag{46.122}$$

i.e. the empirical quantiles are replaced by ‘‘inflated’’ quantiles. When using Equations (46.118) and (46.119) to determine the ICP region, one is essentially including all points (x^*, y^*) such that the nonconformity score is smaller than or equal to the $(1 - \alpha)$-quantile of the enhanced calibration curve $A(\mathcal{C}) \cup \{A(x^*, y^*)\}$. It is not hard to show that this quantile is equivalent to the inflated quantile of the ordinary calibration curve.

The smoothing of Property 46.10.10 can also be implied in this case. If one does not use a smoothened conformal predictor but assumes that all calibration scores are distinct, the following property is obtained:

$$P(y^* \in \Gamma^\alpha(x, \mathcal{C})) \geq 1 - \alpha + \frac{1}{|\mathcal{C}|}. \tag{46.123}$$

So, in the limit of large calibration sets, the exact validity is recovered.

Remark 46.10.14. One does have to pay attention when interpreting the above statement. The validity property holds in probability with respect to both the new instance (x^*, y^*) and the calibration set $\mathcal{C}$. However, this does not not mean that for a fixed calibration set $\mathcal{C}$ the error fraction

$$\frac{|\{1 \leq i \leq k : y_i \in \Gamma^\alpha(x_i, \mathcal{C})\}|}{k} \tag{46.124}$$

is bounded by α for $k \longrightarrow \infty$. Because the events are not independent, the error can be much larger than α.

The above offline ICP algorithm can be generalized to an online algorithm:

Construction 46.10.15 (Online ICP). Consider an increasing sequence of positive integers $(m_n)_{n\in\mathbb{N}_0}$ of "update thresholds". The prediction region $\Gamma^\alpha(\mathcal{S})$ for the data sample $\mathcal{S} := \{(x_1, y_1), \ldots, (x_n, y_n)\}$ is defined as follows:

- If $n \leq m_1$, use a fixed conformal predictor to construct $\Gamma^\alpha(\mathcal{S})$.
- If $m_k < n \leq m_{k+1}$, construct $\Gamma^\alpha(\mathcal{S})$ as follows:

$$\Gamma^\alpha(\mathcal{S}) := \left\{ y \in Y \,\middle|\, \frac{|\{m_k < j \leq n \mid \mu_{z_j} \geq \mu_{z_n}\}|}{n - m_k} > \varepsilon \right\}, \tag{46.125}$$

  where

$$\begin{aligned} \mu_{z_j} &:= A(\mathcal{B}(z_1, \ldots, z_{m_k}), (x_j, y_j)) \\ \mu_{z_n} &:= A(\mathcal{B}(z_1, \ldots, z_{m_k}), (x_n, y)). \end{aligned}$$

It is clear that for $k \ll |\mathcal{C}|$ the offline ICP algorithm approximates the online version. The major benefit of the online algorithm is that one does not need to use the inflated quantiles to obtain valid prediction regions, because the calibration set grows every time new data is observed and, hence, the finite-size fluctuations get supressed.

?? COMPLETE ??

46.10.3 Classifier calibration

A specific instance of regression problems are classification tasks, where the one tries to model a function $f : X \to Y$ with Y finite (or discrete). For the case of probabilistic classifiers, the notion of validity 46.10.1 admits the following formulation:

Definition 46.10.16 (Calibration). A probabilistic (multiclass) classifier $\hat{P}(\cdot \mid \cdot) : Y \times X \to [0, 1]$ is said to be (well-)calibrated if

$$\Pr\left(y \,\middle|\, \hat{P}(y \mid x) = p\right) = p \tag{46.126}$$

for all $p \in [0, 1]$. Here, the confidence level $1 - \alpha$ is the output of the classifier $\hat{P}$, 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 $\mathcal{P} := \{0 = p_1, p_2, \ldots, p_n = 1\}$ of the interval $[0, 1]$ the i^{th} 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: $X \to \{0, 1\}$. Here one can easily apply the conformal prediction framework from the previous section. To this end, define the following nonconformity measure:

$$A_{\text{binary}}(\mathcal{B}, (x, y)) := 1 - \hat{P}(y \mid x; \mathcal{B}), \tag{46.127}$$

where the classifier is possibly estimated on $\mathcal{B}$. 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.

In fact, for classification tasks, where the codomain is finite (or at least discrete), it makes sense to slightly change the approach. Instead of fixing the significance level α, one could (or should) consider some specific values:

- **Confidence**: $\sup\{1-\alpha \mid |\Gamma^\alpha(x)| \leq 1\}$
- **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 othe 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 $\hat{P}$ and fits a sigmoidal layer of the form

$$\hat{P}_{\text{Platt}}(y \mid x; A, B) = \frac{1}{1+\exp(-A\hat{P}(y \mid x) - B)}. \tag{46.128}$$

For the latter one modifies the existing logits as follows for some parameter $T > 0$:

$$\hat{P}_{\text{temp}}(y \mid x; T) = \frac{1}{1+\exp(-\hat{z}(y \mid x)/T)}, \tag{46.129}$$

where $\hat{z}$ represents the logits of the estimator $\hat{P}$.

Remark 46.10.17 (Accuracy). Because temperature scaling does not change the maximum of the softmax function, the eventual predictions do not change.

46.10.4 Normalizing flows

One of the main problems for obtaining valid uncertainty estimates is that the underlying distribution of a given data set is often 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 way around this problem is by first transforming the data such that it is sampled according to a well-known distribution.

46.10.5 Conditionality

The above sections considered confidence predictors that were valid in a global sense, i.e. averaged over the whole instance space $X \times Y$. However, as in ordinary probability theory, in many settings it makes more sense to consider conditional statements:

$$P(y \in \Gamma^\alpha(x) \mid \kappa(x, y)) \geq 1 - \alpha, \tag{46.130}$$

where the function $\kappa : X \times Y \to K$ 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 $X \times Y$. 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:

$$P(y \in \Gamma^\alpha(x) \mid x) \geq 1 - \alpha. \tag{46.131}$$

However, in general, one can show that this cannot be attained:

Property 46.10.18 (No-go theorem). Let 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 $X \times Y$ and P_X-almost all non-atoms $x \in X$ one has

$$\Pr(\lambda(\Gamma^\alpha(x)) = +\infty) \geq 1 - \alpha \tag{46.132}$$

and

$$\Pr(\mathrm{co}(\Gamma^\alpha(x)) = \mathbb{R}) \geq 1 - 2\alpha, \tag{46.133}$$

where λ and co denote the Lebesgue measure and convex hull, respectively.

?? FINISH ??

46.10.6 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 43.6.6 from the p-values produced by a conformal predictor. For every (reasonable) martingale $(X_n)_{n\in\mathbb{N}}$ the Doob-Ville inequality 43.6.8 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 [17].

The general expression of the martingale will be of the form

$$X_i \equiv \prod_{k=0}^{i} f_k(p_k), \tag{46.134}$$

where p_k is the p-value of the k^{th} 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

$$\int_0^1 f_k(p)dp = 1. \tag{46.135}$$

Method 46.10.19 (Power martingale). For every constant $\varepsilon \in [0,1]$ one defines the power martingale as

$$X_i^\varepsilon := \prod_{k=0}^{i} \varepsilon p_k^{\varepsilon-1}. \tag{46.136}$$

One can also construct a **simple mixture** martingale by integrating over ε:

$$X_i := \int_0^1 X_i^\varepsilon d\varepsilon. \tag{46.137}$$

Because $\varepsilon - 1 \leq 0$, 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 46.10.20 (Plug-in martingale). Let $\hat{\rho}_i(p)$ denote an estimate of the probability density constructed using the p-values $\{p_1, \ldots, p_i\}$. The plug-in martingale is defined by the betting functions

$$f_i := \hat{\rho}_{i-1}. \tag{46.138}$$

Now, if the empirical distribution functions of the p-values converge weakly 16.1.51 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 46.10.21. 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

$$X_i := \prod_{k=0}^{i} \frac{g(x_i)}{f(x_i)} \tag{46.139}$$

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 46.10.22 (Likelihood nonconformity). For every two probability densities f, g one can define a nonconformity measure as follows:

$$A_{\mathrm{NP}}(\mathcal{B}, z) := \log f(z) - \log g(z), \tag{46.140}$$

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. If the initial distribution is not known, it can be estimated based on the bag $\mathcal{B}$.

[14]The subscript refers to the Neyman-Pearson lemma 45.7.12 for which this function gives the (logarithm) of the test statistic.

Chapter 47

Fuzzy Set Theory & Imprecise Probabilities ♣

The main reference for the basics on fuzzy sets is the original paper [22]. For the basics of (ordered) sets, see Section 2.6 at the beginning of this compendium.

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 here is the characterization of uncertainty in statistics and machine learning. For that reason it was added here.

47.1 Fuzzy sets

Definition 47.1.1 (Fuzzy set). Consider a set X (this set corresponds to the universe of discours 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 grade 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 obtained.

A fuzzy set is said to be **empty** if its defining function is identically zero.

Remark 47.1.2. One can generalize this definition by replacing $[0,1]$ by a more general poset (with the necessary properties).

Definition 47.1.3 (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:

$$f^*A(x) := A(f(x)). \tag{47.1}$$

The following definition is an immediate generalization of Definition 2.2.4:

Definition 47.1.4 (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 2.2.6 can be extended through the following formula:

$$S \circ R(x,z) := \sup_{y \in X} \min\big(R(x,y), S(y,z)\big). \tag{47.2}$$

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 47.1.5 (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 14.8.1:

$$C(x) := \Lambda(x)A(x) + (1 - \Lambda(x))B(x). \tag{47.3}$$

47.2 Fuzzy measure theory

In this section some of the content of Chapter 16 is generalized to fuzzy set theory. Unless explicitly stated, all concepts will be defined over a general measurable space (X, Σ).

Definition 47.2.1 (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 $A, B \in \mathcal{C}$.

A capacity is said to be **normalized** (or **regular**) if $\mu(X) = 1$. If one drops the monotonicity condition, the notion of a **game** is obtained.

Definition 47.2.2 (Alternating capacity). A k-alternating capacity μ satisfies

$$\mu\left(\bigcap_{i=1}^{k} A_i\right) \leq \sum_{I \subset \{1,\ldots,k\}} (-1)^{|I|+1} \mu\left(\bigcup_{i \in I} A_i\right) \tag{47.4}$$

for all measurable sets $A_1, \ldots, A_k$. 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 47.2.3. A capacity $\mu : \Sigma \to \mathbb{R}$ is k-monotone (resp. k-alternating) if and only if its **dual capacity**

$$\overline{\mu}(A) := \mu(X) - \mu(A^c) \tag{47.5}$$

is k-alternating (resp. k-monotone).

Definition 47.2.4 (Choquet integral). Consider a capacity μ on (X, Σ) and a measurable function $f : \mathcal{U} \to \mathbb{R}$, 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:

$$\int_X f \, d\mu := \int_{-\infty}^{0} \big(\mu(\{y \mid f(y) \geq x\}) - \mu(X)\big) \, dx + \int_0^{+\infty} \mu(\{y \mid f(y) \geq x\}) \, dx. \tag{47.6}$$

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:

$$\int_X f^+ \, d\mu := \int_0^{+\infty} \mu(\{y \mid f^+(y) \geq x\}) \, dx. \tag{47.7}$$

The general Choquet integral then takes the form

$$\int_X f \, d\mu := \int_X f^+ \, d\mu - \int_X f^- \, d\overline{\mu}. \tag{47.8}$$

[1] Also called a **fuzzy measure**.

Property 47.2.5. A capacity is 2-alternating if and only if the associated Choquet integral is subadditive.

Example 47.2.6 (Possibility and necessity). A normalized capacity μ satisfying

$$\mu(A, B) = \max\big(\mu(A), \mu(B)\big) \tag{47.9}$$

for all $A, B \in \Sigma$. If one replaces the maximum by a minimum, the definition of a **necessity measure** is obtained.

Example 47.2.7 (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 47.2.8 (Consonant capacity). A capacity $\mu : \Sigma \to \mathbb{R}$ such that the **focal sets**, the sets $A \in \Sigma$ such that $\mu(A) > 0$, admit a total order (i.e. they are nested).

Remark 47.2.9. Consonant belief and plausibility measures are called **necessity** and **possibility measures**, respectively.

47.3 Imprecise probabilities

Definition 47.3.1 (Gamble). Consider a set X. The set of gambles over X is the (Banach) space of bounded real-valued functions $\mathcal{B}(X) := \{f : X \to \mathbb{R} \mid \sup_{x\in X} f(x) < +\infty\}$. 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 $\mathcal{B}^+(X) := \{f \in \mathcal{B}(X) \mid f > 0\}$, and
4. **Avoiding partial losses:** $\mathcal{B}^-(X) \cap \mathcal{D} = \emptyset$, $\mathcal{B}^-(X) := \{f \in \mathcal{B}(X) \mid f < 0\}$.

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 47.3.2 (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 $\mathcal{D}'$.

Definition 47.3.3 (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 $P_1, P_2 \in K \subseteq \mathbb{P}(X)$, one is also uncertain about the mixtures $\lambda P_1 + (1-\lambda)P_2, \lambda \in [0, 1]$.

Part VII

Classical Physics

Chapter 48

Classical Mechanics

The section about the geometric framework is mainly based on [23, 24]. For an introduction to differential geometry, see Chapter 29 and onwards.

48.1 Newtonian mechanics

48.1.1 Linear motion

Axiom 48.1 (Newton's second law). The force acting on a system can be related to the change of momentum in the following way:

$$\vec{\boldsymbol{F}} = \frac{d\vec{\boldsymbol{p}}}{dt}. \tag{48.1}$$

Definition 48.1.1 (Work).

$$W := \int \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{l}} \tag{48.2}$$

Definition 48.1.2 (Conservative force). If the work done by a force is independent of the path taken, the force is said to be **conservative**:

$$\oint_C \vec{\boldsymbol{F}} \cdot d\vec{\boldsymbol{l}} = 0. \tag{48.3}$$

The Kelvin-Stokes theorem 21.2.5 together with relation (21.13) allows to rewrite the conservative force as the gradient of a scalar field:

$$\vec{\boldsymbol{F}} = -\nabla V. \tag{48.4}$$

Definition 48.1.3 (Central force). A force that only depends on the relative position of two objects:

$$\vec{\boldsymbol{F}}_c \equiv F\big(\|\vec{\boldsymbol{r}}_2 - \vec{\boldsymbol{r}}_1\|\big)\hat{\boldsymbol{e}}_{\boldsymbol{r}}. \tag{48.5}$$

Formula 48.1.4 (Momentum of point mass). Consider a mass m with velocity $\vec{\boldsymbol{v}}$. Its momentum is given by

$$\vec{\boldsymbol{p}} = m\vec{\boldsymbol{v}}. \tag{48.6}$$

If the mass is constant along its trajectory, Newton's second law 48.1 can be rewritten as follows:

$$\vec{\boldsymbol{F}} = m\frac{d\vec{\boldsymbol{v}}}{dt} = m\vec{\boldsymbol{a}}. \tag{48.7}$$

Formula 48.1.5 (Kinetic energy). For a free particle with total momentum p, the kinetic energy is given by the following formula:

$$E_{\text{kin}} := \frac{p^2}{2m}. \tag{48.8}$$

48.1.2 Rotational motion

In this section r always denotes the distance from the object's center of mass to the axis around which the object rotates.

Definition 48.1.6 (Angular velocity).

$$\vec{\boldsymbol{\omega}} := \frac{\vec{\boldsymbol{r}} \times \vec{\boldsymbol{v}}}{r^2} \tag{48.9}$$

Definition 48.1.7 (Angular frequency).

$$\nu := \frac{\|\vec{\boldsymbol{\omega}}\|}{2\pi} \tag{48.10}$$

Definition 48.1.8 (Moment of inertia). For a (spherically) symmetric object the moment of inertia is given by

$$I := \int_V r^2 \rho(r) dV, \tag{48.11}$$

where ρ denotes the mass density function. For a general body the moment of inertia tensor is given by

$$\mathcal{I} := \int_V \rho(\vec{\boldsymbol{r}})\big(r^2 \mathbb{1} - \vec{\boldsymbol{r}} \otimes \vec{\boldsymbol{r}}\big) dV. \tag{48.12}$$

Example 48.1.9 (Objects with azimuthal symmetry). Let m, r denote the mass and the radius of the object, respectively.

- Solid disk: $I = \frac{1}{2}mr^2$
- Cylindrical shell: $I = mr^2$
- Hollow sphere: $I = \frac{2}{3}mr^2$
- Solid sphere: $I = \frac{2}{5}mr^2$

Proof for the solid disk and sphere. The volume of a (solid) disk is given by

$$V_{\text{disk}} = \pi R^2 d, \tag{48.13}$$

where R denotes the radius and d denotes the thickness. The mass density is then given by

$$\rho = \frac{M}{\pi R^2 d}. \tag{48.14}$$

Using cylindrical coordinates the moment of inertia becomes

$$\begin{aligned} I &= \frac{M}{\pi R^2 d} \int_0^{2\pi} d\varphi \int_0^d dz \int_0^R r^3 dr \\ &= \frac{M}{\pi R^2 d} 2\pi d \frac{R^4}{4} \\ &= \frac{1}{2} M R^2. \end{aligned} \tag{48.15}$$

The volume of a solid sphere is given by

$$V_{\text{sphere}} = \frac{4}{3}\pi R^3, \tag{48.16}$$

where R denotes the radius. The mass density is then given by

$$\rho = \frac{M}{\frac{4}{3}\pi R^3}. \tag{48.17}$$

Spherical coordinates will be used to derive the moment of inertia, but one has to be careful. The r in Formula 48.1.8 is the distance between a point in the body and the axis of rotation. So it is not the same as the 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:

$$r = r' \sin\theta, \tag{48.18}$$

where r' is the spherical coordinate. Now, one can calculate the moment of inertia as follows:

$$\begin{aligned} I &= \frac{M}{\frac{4}{3}\pi R^3}\int_0^{2\pi} d\varphi \int_0^R r'^4 dr' \int_0^\pi \sin^3\theta d\theta \\ &= \frac{M}{\frac{4}{3}\pi R^3} 2\pi \frac{R^5}{5}\frac{4}{3} \\ &= \frac{2}{5}MR^2. \end{aligned} \tag{48.19}$$

□

Definition 48.1.10 (Principal axes of inertia). Let I be the matrix of inertia, i.e. the matrix associated with the inertia tensor (48.12). This is a real symmetric matrix and, by Property 20.5.16, admits an eigendecomposition of the form

$$I = Q\Lambda Q^T. \tag{48.20}$$

The columns of Q determine the principal axes of inertia. The eigenvalues are called the **principal moments of inertia**.

Theorem 48.1.11 (Parallel axis theorem[1]). *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:*

$$I_A = I_{\text{CM}} + m\|\vec{\boldsymbol{r}}_A - \vec{\boldsymbol{r}}_{\text{CM}}\|^2, \tag{48.21}$$

where m is the mass of the rotating body.

Definition 48.1.12 (Angular momentum).

$$\vec{\boldsymbol{L}} := \vec{\boldsymbol{r}} \times \vec{\boldsymbol{p}} \tag{48.22}$$

Given the angular velocity vector $\vec{\boldsymbol{\omega}}$ one can compute the angular momentum as follows:

$$\vec{\boldsymbol{L}} = \mathcal{I}(\vec{\boldsymbol{\omega}}), \tag{48.23}$$

[1] Also called **Steiner's theorem**.

where $\mathcal{I}$ is the inertia tensor. If $\vec{\boldsymbol{\omega}}$ is parallel to a principal axis, the formula reduces to

$$\vec{\boldsymbol{L}} = I\vec{\boldsymbol{\omega}}, \tag{48.24}$$

where I is the corresponding principal moment of inertia.

Formula 48.1.13 (Torque). For angular momenta there exists a formula analogous to Newton's second law:

$$\vec{\boldsymbol{\tau}} := \frac{d\vec{\boldsymbol{L}}}{dt}. \tag{48.25}$$

For a constant mass this formula can be rewritten as follows:

$$\vec{\boldsymbol{\tau}} = I\vec{\boldsymbol{\alpha}} = \vec{\boldsymbol{r}} \times \vec{\boldsymbol{F}}. \tag{48.26}$$

Remark 48.1.14. 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 $\det = -1$.

Formula 48.1.15 (Rotational energy).

$$E_{\mathrm{rot}} := \frac{1}{2}\mathcal{I}(\vec{\boldsymbol{\omega}}) \cdot \vec{\boldsymbol{\omega}} \tag{48.27}$$

48.2 Lagrangian mechanics

48.2.1 Action

Definition 48.2.1 (Generalized coordinates). The generalized coordinates q_k 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 49 it is explained how the relevant degrees of freedom can be extracted.)

Definition 48.2.2 (Generalized velocities). The generalized velocities $\dot{q}_k$ are the derivatives of the generalized coordinates with respect to time.

Definition 48.2.3 (Conjugate momentum).

$$p_k := \frac{\partial L}{\partial \dot{q}^k} \tag{48.28}$$

Notation 48.2.4. Given a Lagrangian function, depending on n generalized coordinates and their associated velocities, the following shorthand notation is often used:

$$L(q, \dot{q}, t) \equiv L(q_1(t), \ldots, q_n(t), \dot{q}_1(t), \ldots, \dot{q}_n(t), t) \tag{48.29}$$

Definition 48.2.5 (Action). Given a Lagrangian function L, the action is a functional on the space of paths in configuration space defined by integrating L:

$$S[q] := \int_{t_1}^{t_2} L(q, \dot{q}, t)dt. \tag{48.30}$$

48.2.2 Euler-Lagrange equations

Axiom 48.2 (d'Alembert's principle).

$$\sum_i (\vec{\boldsymbol{F}}_i - \dot{\vec{\boldsymbol{p}}}_i) \cdot \delta\vec{\boldsymbol{q}}_i = 0, \tag{48.31}$$

where the δq^i denote the **virtual displacement** vectors, i.e. the infinitesimal variations consistent with the contraints. In the spirit of calculus of variations (Section 38.1) these are defined as follows. Consider a trajectory $\vec{\boldsymbol{q}} : [0,1] \to \mathbb{R}^n$ and a variation $\vec{\boldsymbol{\gamma}}$ of $\vec{\boldsymbol{q}}$, i.e. a smooth function $\vec{\boldsymbol{\gamma}} : [0,1] \times [-\varepsilon, \varepsilon] \to \mathbb{R}^n$ such that $\vec{\boldsymbol{\gamma}}(t,0) = \vec{\boldsymbol{q}}(t)$. This function encodes how $\vec{\boldsymbol{q}}$ can vary given the constraints of the system. The virtual displacement vector is then defined as the tangent vector

$$\delta\vec{\boldsymbol{q}} := \left.\frac{d\vec{\boldsymbol{\gamma}}}{d\varepsilon}\right|_{\varepsilon=0}. \tag{48.32}$$

Formula 48.2.6 (Euler-Lagrange equation of the first kind).

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}^k}\right) - \frac{\partial T}{\partial q^k} = Q_k, \tag{48.33}$$

where T is the total kinetic energy and Q_k are the **generalized forces**:

$$Q_k := \sum_i \vec{\boldsymbol{F}}_i \cdot \frac{\partial \vec{\boldsymbol{r}}_i}{\partial q^k}. \tag{48.34}$$

Note that the constraint forces do not contribute to this quantity because they are always perpendicular to the motion.

Formula 48.2.7 (Euler-Lagrange equation of the second kind).

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^k}\right) - \frac{\partial L}{\partial q^k} = 0 \tag{48.35}$$

Proof based on d'Alembert's principle. In the following derivation the mass is assumed to be constant.

$$\begin{aligned}
&\sum_k \left(\vec{\boldsymbol{F}}_k - \dot{\vec{\boldsymbol{p}}}_k\right) \cdot \dot{\vec{\boldsymbol{r}}}_k = 0 \\
\iff &\sum_k \left(\vec{\boldsymbol{F}}_k - \dot{\vec{\boldsymbol{p}}}_k\right) \cdot \left(\sum_l \frac{\partial \vec{\boldsymbol{r}}_k}{\partial q_l}\dot{q}_l\right) = 0 \\
\iff &\sum_l \left(\sum_k \vec{\boldsymbol{F}}_k \cdot \frac{\partial \vec{\boldsymbol{r}}_k}{\partial q_l} - \sum_k m\ddot{\vec{\boldsymbol{r}}}_k \cdot \frac{\partial \vec{\boldsymbol{r}}_k}{\partial q_l}\right)\dot{q}_l = 0 \\
\iff &\sum_l \left(Q_l - \sum_k m\ddot{\vec{\boldsymbol{r}}}_k \cdot \frac{\partial \vec{\boldsymbol{r}}_k}{\partial q_l}\right)\dot{q}_l = 0.
\end{aligned} \tag{48.36}$$

Now, consider at the following derivative:

$$
\begin{aligned}
\frac{d}{dt}\left(\dot{\vec{r}}\cdot\frac{\partial\vec{r}}{\partial q_l}\right) &= \ddot{\vec{r}}\cdot\frac{\partial\vec{r}}{\partial q_l} + \dot{\vec{r}}\cdot\frac{d}{dt}\left(\frac{\partial\vec{r}}{\partial q_l}\right) \\
\iff \ddot{\vec{r}}\cdot\frac{\partial\vec{r}}{\partial q_l} &= \frac{d}{dt}\left(\dot{\vec{r}}\cdot\frac{\partial\vec{r}}{\partial q_l}\right) - \dot{\vec{r}}\cdot\frac{d}{dt}\left(\frac{\partial\vec{r}}{\partial q_l}\right) \\
\iff \ddot{\vec{r}}\cdot\frac{\partial\vec{r}}{\partial q_l} &= \frac{d}{dt}\left(\dot{\vec{r}}\cdot\underbrace{\frac{\partial\vec{r}}{\partial q_l}}_{A}\right) - \dot{\vec{r}}\cdot\left(\frac{\partial\dot{\vec{r}}}{\partial q_l}\right).
\end{aligned}
\tag{48.37}
$$

To evaluate the factor indicated by A, one can consider another derivative:

$$
\begin{aligned}
\frac{\partial\dot{\vec{r}}}{\partial\dot{q}_l} &= \frac{\partial}{\partial\dot{q}_l}\left(\sum_k\frac{\partial r}{\partial q_k}\dot{q}_k\right) \\
&= \sum_k\frac{\partial r}{\partial q_k}\delta_{kl} \\
&= \frac{\partial\vec{r}}{\partial q_l} \\
&= A.
\end{aligned}
$$

Substituting this in Equation (48.37) gives

$$
\begin{aligned}
\ddot{\vec{r}}\cdot\frac{\partial\vec{r}}{\partial q_l} &= \frac{d}{dt}\left(\dot{\vec{r}}\cdot\frac{\partial\dot{\vec{r}}}{\partial\dot{q}_l}\right) - \dot{\vec{r}}\cdot\left(\frac{\partial\dot{\vec{r}}}{\partial q_l}\right) \\
&= \frac{d}{dt}\left(\frac{1}{2}\frac{\partial\dot{\vec{r}}^2}{\partial\dot{q}_l}\right) - \frac{1}{2}\frac{\partial\dot{\vec{r}}^2}{\partial q_l}.
\end{aligned}
\tag{48.38}
$$

If one multiplies this by the mass m and sums over all masses, the following expression is obtained:

$$
\begin{aligned}
\sum_k m_k\ddot{\vec{r}}_k\cdot\frac{\partial\vec{r}_k}{\partial q_l} &= \frac{d}{dt}\frac{\partial}{\partial\dot{q}_l}\left(\sum_k\frac{1}{2}m\dot{\vec{r}}_k^2\right) - \frac{\partial}{\partial q_l}\left(\sum_k\frac{1}{2}m\dot{\vec{r}}_k^2\right) \\
&= \frac{d}{dt}\frac{\partial T}{\partial\dot{q}_l} - \frac{\partial T}{\partial q_l},
\end{aligned}
\tag{48.39}
$$

where the total kinetic energy is denoted by T in the last line. Plugging this result into Equation (48.36) gives

$$
\sum_l\left(Q_l - \frac{d}{dt}\frac{\partial T}{\partial\dot{q}_l} - \frac{\partial T}{\partial q_l}\right)\dot{q}_l = 0.
\tag{48.40}
$$

Because all the coordinates q_l are independent, the following relation should hold for all l:

$$
\begin{aligned}
Q_l - \frac{d}{dt}\left(\frac{\partial T}{\partial\dot{q}_l}\right) - \frac{\partial T}{\partial q_l} &= 0 \\
\iff \frac{d}{dt}\left(\frac{\partial T}{\partial\dot{q}_l}\right) - \frac{\partial T}{\partial q_l} &= Q_l.
\end{aligned}
\tag{48.41}
$$

This last equation is known as a **Lagrange equation of the first kind**.

If the system only contains conservative forces, the force on the i^{th} mass can be written as

$$F_i = -\nabla_i V. \tag{48.42}$$

With this in mind, one can relate the partial derivatives of the potential to the generalized forces:

$$\begin{aligned}\frac{\partial V}{\partial q_l} &= \sum_i (\nabla_i V) \cdot \frac{\partial \vec{\boldsymbol{r}}_i}{\partial q_l} \\ &= -Q_l.\end{aligned} \tag{48.43}$$

Furthermore, the derivative of V with respect to the generalized velocities vanishes. This combined with Equation (48.41) gives

$$\begin{aligned} &\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_l}\right) - \frac{\partial T}{\partial q_l} = Q_l \\ \iff &\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_l}\right) - \frac{\partial T}{\partial q_l} = -\frac{\partial V}{\partial q_l} + \frac{\partial V}{\partial \dot{q}_l} \\ \iff &\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_l} - \frac{\partial V}{\partial \dot{q}_l}\right) - \frac{\partial}{\partial q_l}(T - V) = 0. \end{aligned} \tag{48.44}$$

If one introduces a new variable $L := T - V$, called the **Lagrangian**, one gets the **Lagrangian equation of the second kind**:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_l}\right) - \frac{\partial L}{\partial q_l} = 0. \tag{48.45}$$

□

Proof based on the principle of least action. First, recall the definition of the **action**:

$$I[q] := \int_{t_1}^{t_2} L(q(t), \dot{q}(t), t)dt. \tag{48.46}$$

The principle of least action (**Hamilton's principle**) postulates that the action is minimal for the physically relevant path. To this end, define a family of paths

$$q(t, \alpha) = q(t) + \alpha\eta(t), \tag{48.47}$$

where $\eta(t)$ is an arbitrary function satisfying the following boundary conditions:

$$\begin{cases} \eta(t_1) = 0, \\ \eta(t_2) = 0. \end{cases} \tag{48.48}$$

If the action integral is extended to such a family, the integral (48.46) becomes a function of α:

$$I(\alpha) = \int_{t_1}^{t_2} L(q(t, \alpha), \dot{q}(t, \alpha), t)dt. \tag{48.49}$$

Requiring that the action integral is stationary for the physical path $q(t)$, i.e. $\alpha = 0$, is equivalent to requiring that the derivative at $\alpha = 0$ vanishes:

$$\left.\frac{dI}{d\alpha}\right|_{\alpha=0} = 0. \tag{48.50}$$

As one evaluates this derivative at $\alpha = 0$, $q(t,\alpha)$ can be replaced by $q(t)$ due to (48.47):

$$\begin{aligned}\frac{dI}{d\alpha} &= \int_{t_1}^{t_2}\left[\frac{\partial L}{\partial q}\frac{\partial q}{\partial \alpha} + \frac{\partial L}{\partial \dot{q}}\frac{\partial \dot{q}}{\partial \alpha}\right]dt \\ &= \int_{t_1}^{t_2}\left[\frac{\partial L}{\partial q}\eta(t) + \frac{\partial L}{\partial \dot{q}}\dot{\eta}(t)\right]dt.\end{aligned} \tag{48.51}$$

By applying integration by parts to the second term in this integral, one obtains

$$\begin{aligned}\frac{dI}{d\alpha} &= \int_{t_1}^{t_2}\left[\frac{\partial L}{\partial q}(t)\eta(t) + \frac{\partial L}{\partial \dot{q}}(t)\dot{\eta}(t)\right]dt \\ &= \int_{t_1}^{t_2}\left[\frac{\partial L}{\partial q}(t)\eta(t) + \frac{\partial L}{\partial \dot{q}}(t)\frac{d\eta}{dt}\right]dt \\ &= \int_{t_1}^{t_2}\frac{\partial L}{\partial q}(t)\eta(t)dt + \eta(t_2)\frac{\partial L}{\partial \dot{q}}(t_2) - \eta(t_1)\frac{\partial L}{\partial \dot{q}}(t_1) - \int_{t_1}^{t_2}\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right)\eta(t)dt.\end{aligned} \tag{48.52}$$

Due to the initial conditions (48.50) for the function η, the second and third term vanish:

$$\frac{dI}{d\alpha} = \int_{t_1}^{t_2}\left[\frac{\partial L}{\partial q} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right)\right]\eta(t)dt. \tag{48.53}$$

Furthermore, because the function η was arbitrary, the only possible way that this derivative zero is when the integrand itself is zero:

$$\frac{\partial L}{\partial q} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) = 0. \tag{48.54}$$

Comparing this result to Equation (48.45) shows that one can also obtain the **Lagrangian equations of the second kind** by starting from the principle of least action. □

Definition 48.2.8 (Cyclic coordinate). If the Lagrangian L does not explicitly depend on a coordinate q_k, the coordinate is said to be cyclic.

Theorem 48.2.9 (Noether). *The conjugate momentum of a cyclic coordinate is a conserved quantity:*

$$\dot{p}_k \overset{48.2.3}{=} \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^k}\right) \overset{(48.35)}{=} \frac{\partial L}{\partial q^k} = 0. \tag{48.55}$$

48.2.3 Kepler problem

Formula 48.2.10 (Potential for a point mass).

$$V = -G\frac{M}{r}, \tag{48.56}$$

where $G = 6.67 \times 10^{-11}\frac{\text{Nm}^2}{\text{kg}^2}$ is the **gravitational constant**.

To solve the Kepler problem it is useful to perform a coordinate transformation. At the same time one passes to the center-of-mass reference frame

$$\vec{\boldsymbol{R}} := \frac{m_1\vec{\boldsymbol{r}}_1 + m_2\vec{\boldsymbol{r}}_2}{m_1 + m_2} \qquad M := m_1 + m_2 \tag{48.57}$$

and also introduces the "reduced mass":

$$\vec{\boldsymbol{r}} := \vec{\boldsymbol{r}}_1 - \vec{\boldsymbol{r}}_2 \qquad \mu := \frac{m_1 m_2}{m_1 + m_2}. \tag{48.58}$$

With these coordinates, the Kepler Lagrangian becomes:

$$L_{\text{Kepler}} = \frac{1}{2}M\dot{R}^2 + \frac{1}{2}\mu\dot{r}^2 + \frac{GM}{r}. \tag{48.59}$$

The first benefit of this transformation is that the equations of motion decouple. The center of mass behaves as a free particle:

$$M\ddot{\vec{\boldsymbol{R}}} = 0. \tag{48.60}$$

The displacement vector and reduced mass are those of a particle influenced by a gravitational force:

$$\mu\ddot{\vec{\boldsymbol{r}}} = \frac{GM}{r^2}\hat{\boldsymbol{e}}_{\boldsymbol{r}}. \tag{48.61}$$

Theorem 48.2.11 (Bertand). *The only central force problems with bound orbits, whose orbits are all closed, are given by the inverse potential*

$$V(r) \sim \frac{1}{r^2} \tag{48.62}$$

and harmonic potential

$$V(r) \sim r^2. \tag{48.63}$$

Definition 48.2.12 (Laplace-Runge-Lenz vector). Although all conservative central force problems have some common symmetries (time translation and rotational symmetries), 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 35.4.5 instead of a kinematical one 35.4.4, i.e. it cannot simply be deduced from the Lagrangian description.

Before explaining the symmetry itself, the associated Noether charge is given:

$$\vec{\boldsymbol{A}} := \vec{\boldsymbol{p}} \times \vec{\boldsymbol{L}} - mk\vec{\boldsymbol{r}}, \tag{48.64}$$

where m denotes the mass and k fixes the scale of the potential:

$$V(r) = \frac{k}{r}. \tag{48.65}$$

The reason why the LRL vector 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.

?? FINISH ??

48.3 Hamiltonian mechanics

Definition 48.3.1 (Canonical coordinates). Consider the Lagrangian 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 48.3.2 (Hamiltonian function). Given a Lagrangian L, the Hamiltonian function is defined by the following Legendre transformation 14.8.6:

$$H(q, p, t) := \sum_i p_i \dot{q}^i - L(q, p, t). \tag{48.66}$$

Formula 48.3.3 (Hamilton's equations). Inserting the above definition in the action 48.2.5 and applying the variational principle results in the following equations:

$$\dot{q}^i = \frac{\partial H}{\partial p_i}, \tag{48.67}$$

$$\dot{p}_i = -\frac{\partial H}{\partial q^i}. \tag{48.68}$$

Systems obeying these equations are called **Hamiltonian systems**. (See Section 35.4 for a formal introduction).

48.4 Poisson brackets

Definition 48.4.1 (Poisson bracket). To stick with the conventions of Definition 35.2.2, the Poisson bracket is defined as

$$\{A, B\} = \frac{\partial A}{\partial p}\frac{\partial B}{\partial q} - \frac{\partial A}{\partial q}\frac{\partial B}{\partial p}, \tag{48.69}$$

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 48.4.2 (Total time derivative). Hamilton's equations imply the following expression for the total time-derivative:

$$\frac{dF}{dt} = \frac{\partial F}{\partial t} + \{H, F\}, \tag{48.70}$$

where H is the Hamiltonian 48.3.2.

48.5 Hamilton-Jacobi equation

For a geometrical interpretation see Section 35.4.2.

Definition 48.5.1 (Canonical transformations). A canonical transformation is a transformation that leaves the Hamiltonian equations of motion unchanged. Mathematically this means that the transformations leave the action invariant up to a constant or, equivalently, they leave the Lagrangian invariant up to a complete time-derivative:

$$\sum_i p_i \dot{q}^i - H(q, p, t) = \sum_i P_i \dot{Q}^i - K(Q, P, t) - \frac{dS}{dt}(Q, P, t). \tag{48.71}$$

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 48.5.2 (Hamilton-Jacobi equation). Sufficient conditions for a function S to be a generating function of canonical transformations are:

$$p_i = \frac{\partial S}{\partial q^i},$$
$$Q^i = \frac{\partial S}{\partial P_i},$$

and

$$K = H + \frac{\partial S}{\partial t}.$$

Choosing the new Hamiltonian function K to be 0 (this function is sometimes called the **Kamiltonian**) gives the Hamilton-Jacobi equation:

$$H\left(q, \frac{\partial S}{\partial q}, t\right) + \frac{\partial S}{\partial t} = 0. \tag{48.72}$$

The function S is called **Hamilton's principal function**.

Property 48.5.3. The new coordinates P_i and Q^i are all constants of motion. This follows immediately from the choice $K = 0$.

Definition 48.5.4 (Hamilton's characteristic function). For time-independent systems the HJE can be rewritten as follows:

$$H\left(q, \frac{\partial S}{\partial q}\right) = -\frac{\partial S}{\partial t} =: E. \tag{48.73}$$

After a redefinition of H this is the same as Equation (35.39). One thus obtains the classical result that for time-independent systems the Hamiltonian function is a constant of motion (often called the **energy**[2]). Integration with respect to time gives the following form of the principal function:

$$S(q, p, t) = W(q, p) - Et. \tag{48.74}$$

The time-independent function W is called Hamilton's characteristic function.

Property 48.5.5 (Stäckel condition). The Hamilton-Jacobi equation is separable if and only if the potential is of the form

$$V(q) = \sum_{i=1}^{n} \frac{W_i(q^i)}{G_i^2(q)} \tag{48.75}$$

whenever the Hamiltonian function can be written as

$$H(q, p) = \frac{1}{2}\sum_i \frac{p_i^2}{G_i^2(q)} + V(q). \tag{48.76}$$

Potentials of this form are called **Stäckel potentials**.

[2]Note that this is not a general fact.

48.6 Analytical mechanics

48.6.1 Phase space

Definition 48.6.1 (Phase space). The set of all possible n-tuples (q^i, p_i) of generalized coordinates and associated momenta is called the phase space of the system. Note that this also includes the tuples that are not solutions of the equations of motion.

Definition 48.6.2 (Libration). A closed trajectory 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 48.6.3 (Rotation). A closed trajectory for which at least one of the variables takes on all possible values. Topologicaly it is characterized by a noncontractible, closed trajectory.

Definition 48.6.4 (Separatrix). When plotting different (closed) trajectories in the phase space of a system, the curve that separates regions of librations and rotations is called the separatrix.[3]

Definition 48.6.5 (Lagrangian derivative[4]). Let $a(q, \dot{q}, t)$ be a phase space quantity. Its Lagrangian derivative along a path $(q(t), \dot{q}(t))$ in phase space is given by

$$\begin{aligned}\frac{Da}{Dt} &:= \lim_{\Delta t\to 0}\frac{a(q+\Delta q, \dot{q}+\Delta\dot{q}, t+\Delta t) - a(q,\dot{q},t)}{\Delta t}\\ &= \frac{\partial a}{\partial t} + \frac{dq^i}{dt}\frac{\partial a}{\partial q^i} + \frac{d\dot{q}^i}{dt}\frac{\partial a}{\partial \dot{q}^i}\\ &= \frac{\partial a}{\partial t} + \dot{q}^i\frac{\partial a}{\partial q^i} + \frac{d\dot{q}^i}{dt}\frac{\partial a}{\partial \dot{q}^i}\,.\end{aligned} \tag{48.77}$$

The second term $\dot{\vec{q}}\cdot\nabla a$ in this equation is called the **advective** term.

Remark 48.6.6. In the case that $a(q, \dot{q}, t)$ is a tensor field, the gradient ∇ has to be replaced by the covariant derivative. The advective term is then sometimes called the **convective** term.

Corollary 48.6.7. For $a(q, \dot{q}, t) = \vec{q}$ one obtains

$$\frac{D\vec{q}}{Dt} = \dot{\vec{q}}. \tag{48.78}$$

48.6.2 Liouville's theorem

Formula 48.6.8 (Liouville's lemma). Consider a phase space volume element dV_0 moving along a path $(q(t), \dot{q}(t)) \equiv x(t)$. The Jacobian $J(x(t))$ associated with this motion is given by

$$J(x(t)) = \frac{dV}{dV_0} = \det\left(\frac{\partial\vec{x}}{\partial\vec{x}_0}\right) = \sum_{ijklmn}\varepsilon^{ijklmn}\frac{\partial x^1}{\partial x_0^i}\frac{\partial x^2}{\partial x_0^j}\frac{\partial x^3}{\partial x_0^k}\frac{\partial x^4}{\partial x_0^l}\frac{\partial x^5}{\partial x_0^m}\frac{\partial x^6}{\partial x_0^n}. \tag{48.79}$$

The Lagrangian derivative of this Jacobian is

$$\frac{DJ}{Dt} = (\nabla\cdot\vec{x})J. \tag{48.80}$$

Furthermore, using Hamilton's equations 48.3.3, it is easy to prove that

$$\nabla\cdot\vec{x} = 0, \tag{48.81}$$

and, hence, that the material derivative of J vanishes.

[3] In general the separatrix of a dynamical system is a curve that separates regions with different behaviour.

[4] Also known as the **material derivative**, especially when applied to fluid mechanics.

Theorem 48.6.9 (Liouville). *Let $V(t)$ be a moving phase space volume containing a fixed set of particles. Applying Liouville's lemma gives*

$$\frac{DV}{Dt} = \frac{D}{Dt}\int_{\Omega(t)} d^6x = \frac{D}{Dt}\int_{\Omega_0} J(\vec{\boldsymbol{x}},t)d^6x_0 = 0. \tag{48.82}$$

This implies that the phase space volume of a Hamiltonian system is invariant with respect to time-evolution.

Remark 48.6.10 (♣). Any phase space admits a symplectic form ω such that the Hamiltonian equations of motion are encoded in a vector field X_H. By noting that the volume in phase space is calculated through the symplectic volume form $\mathrm{Vol}_\omega \sim \omega^n$, this theorem easily follows from the fact that time evolution is the flow of the Hamiltonian vector field, which preserves the symplectic form.

Formula 48.6.11 (Boltzmann's transport equation). Let $F(q,\dot{q},t)$ be the mass distribution function

$$M_{\text{tot}} = \int_{\Omega(t)} F(q,\dot{q},t)d^6x. \tag{48.83}$$

From the conservation of mass one can derive the following formula:

$$\frac{DF}{Dt} = \frac{\partial F}{\partial t} + \frac{dq^i}{dt}\cdot\frac{\partial F}{\partial q^i} - \nabla V^i\frac{\partial F}{\partial\dot{q}^i} = \left[\frac{\partial F}{\partial t}\right]_{\text{col}}, \tag{48.84}$$

where the right-hand side gives the change of F due to collisions.[5] This partial differential equation in 7 variables can be solved to obtain F.

Consider a Hamiltonian system with a phase space $\mathcal{V}$. By Liouville's theorem, the phase flow generated by the equations of motion is a volume- or measure-preserving map $g:\mathcal{V}\rightarrow\mathcal{V}$ (Definition 16.1.37). This gives rise to the following theorem:

Theorem 48.6.12 (Poincaré recurrence theorem). *Let $\mathcal{V}_0$ be the initial phase space volume of the system. For every point $x_0\in\mathcal{V}_0$ and for every neighbourhood U of x_0, there exists a point $y\in U$ such that $g^n(y)\in U$ for every $n\in\mathbb{N}$.*

Theorem 48.6.13 (Strong Jeans's theorem[6]). *For 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.

48.6.3 Continuity equation

Formula 48.6.14 (Reynolds transport theorem). Consider a quantity

$$F = \int_{V(t)} f(q,\dot{q},t)dV.$$

Combining Equation (48.80) with the divergence theorem 21.2.6 gives

$$\frac{DF}{Dt} = \int_V \frac{\partial f}{\partial t}dV + \oint_{\partial V} f\dot{\vec{\boldsymbol{q}}}\cdot d\vec{\boldsymbol{S}}. \tag{48.85}$$

This formula can be interpreted as a three-dimensional generalization of the *Leibniz integral rule*.

[5]The collisionless form of this equation is sometimes called the **Vlasov equation**.
[6]Actually due to *Donald Lynden-Bell*.

Formula 48.6.15 (Continuity equations). For a conserved quantity the equation above becomes:

$$\frac{Df}{Dt} + (\nabla \cdot \dot{\vec{q}})f = 0 \tag{48.86}$$

$$\frac{\partial f}{\partial t} + \nabla \cdot (f\dot{\vec{q}}) = 0. \tag{48.87}$$

If one sets $f = \rho$ (the mass density), the first equation is called the **Lagrangian continuity equation** and the second equation is called the **Eulerian continuity equation**. Both equations can be found by pulling the Lagrangian derivative inside the integral on the left-hand side of 48.6.14.

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 given point (or particle) in the system and measures how a given quantity evolves around the chosen point as it moves throughout space.

Corollary 48.6.16. Combining the Reynolds transport theorem with the Lagrangian continuity equation gives the following identity for an arbitrary function f:

$$\frac{D}{Dt}\int_V \rho f dV = \int_V \rho \frac{Df}{Dt} dV. \tag{48.88}$$

48.6.4 Fluid mechanics

Theorem 48.6.17 (Cauchy's stress theorem[7]). *Knowing the stress vectors acting on the coordinate planes through a point is sufficient to calculate 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 48.6.18 (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 $\vec{n}$ to that plane:

$$\vec{t}_{(\vec{n})} = \mathbf{T}(\vec{n}). \tag{48.89}$$

Example 48.6.19. For identical particles the stress tensor is given by

$$\mathbf{T} = -\rho\langle \vec{w} \otimes \vec{w}\rangle, \tag{48.90}$$

where $\vec{w}$ is the random component of the velocity vector and $\langle\cdot\rangle$ denotes the expectation value 43.4.1.

Theorem 48.6.20 (Cauchy's lemma). *The stress vectors acting on opposite planes are equal in magnitude but opposite in direction:*

$$\vec{t}_{(-\vec{n})} = -\vec{t}_{(\vec{n})}. \tag{48.91}$$

[7] Also known as **Cauchy's fundamental theorem**.

Formula 48.6.21 (Cauchy momentum equation). From Newton's second law 48.1 it follows that

$$\frac{D\vec{\boldsymbol{P}}}{Dt} = \int_V \vec{\boldsymbol{f}}(x,t)dV + \oint_{\partial V} \vec{\boldsymbol{t}}(x,t)dS, \tag{48.92}$$

where $\vec{\boldsymbol{P}}$ is the momentum density, $\vec{\boldsymbol{f}}$ are the "body" forces and $\vec{\boldsymbol{t}}$ are the surface forces (such as *shear stress*). Using Cauchy's stress theorem and the divergence theorem 21.2.6 one obtains

$$\frac{D\vec{\boldsymbol{P}}}{Dt} = \int_V \Big[\vec{\boldsymbol{f}}(x,t) + \nabla\cdot\mathbf{T}(x,t)\Big]dV. \tag{48.93}$$

The left-hand side can be rewritten using Equation (48.88) as

$$\int_V \rho\frac{D\vec{\boldsymbol{v}}}{Dt}dV = \int_V \Big[\vec{\boldsymbol{f}}(x,t) + \nabla\cdot\mathbf{T}(x,t)\Big]dV. \tag{48.94}$$

48.7 Dynamical systems

The following property, although seemingly innocuous, is very important for the study of mechanical systems:

Property 48.7.1. For dynamical systems governed by ODEs satisfying the Picard-Lindelöf conditions 18.2.1, different trajectories never intersect.

The above property has an important (but nontrivial) consequence

Theorem 48.7.2 (Poincaré-Bendixson). *In a phase plane, i.e. 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 48.7.3. In two-dimensional (Cartesian) phase spaces there cannot exist chaos, i.e. no strange attractors can exist.

Definition 48.7.4 (Lypanuov exponents). Consider two trajectories of a system 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

$$s(t) \approx e^{\lambda t}s_0, \tag{48.95}$$

λ is called the Lyapunov exponent of the system.

Definition 48.7.5 (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 48.7.6 (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 $U \subset S$ is open and connected.
- $P|_{P(U)}$ is a diffeomorphism.

- For every point $u \in U$, the positive semi-orbit of u intersect 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 48.7.7 (Fixed points). Fixed points of the Poincaré map correspond to closed orbits.

48.8 Geometric mechanics

In this section the current chapter is reformulated in a differential geometric framework. 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 $\frac{1}{2}g(v, v)$, together with a second-order ODE in the form of a vector field on TM such that $\pi_*(X_v) = v$, where π denotes the tangent bundle projection.

For integral curves γ 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:

$$\dot{q}^i(t) = \frac{dq^i}{dt}(t). \tag{48.96}$$

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:

$$X = \dot{q}^i \frac{\partial}{\partial q^i} + F^i(q, \dot{q}) \frac{\partial}{\partial \dot{q}^i}. \tag{48.97}$$

By writing this vector field as a system of differential equations, a second-order ODE is obtained (hence the terminology):

$$\frac{d^2 q^i}{dt^2} = F^i(q, \dot{q}). \tag{48.98}$$

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 28.2.47. By adopting the notation of Equation (34.22) one can generalize the geodesic equation to obtain Newton's equations of motion for an arbitrary smooth "potential" $U : M \to \mathbb{R}$:

Formula 48.8.1 (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

$$\nabla_{\dot{\sigma}} \dot{\sigma} = -\mathrm{grad}(U) \tag{48.99}$$

where ∇ indicates the Levi-Civita connection and grad denotes the gradient operator 21.1.1 (here not denoted by ∇ to avoid confusion with the covariant derivative).

48.8.1 Lagrangian mechanics

Now it is time to turn to Noether's theorem and, in particular, the version concerning cyclic coordinates 48.2.9. Any diffeomorphism of M induces a diffeomorphism on TM by pushforward.

Definition 48.8.2 (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:

$$\widehat{X}(v) := g(X_{\pi(v)}, v). \tag{48.100}$$

Using this definition Noether's theorem 48.2.9 can be reformulated as follows:

Theorem 48.8.3 (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:

$$X_U := \dot{q}^i \frac{\partial}{\partial q^i} + \frac{\partial L}{\partial q^i}\frac{\partial}{\partial p_i}. \tag{48.101}$$

After performing the Legendre transformation $E := p_i\dot{q}^i - L$ to obtain the (Hamiltonian) energy function[8], Newton's equations can be rewritten in Hamiltonian form:

$$X_E = \frac{\partial E}{\partial p_i}\frac{\partial}{\partial q^i} - \frac{\partial E}{\partial q^i}\frac{\partial}{\partial p_i}. \tag{48.102}$$

48.8.2 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 48.8.4 (Fibre derivative). Let $L : TM \to \mathbb{R}$ be a smooth Lagrangian. The fibre derivative of L is defined as the directional (Gâteaux) derivative 23.3.12 of L:

$$\langle \mathbb{F}L(v), w\rangle := \left.\frac{d}{dt}\right|_{t=0} L(v + tw). \tag{48.103}$$

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

$$\mathbb{F}L : (q^i, \dot{q}^i) \mapsto \left(q^i, \frac{\partial L}{\partial \dot{q}^i}\right) \equiv (q^i, p_i). \tag{48.104}$$

As such the classical definition 48.2.3 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 Equation (48.100) of conjugate momenta given in the previous paragraph.

Remark 48.8.5 (Legendre transform). The fibre derivative $\mathbb{F}L$ is often called the Legendre transform of L. Although this does not exactly coincide with 14.8.6 or (44.17), the relation is simple enough. The Legendre transformation $L \mapsto L^*$ (on the tangent bundle) is implemented as $L^*(x) = \langle \mathbb{F}L(x), x\rangle - L(x)$.

[8] This function will not be called a Hamiltonian function, because this will be reserved for functions on the cotangent bundle.

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, 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 such that 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**.

48.8.3 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 $T^*M \times \mathbb{R}$:

$$\alpha := p_i dq^i - H dt \equiv \omega - H dt. \tag{48.105}$$

First of all, this endows the extended cotangent bundle with a contact structure 36.1.3. Moreover, it can be shown to be a relative integral invariant 32.5.10 of the following Pfaffian system:

$$\frac{dq^i}{\partial H/\partial p_i} = dt = \frac{-dp_i}{\partial H/\partial q^i}. \tag{48.106}$$

The vector fields that leave the contact form invariant are exactly those generated by the Hamiltonian vector field X_H. Liouville's theorem 48.6.9 is then simply a consequence of the absolute invariance of $d\alpha$.

A canonical transformation 48.5.1 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 48.8.6 (Canonical transformation). A transformation of the trivial line bundle $T^*M \times \mathbb{R}$ that leaves both the fibres and the **symplectization**[9]

$$\Omega := d(e^\lambda \alpha) \tag{48.107}$$

invariant.

?? CHECK THIS STATEMENT (cannot find source) ??

48.8.4 Symplectic structure on infinite-dimensional systems

Although this section could have been included in Chapter 35, it better fits in here since the study of these manifolds is almost always related to the study of physical phenomena such as *solitons*.

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

[9]This form turns $T^*M \times \mathbb{R}^2$ into a symplectic manifold.

is preserved, the content of Remark 35.2.7 applies also to infinite-dimensional systems, 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:

$$\{F, G\} := \omega(X^F, X^G). \tag{48.108}$$

?? COMPLETE (e.g. Palais, cursus Antwerpen) ??

Chapter 49

Constrained mechanics

The foundations for this subject were laid down by *Dirac* in [70]. By introducing constraints, the coordinates and their momenta are not independent anymore. This implies for example that the Hamiltonian equations of motion have to be modified.

49.1 Constraint surface

First, recall the Lagrangian equations of motion (48.35). 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 48.2.3 cannot be inverted to express velocities in terms of momenta. Alternatively, the momenta p and coordinates q are not independent and there must exist relations of the form

$$\phi(q, p) = 0. \tag{49.1}$$

Constraints of this type constraints 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 49.1 (Regularity conditions). It will always be assumed that the independent constraints, i.e. the minimal subset of constraints that imply the others, satisfy the following (equivalent) conditions:

- The constraints can locally serve as the first coordinates of a (regular) coordinate system.
- The differentials (gradients) $d\phi_m$ are locally linearly independent.
- The variations $\delta\phi_m$ 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 constrained equations $\phi_m(q, p) = 0, m \in I$, the **primary constraints**. If this sytem was 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:

$$\dot{q}^i = \frac{\partial H}{\partial p_i} + \sum_{m\in I} u_m \frac{\partial \phi_m}{\partial p_i} \tag{49.2}$$

$$\dot{p}_i = -\frac{\partial H}{\partial q^i} - \sum_{m\in I} u_m \frac{\partial \phi_m}{\partial q^i}, \tag{49.3}$$

where the u_m are functions of the coordinates and velocities that play a role similar to ordinary Lagrange multipliers.

Remark 49.1.1. The above relations follow from the general 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

$$\begin{cases} \lambda_i = \sum_{m\in I} u_m \dfrac{\partial\phi_m}{\partial q^i} \\ \mu^i = \sum_{m\in I} u_m \dfrac{\partial\phi_m}{\partial p_i}. \end{cases} \tag{49.4}$$

Combining this result with the usual derivation of Hamilton's equations from a Lagrangian action principle gives the above modified equations.

In terms of Poisson brackets the constrained time evolution of a (time-independent) function is given by

$$\dot{f} = \{H, f\} + \sum_{m\in I} u_m\{\phi_m, f\}. \tag{49.5}$$

Method 49.1.2 (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 30.2.3. Furthermore, after working out the Poisson brackets one can use the constraint equations to drop all terms that are proportional to ϕ_m.

For example, Equation (49.5) can be rewritten as

$$\dot{f} = \{H + \sum_{m\in I} u_m\phi_m, f\}. \tag{49.6}$$

To prove the equivalence, one can use the linearity and Leibniz properties. This involves the following equality

$$\{u_m\phi_m, f\} = \{u_m, f\}\phi_m + u_m\{\phi_m, f\}. \tag{49.7}$$

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 49.1.3 (Weak equality). The constraints ϕ_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 $\phi_m \approx 0$ is only a weak equality.

Using the above definitions one can write an arbitrary time derivative as

$$\dot{f} \approx \{H_T, f\}, \tag{49.8}$$

where $H_T := H + \sum_{m\in I} u_m\phi_m$.

Remark 49.1.4 (Closure). An important remark regarding weak equalities can be found by taking a Poisson bracket of 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 $f \approx 0$, then $\{f, g\}$ does not even have to vanish at all.

Property 49.1.5 (Consistency conditions). By taking $f = \phi_n$ for any $n \in I$ in Equation (49.5) a set of consistency conditions is obtained:

$$\{H, \phi_n\} + \sum_{m \in I} u_m \{\phi_m, \phi_n\} \approx 0. \tag{49.9}$$

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

  $$\chi(q, p) = 0. \tag{49.10}$$

  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

$$u_m = U_m + v_a V_m^a, \tag{49.11}$$

where U_m is a solution of the inhomogeneous equation and the V_m^a are linearly independent solutions of the homogeneous equation

$$\sum_{m \in I} u_m \{\phi_m, \phi_n\} = 0. \tag{49.12}$$

The resulting coefficients v_a are completely arbitrary functions of time. Therefore, the total Hamiltonian can be written in the form

$$H_T = H'(q, p) + v^a(t)\phi_a(q, p), \tag{49.13}$$

where $\phi_a := \sum_{m \in I} V_m^a \phi_m$. 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 49.1.6 (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 ϕ_a are first-class. The number of arbitrary coefficients v^a is equal to the number of primary first-class constraints.

Notation 49.1.7. To distinguish between first- and second-class constraints, the latter are often denoted by a separate symbol χ.

Property 49.1.8 (Closure). The Poisson bracket of two primary first-class functions is first-class. So is the Poisson bracket of the total Hamiltonian and a first-class primary constraint.

Remark 49.1.9 (Dirac conjecture). The primary first-class constraints ϕ_a generate gauge transformations in the sense that variations in the coefficients v^a, 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

$$H_E := H_T + v^b(t)\phi_b(q,p), \tag{49.14}$$

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]

Corollary 49.1.10. 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's theorem 32.5.5.

Formula 49.1.11 (Degrees of freedom). The number of degrees of freedom is given by the following formula:

$$\begin{aligned} 2\times\text{number of d.o.f.} = {} & \text{number of canonical coordinates} \\ & - \text{number of second-class constraints} \\ & - 2\times\text{number of first-class constraints.} \end{aligned}$$

Definition 49.1.12 (Dirac bracket). To take care of second-class constraints, *Dirac* introduced a modification of the Poisson bracket:

$$\{f,g\}_D := \{f,g\} - \{f,\chi_a\}C^{ab}\{\chi_b,g\}, \tag{49.15}$$

where the χ_a's 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 defines a Poisson algebra 30.2.3. 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 49.1.13. 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 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 49.1.14 (Gauge-invariant functions). Consider the algebra of smooth functions on phase space $C^\infty(M)$. 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

[1] Note that H_T is the Hamiltonian that corresponds to a Lagrangian approach. H_E gives a more general theory.

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**.

The restriction to gauge-invariant functions is also imperative if one wants to extend the bracket operation 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.

Property 49.1.15 (Geometric characterization). Restricted to a first-class constraint surface, the "symplectic" form becomes maximally degenerate with rank $\mathrm{rk}(\omega) = \dim(M) - 2\dim(\Sigma)$. 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 $\mathrm{rk}(\omega) = \dim(M) - \dim(\Sigma)$, which implies that the surface is isotropic. Furthermore, the induced Poisson bracket coincides with the restriction of the Dirac bracket.

Remark 49.1.16 (Algebraic characterization). The fact that first-class constraints define a coisotropic submanifold is not a peculiarity. A multiplicative ideal of a Poisson algebra 30.2.3 that is closed under the Poisson bracket, is often called a **coisotrope** (or coisotropic ideal). Coisotropic submanifolds of a Poisson manifold 35.2.4 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 $\mathcal{I}$, one can pass to the quotient $N(\mathcal{I})/\mathcal{I}$, 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.)

49.2 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 27.1.2. In general the phase space will be assumed to be a *supermanifold*.

First, the ordinary Poisson bracket is extended to Grassmann-odd coordinates as follows:

$$\{\theta^i, \theta^j\} = 0 = \{\pi_i, \pi_j\} \tag{49.16}$$

and

$$\{\theta^i, \pi_j\} = \delta^i_j = \{\pi_j, \theta^i\}. \tag{49.17}$$

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:

$$\{f, g\} := \frac{\partial^R f}{\partial z^i}\sigma^{ij}\frac{\partial^L g}{\partial z^j}. \tag{49.18}$$

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

$$\{f, g\} = \left(\frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q^i} - \frac{\partial f}{\partial q^i}\frac{\partial g}{\partial p_i}\right) + (-1)^{\deg(f)}\left(\frac{\partial f}{\partial \pi_i}\frac{\partial g}{\partial \theta^i} + \frac{\partial f}{\partial \theta^i}\frac{\partial g}{\partial \pi_i}\right). \tag{49.19}$$

The algebraic properties of this generalized Poisson bracket are graded generalizations of those of the ordinary one:

$$\{f, g\} = -(-1)^{\deg(f)\deg(g)}\{g, f\} \tag{49.20}$$

$$\begin{aligned}0 = \{f, \{g, h\}\} &+ (-1)^{[\deg(f)+\deg(g)]\deg(h)}\{h, \{f, g\}\}\\ &+ (-1)^{\deg(f)[\deg(g)+\deg(h)]}\{g, \{h, f\}\}\end{aligned} \tag{49.21}$$

$$\{f, gh\} = \{f, g\}h + (-1)^{\deg(f)\deg(g)}f\{g, h\} \tag{49.22}$$

$$\deg(\{f, g\}) = \deg(f) + \deg(g). \tag{49.23}$$

The first two properties state that the generalized Poisson bracket gives rise to a Lie superalgebra 27.7.2. 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*.

49.3 BRST symmetry

49.3.1 Introduction

Consider a dynamical system (M, ω, H) together with a set of first-class constraints $\{\phi_a\}_{a\in I}$. For further convenience the constraints will be assumed to be **irreducible**, i.e. their Jacobian is assumed to be of full rank on the constraint surface. As was shown before, these constraints generate an algebra under the Poisson bracket. However, more structure exists. To explore this structure, enlarge the phase space 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.

$$\{\mathcal{P}_a, \eta^b\} := -\delta^b_a, \tag{49.24}$$

of degrees $\mathrm{gh}(\mathcal{P}_a) = -1$ and $\mathrm{gh}(\eta^a) = 1$.

Remark 49.3.1 (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, which is the BRST-partner of the *Faddeev-Popov antighost field* $\bar{c}$, i.e. $\{\bar{c}, \Omega\} = B$.

Definition 49.3.2 (BRST operator). To any dynamical system governed by first-class constraints $\{\phi_a\}_{a\in I}$ one can associate a BRST operator (or function) defined by the following conditions:

1. It is of (ghost) degree 1:
$$\mathrm{gh}(\Omega) = 1. \tag{49.25}$$
2. It is nilpotent with respect to the Poisson bracket:
$$\{\Omega, \Omega\} = 0. \tag{49.26}$$
3. It is Hermitian:
$$\Omega^* = \Omega. \tag{49.27}$$

[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 constraint.

4. It is proportional to the constraints in lowest order:

$$\Omega = \eta^a\phi_a + \text{terms in ghost momenta.} \tag{49.28}$$

Example 49.3.3 (Abelian constraint algebra). In the case where the constraints form an Abelian algebra, i.e. $\{\phi_a, \phi_b\} = 0$, the BRST operator has a simple expression:

$$\Omega = \eta^a\phi_a. \tag{49.29}$$

Example 49.3.4 (Closed constraint algebra). In the case where the constraints form a closed algebra, i.e. $\{\phi_a, \phi_b\} = C^c_{ab}\phi_c$ with C^c_{ab} constant, the BRST operator has a slightly more 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:

$$\Omega = \eta^a\phi_a - \frac{(-1)^{\varepsilon_b}}{2}\eta^b\eta^c C^a_{bc}\mathcal{P}_a. \tag{49.30}$$

Remark 49.3.5 (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.)

Because of the algebraic properties above, the BRST operator defines a cohomology theory 5.1.3:

Property 49.3.6 (BRST cohomology). For all functions f on the extended phase space, one has the following equality:

$$\{\{f, \Omega\}, \Omega\} = 0. \tag{49.31}$$

In view of this structure one defines a BRST-closed function as a function f satisfying

$$\{f, \Omega\} = 0 \tag{49.32}$$

and a BRST-exact function as a function f that can be written as

$$f = \{g, \Omega\} \tag{49.33}$$

for some function g. It is clear that the resulting BRST cohomology theory is gauge-invariant, since Ω is gauge-invariant.

It can also be shown that the BRST operator only depends on the constraint surface and not on the choice of a local description:

Property 49.3.7 (Uniqueness). Any two BRST generators Ω, Ω' associated to the same constraint surface are related by a canonical transformation on the extended phase space.

For negative ghost numbers it can be shown that BRST cohomology vanishes. In degree 0 the BRST cohomology is characterized as follows:

Property 49.3.8 (Gauge-invariant functions). $H^0(\Omega)$ 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.

Property 49.3.9 (Poisson bracket). The Poisson bracket descends to $H^0(\Omega)$ 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_0g_0 and $\{f_0,g_0\}$, respectively.

Example 49.3.10 (Extension of constraints). Consider a constraint ϕ_a. A BRST-invariant extension is given by the Poisson bracket

$$G_a := \{-\mathcal{P}_a, \Omega\}. \tag{49.34}$$

This immediately shows that the extension G_a corresponds to the observable 0, since it is Ω-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. The interpretation of higher cohomology groups will be adressed in Chapter 57.

Definition 49.3.11 (Gauge-fixing). Consider a classical Hamiltonian H_0 and its BRST extension H. 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**.

49.3.2 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 5.4.15 associated to the first-class constraints ϕ_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 49.3.12 (Antighost number). The degree induced by δ is called the **antighost number**. The Koszul generators $\mathcal{P}_a$, associated to the constraints ϕ_a, have antighost number 1 and have Grassmann parity $\varepsilon_a + 1$. This also implies that $\text{antigh}(\delta) = -1$.

A second step is to characterize the gauge structure on 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 ϕ_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**.

The Hamiltonian vector fields associated to the first-class constraints are tangent to the gauge orbits. Furthermore, by the irreducibility of the constraints, the vector fields form a frame field for the tangent bundle of the gauge orbits.

Definition 49.3.13 (Longitudinal complex). Vector fields that are tangent to the gauge orbits are called longitudinal or vertical vector fields (not to be confused with the vertical vector fields from Section 33.3). A frame is given by the Hamiltonian vector fields $X_a := \{\phi_a, \cdot\}$.

Longitudinal forms η^a are defined as the multilinear duals to the longitudinal vector fields. The tangent bundle of Σ in M can be decomposed as follows:

$$TM|_\Sigma = T\mathcal{F} \oplus N\Sigma, \tag{49.35}$$

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 $\Omega^{\bullet,0}(\Sigma)$ and the longitudinal exterior 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

$$\mathbf{d}f := df = \{\phi_a, f\}\,\eta^a, \tag{49.36}$$

$$\mathbf{d}\eta^a := -\frac{1}{2}C^a_{bc}(q,p)\eta^b \wedge \eta^c, \tag{49.37}$$

where $C^a_{bc}(q,p)$ 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 30.6. (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

$$A \equiv A_{i_1\dots i_k}(q,p)\eta^{i_1} \wedge \cdots \wedge \eta^{i_k}, \tag{49.38}$$

where the coefficients are equivalence classes of weakly equal functions in $C^\infty(M)$.

Definition 49.3.14 (Ghost number). The form degree of a longitudinal differential form is called the **(pure) ghost number**. It is denoted by ‘‘pure gh’’. This also implies that pure $\mathrm{gh}(\mathbf{d}) = 1$.

Note that the number of Koszul generators is the same as the number of ghost fields since both of these are induced by the constraints ϕ^a. To extend the Poisson bracket to the **extended phase space** containing phase space functions, ghost fields and ghost momenta, the following convention is introduced:

$$\{\mathcal{P}_a, \eta^b\} = -(-1)^{(\varepsilon_a+1)(\varepsilon_b+1)}\{\eta^b, \mathcal{P}_a\} := -\delta^b_a. \tag{49.39}$$

Definition 49.3.15 (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:

$$\mathrm{gh}(A) := \text{pure gh}(A) - \text{antigh}(A). \tag{49.40}$$

It satisfies

$$\mathrm{gh}(AB) = \mathrm{gh}(A) + \mathrm{gh}(B). \tag{49.41}$$

It can be obtained as the eigenvalue of the operator

$$\mathcal{G} := i\eta^a\mathcal{P}_a. \tag{49.42}$$

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 $\mathrm{gh}(\mathbf{d}) = 1$ and $\mathrm{antigh}(\mathbf{d}) = 0$.[3] By making the choice

$$\mathbf{d}\mathcal{P}_a := (-1)^{\varepsilon_a}\eta^c C^b_{ca}\mathcal{P}_b, \tag{49.43}$$

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 $\mathbb{C}[\eta^a]$. Because the latter is a free and, in particular, projective module, the homology of the tensor product is the tensor product of the homology with this module. The cohomology of $\mathbf{d}$ modulo δ on M can be shown to coincide with the cohomology of $\mathbf{d}$ on Σ. This cohomology theory is exactly the **BRST cohomology** from the previous section.

Homological perturbation theory 5.3.3 now also says that there exists a true differential s on the extended phase space generating BRST cohomology with the following additional properties:

$$\begin{aligned} s &= \delta + \mathbf{d} + \cdots \\ \varepsilon(s) &= 1 \\ \mathrm{gh}(s) &= 1. \end{aligned} \tag{49.44}$$

This operator can be induced by a **BRST function** Ω such that

$$sA := \{A, \Omega\} \tag{49.45}$$

as before.

49.3.3 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 for $C^\infty(\Sigma)$ anymore. Because higher-order relations exist among the constraints, the higher-order homology does 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 5.4.16 exists. 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, i.e. the Chevalley-Eilenberg algebra associated to a L_∞-algebroid.

Formula 49.3.16 (Ghost number). Due to the introduction of ghost-of-ghosts, Equation (49.42) 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

$$\mathcal{G} := i\sum_{i=1}^{\infty}\eta^{a_i}\mathcal{P}_{a_i} + \text{constants due to operator ordering}, \tag{49.46}$$

where $\mathcal{P}_{a_i}$ are the ghost-of-ghost momenta of antighost number i.

[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.

Remark 49.3.17 (Irreducible constraints). The case of irreducible constraints can be recovered from this general situation by throwing away all higher-order generators.

Remark 49.3.18 (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

$$\{\phi_a, \phi_b\} = C^c_{ab}\phi_c \tag{49.47}$$

defines a Lie algebroid in the case of irreducible constraints and a (potentially truncated) L_∞-algebroid in the case of reducible constraints. Similar to Equation (30.77) 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.

Chapter 50

Electromagnetism

50.1 Electricity

Property 50.1.1 (Conservation of charge). Charge is a conserved quantity, i.e. the total charge of a system is a constant.

Definition 50.1.2 (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.

Formula 50.1.3 (Coulomb's law). The force between two point charges q_1, q_2 is given by

$$F = \frac{q_1 q_2}{4\pi\varepsilon_0 r^2}, \tag{50.1}$$

where r is the distance between the charges and ε_0 is the permittivity of the vacuum ($\varepsilon_0 \approx 8.854 \times 10^{-12}\,\frac{\mathrm{F}}{\mathrm{m}}$).

Definition 50.1.4 (Electric potential). The electric field gives the force experienced by a single (positive) charge. The electric potential is then defined as the work done by this charge when moving through the field:

$$V(\vec{\boldsymbol{a}}, \vec{\boldsymbol{b}}) := \int_{\vec{\boldsymbol{a}}}^{\vec{\boldsymbol{b}}} \vec{\boldsymbol{E}} \cdot d\vec{\boldsymbol{l}}. \tag{50.2}$$

By fixing the endpoint (some fixed reference location[1]), the electric field and electric potential are related in the same way as the ordinary potential and the associated (conservative) force (cf. Equation (48.4)):

$$E = -\nabla V. \tag{50.3}$$

Definition 50.1.5 (Capacitance). The capacitance is a (geometrical) value that reflects the amount of charge an object can store:

$$C := \frac{Q}{V}. \tag{50.4}$$

[1] As usual the potential is only determined up to some constant.

50.1.1 Conductors and resistors

Definition 50.1.6 (Current). The amount of charge that passes a given point in one second:

$$I := \frac{dQ}{dt}. \tag{50.5}$$

Formula 50.1.7 (Power). Given that the electric potential gives the work done by a positive charge, the electric power of a current I is given by

$$P = VI. \tag{50.6}$$

Property 50.1.8 (Electric field of conductor). The electric field inside a conductor is vanishes at equilibrium. All free charges are located at the surface. Moreover, the electric field at the surface points in the direction of the normal vector.

Definition 50.1.9 (Drift velocity). The average speed of independent charge carriers is the drift velocity $\vec{\boldsymbol{v}}_d$. It is important to remark 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.

Definition 50.1.10 (Mobility).

$$\mu := \frac{v_d}{E}, \tag{50.7}$$

where E is the external electric field.

Definition 50.1.11 (Conductivity).

$$\sigma := nq\mu \tag{50.8}$$

Definition 50.1.12 (Resistivity).

$$\rho := \frac{1}{\sigma} \tag{50.9}$$

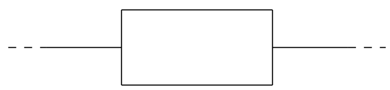
Formula 50.1.13 (Pouillet's law). The resistance of a wire is given by the following formula:

$$R = \frac{l}{A}\rho, \tag{50.10}$$

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 50.1.14 (Resistor circuits). Consider an electrical circuit of the form



The potential difference over both resistors is equal or, equivalently,

$$R_1I_1 = R_2I_2. \tag{50.11}$$

?? COMPLETE (use correct icon for resistor)??

Theorem 50.1.15 (Kirchhoff's first law). *The directed*[2] *sum of currents meeting at any point is zero.*

Theorem 50.1.16 (Kirchhoff's second law). *The directed sum of potential differences in a closed circuit is zero.*

[2] Currents leaving the point are counted with a minus sign.

50.1.2 Ohm's law

Formula 50.1.17 (Free current density). The current density generated by free charges is given by

$$\vec{\boldsymbol{J}} = nq\vec{\boldsymbol{v}}_d. \tag{50.12}$$

Formula 50.1.18 (Ohm's law).

$$\vec{\boldsymbol{J}} = \sigma \cdot \vec{\boldsymbol{E}}, \tag{50.13}$$

where σ is the conductivity tensor. (Compare this to Definitions 50.1.10 and 50.1.11 for 1D systems.)

Formula 50.1.19 (Ohm's law in wires). The following formula can be found by combining equations 50.1.11, 50.1.12 and 50.1.18 and by assuming that the conductivity tensor is a scalar (this follows from the isotropic behaviour of most resistors):

$$U = RI. \tag{50.14}$$

?? CHECK THIS ARGUMENT ??

50.1.3 Electric dipoles

Formula 50.1.20 (Electric dipole).

$$\vec{\boldsymbol{p}} := q\vec{\boldsymbol{l}} \tag{50.15}$$

where

- q is the positive charge, and
- $\vec{\boldsymbol{l}}$ 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 $\vec{\boldsymbol{l}}$ points in the direction of the x-axis, i.e. $\vec{\boldsymbol{l}} = \lambda\hat{\boldsymbol{e}}_{\boldsymbol{x}}$ for some $\lambda > 0$. The electric potential of the dipole is then given by:

$$V(\vec{\boldsymbol{r}}) = \frac{1}{4\pi\varepsilon_0}\left(\frac{q}{r} - \frac{q}{r + l\cos(\theta)}\right) = \frac{q}{4\pi\varepsilon_0}\frac{l\cos(\theta)}{r(r + l\cos(\theta))}, \tag{50.16}$$

where θ is the angle between the $\hat{\boldsymbol{e}}_{\boldsymbol{x}}$ and $\vec{\boldsymbol{r}} + \vec{\boldsymbol{l}}$ (see Figure 50.1). In the far-field regime, where $l\cos(\theta) \ll r$, the potential is given by

$$V(\vec{\boldsymbol{r}}) \approx \frac{q}{4\pi\varepsilon_0}\frac{l\cos(\theta)}{r^2}. \tag{50.17}$$

Formula 50.1.21 (Energy). If an electric dipole is placed in an external electric field, its potential energy is given by

$$U = -\vec{\boldsymbol{p}} \cdot \vec{\boldsymbol{E}}. \tag{50.18}$$

Formula 50.1.22 (Torque). If an electric dipole is placed in an external electric field, the torque on this system is given by

$$\vec{\boldsymbol{\tau}} = \vec{\boldsymbol{p}} \times \vec{\boldsymbol{E}}. \tag{50.19}$$

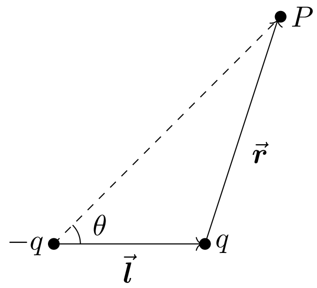


Figure 50.1: Electric dipole configuration.

Formula 50.1.23 (Multipole expansion). Consider a general charge distribution $\rho : \mathbb{R}^3 \to \mathbb{R}$ with support in a region Γ and assume that $\|\vec{\boldsymbol{r}}\| \gg \mathrm{diam}(\Gamma)$. If $\vec{\boldsymbol{r}}$ is oriented along the z-axis, one can write the (squared) distance to an infinitesimal volume element as

$$R^2 = r^2 + (r')^2 - 2rr'\cos(\theta) \tag{50.20}$$

or

$$\frac{1}{R} = \frac{1}{r}\frac{1}{\sqrt{1 - \left(\frac{r'}{r}\right)^2 - 2\left(\frac{r'}{r}\right)\cos(\theta)}}, \tag{50.21}$$

where $\vec{\boldsymbol{r}}'$ is the position of the volume element and θ is the polar angle of the volume element. In the general integral

$$V(\vec{\boldsymbol{r}}) = \frac{1}{4\pi\varepsilon_0}\int \frac{\rho(\vec{\boldsymbol{r}}')}{\|\vec{\boldsymbol{r}} - \vec{\boldsymbol{r}}'\|}d^3V' \tag{50.22}$$

one can now expand the denominator in powers of $1/r$, since the second factor in (50.21) is the generating function of Legendre polynomials:

$$V(\vec{\boldsymbol{r}}) = \frac{1}{4\pi\varepsilon_0}\left(\frac{K_0}{r} + \frac{K_1}{r^2} + \frac{K_2}{r^3} + \cdots\right), \tag{50.23}$$

where:

- $K_0 := \displaystyle\int \rho(\vec{\boldsymbol{r}})dV$
- $K_1 := \displaystyle\int r\cos(\theta)\rho(\vec{\boldsymbol{r}})dV$
- $K_2 := \displaystyle\int \frac{3\cos^2(\theta) - 1}{2}r^2\rho(\vec{\boldsymbol{r}})dV$
- $\cdots$
- $K_n := \displaystyle\int P_n(\cos(\theta))r^n\rho(\vec{\boldsymbol{r}})dV$,

where P_n is the n^{th} *Legendre polynomial.* The first term gives an approximation as if all charge was concentrated in a single point (its 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 $\vec{\boldsymbol{r}}$ is not assumed to be oriented along the z-axis, the Legendre polynomials have to be replaced by *spherical harmonics.*

50.2 Magnetism

Definition 50.2.1 (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 moving at one meter per second experiences a force of one newton when the magnetic induction equals one tesla: $\mathrm{T} = \frac{\mathrm{N}}{\mathrm{Am}}$.

The **magnetizing field** $\vec{\boldsymbol{H}}$ 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**:

$$\vec{\boldsymbol{M}} := \chi\vec{\boldsymbol{H}}, \tag{50.24}$$

where χ is the **magnetic susceptibility**.

The **magnetic induction** $\vec{\boldsymbol{B}}$ is the field generated by both the external sources and the internal magnetization. It is only this field that can be measured. In vacuum, the following relation between the magnetic induction, the magnetizing field and the magnetization exists:

$$\vec{\boldsymbol{B}} = \mu_0\left(\vec{\boldsymbol{H}} + \vec{\boldsymbol{M}}\right). \tag{50.25}$$

By combining the previous two formulas one obtains (this equation is only valid in linear media)

$$\vec{\boldsymbol{B}} = \mu_0(1+\chi)\vec{\boldsymbol{H}}. \tag{50.26}$$

The proportionality constant in this formula is called the **magnetic permeability**:

$$\mu := \mu_0(1+\chi), \tag{50.27}$$

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 50.2.2 (Tensorial formulation). In anisotropic materials one has to use a tensorial formulation:

$$B_i = \sum_j \mu_{ij}H_j, \tag{50.28}$$

$$M_i = \sum_j \chi_{ij}H_j. \tag{50.29}$$

In this case both μ and χ are $(1,1)$-tensors, i.e. linear maps.

50.2.1 Electric charges in a magnetic field

Formula 50.2.3 (Gyroradius). A charge q moving in a magnetic field B with velocity $\vec{\boldsymbol{v}}$ will follow a circular path with radius

$$r = \frac{mv_\perp}{|q|B}. \tag{50.30}$$

Formula 50.2.4 (Gyrofrequency[3]). The circular velocity is given by

$$\omega = \frac{|q|B}{m}. \tag{50.31}$$

[3] Also called the **Larmor** or **cyclotron frequency**.

50.3 Differential Maxwell equations

Formula 50.3.1 (Gauss's law for electricity).

$$\nabla \cdot \vec{\boldsymbol{E}} = \frac{\rho}{\varepsilon} \tag{50.32}$$

Formula 50.3.2 (Gauss's law for magnetism).

$$\nabla \cdot \vec{\boldsymbol{B}} = 0 \tag{50.33}$$

If magnetic monopoles would exist, the right-hand side would contain a (magnetic) charge density as in the case of electricity.

Formula 50.3.3 (Faraday's law).

$$\nabla \times \vec{\boldsymbol{E}} = -\frac{\partial \vec{\boldsymbol{B}}}{\partial t} \tag{50.34}$$

Formula 50.3.4 (Maxwell's law[4]).

$$\nabla \times \vec{\boldsymbol{B}} = \varepsilon\mu\frac{\partial \vec{\boldsymbol{E}}}{\partial t} + \mu\vec{\boldsymbol{J}} \tag{50.35}$$

50.4 Potentials

50.4.1 Decomposition in potentials

The Helmholtz decomposition 21.1.15 together with Gauss's law 50.3.2 imply the following general form for $\vec{\boldsymbol{B}}$:

$$\vec{\boldsymbol{B}} = \nabla \times \vec{\boldsymbol{A}}, \tag{50.36}$$

where $\vec{\boldsymbol{A}}$ is called the **magnetic potential**.

Combining Equation (50.36) with Faraday's law 50.3.3 (and doing some rewriting) gives the following general form for $\vec{\boldsymbol{E}}$:

$$\vec{\boldsymbol{E}} = -\nabla V - \frac{\partial \vec{\boldsymbol{A}}}{\partial t}, \tag{50.37}$$

where V is called the **electrostatic potential**.

Property 50.4.1. Substituting the previous two expressions into 50.3.1 and 50.3.4 gives the following two (coupled) conditions for the electromagnetic potentials:

$$\Delta\vec{\boldsymbol{A}} - \varepsilon\mu\frac{\partial^2 \vec{\boldsymbol{A}}}{\partial t^2} = \nabla\left(\nabla \cdot \vec{\boldsymbol{A}} + \varepsilon\mu\frac{\partial V}{\partial t}\right) - \mu\vec{\boldsymbol{J}} \tag{50.38}$$

$$\Delta V - \varepsilon\mu\frac{\partial^2 V}{\partial t^2} = -\frac{\partial}{\partial t}\left(\nabla \cdot \vec{\boldsymbol{A}} + \varepsilon\mu\frac{\partial V}{\partial t}\right) - \frac{\rho}{\varepsilon}. \tag{50.39}$$

[4] Also called the **Maxwell-Ampère law**.

50.5 Lorentz force

The relation between the force acting on a charged particle and the strength of electric and magnetic fields is given by the Lorentz force law:

Formula 50.5.1 (Lorentz force).

$$\vec{\boldsymbol{F}} = q\left(\vec{\boldsymbol{E}} + \vec{\boldsymbol{v}} \times \vec{\boldsymbol{B}}\right) \tag{50.40}$$

The charge of a particle acts as the constant of proportionality.

Formula 50.5.2 (Lorentz force density).

$$\vec{\boldsymbol{f}} = \rho\vec{\boldsymbol{E}} + \vec{\boldsymbol{J}} \times \vec{\boldsymbol{B}} \tag{50.41}$$

50.5.1 Gauge transformations

Looking at Equation (50.36), it is clear that a transformation $\vec{\boldsymbol{A}} \longrightarrow \vec{\boldsymbol{A}} + \nabla\psi$ has no effect on $\vec{\boldsymbol{B}}$ due to Equation (21.13). To compensate for this transformation in Equation (50.37), one also has to perform a transformation $V \longrightarrow V - \frac{\partial\psi}{\partial t}$. These transformations are called **gauge transformations**.

Definition 50.5.3 (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 50.5.4. 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 50.5.5 (Lorenz gauge). A first example of a gauge fixing condition is the Lorenz gauge[5]:

$$\nabla \cdot \vec{\boldsymbol{A}} + \varepsilon\mu\frac{\partial V}{\partial t} = 0. \tag{50.42}$$

When using this gauge fixing condition, Equations (50.38) and (50.39) uncouple. This allows to rewrite them as

$$\Box\vec{\boldsymbol{A}} = -\mu\vec{\boldsymbol{J}} \tag{50.43}$$

$$\Box V = -\frac{\rho}{\varepsilon}. \tag{50.44}$$

To see which gauge functions ψ are valid in this case, perform a transformation as explained above:

$$\vec{\boldsymbol{A}}' = \vec{\boldsymbol{A}} + \nabla\psi \qquad \text{and} \qquad V' = V - \frac{\partial\psi}{\partial t}. \tag{50.45}$$

Substituting these transformations in Equation (50.42) and using the fact that both sets of potentials $(\vec{\boldsymbol{A}}, V)$ and $(\vec{\boldsymbol{A}}', V')$ satisfiy the Lorenz gauge (50.42) gives the following condition for the gauge function ψ:

$$\Box\psi = 0. \tag{50.46}$$

Example 50.5.6 (Coulomb gauge). Apart from the Lorenz gauge there is also the Coulomb gauge:

$$\nabla \cdot \vec{\boldsymbol{A}} = 0. \tag{50.47}$$

[5]Named after *Ludvig Lorenz*. Not to be confused with *Hendrik Lorentz*.

50.6 Energy and momentum

Definition 50.6.1 (Poynting vector).

$$\vec{\boldsymbol{S}} := \vec{\boldsymbol{E}} \times \vec{\boldsymbol{H}} \tag{50.48}$$

Formula 50.6.2 (Energy density).

$$W = \frac{1}{2}\left(\vec{\boldsymbol{E}} \cdot \vec{\boldsymbol{D}} + \vec{\boldsymbol{B}} \cdot \vec{\boldsymbol{H}}\right) \tag{50.49}$$

50.7 Differential-geometric perspective

Using the tools introduced in Chapter 32 (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 [60] for a complete derivation and interpretation. (It should be noted that *Gaussian units* are used throughout this section.)

Definition 50.7.1 (Field strength). Define

$$\mathbf{E} = E_1dx^1 + E_2dx^2 + E_3dx^3$$

and

$$\mathbf{B} = B_1dx^2 \wedge dx^3 + B_2dx^3 \wedge dx^1 + B_3dx^1 \wedge dx^2$$

as the electric and magnetic field forms. Using these differential forms one can write the field strength as follows:

$$\mathbf{F} = \mathbf{B} - dt \wedge \mathbf{E}. \tag{50.50}$$

Formula 50.7.2 (Maxwell's equations). Define the electric 4-current as

$$\mathbf{J} = \rho dt - J_1dx^1 - J_2dx^2 - J_3dx^3.$$

Maxwell's equations can now be rewritten as follows:

$$d\mathbf{F} = 0 \tag{50.51}$$

$$*d(*\mathbf{F}) = 4\pi\mathbf{J}, \tag{50.52}$$

where $*$ is the Hodge operator (21.61).

Definition 50.7.3 (Potential). The homogeneous equation (50.51) and Poincaré's lemma 32.8.8 imply that there exists a differential one-form $\mathbf{A}$ such that

$$\mathbf{F} = d\mathbf{A}. \tag{50.53}$$

This one-form is called the potential or **gauge field**. It can be related to the ordinary scalar potential V and vector potential $\vec{\boldsymbol{A}}$ as follows:

$$\mathbf{A} = -V dt + A_1dx^1 + A_2dx^2 + A_3dx^3. \tag{50.54}$$

Property 50.7.4 (Gauge transformation). Because $d^2 \equiv 0$, the above equation is invariant under a transformation $\mathbf{A} \longrightarrow \mathbf{A} + df$ for any $f \in C^\infty(\mathbb{R}^3)$. When written out in coordinates, this gives exactly the gauge transformations from Equation (50.45).

?? COMPLETE ??

Chapter 51

Optics

51.1 General

51.1.1 Conservation of energy

From the law of conservation of energy one can derive the following formula:

$$T + R + A = 1, \tag{51.1}$$

where

- T is the transmission coefficient,
- R is the reflection coefficient, and
- A is the absorption coefficient.

51.2 Wave mechanics

Formula 51.2.1 (Wave equation). In one spatial dimension the wave equation reads as follows:

$$\frac{\partial^2 f}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 f}{\partial t^2}. \tag{51.2}$$

In higher dimensions this can be rewritten using the Laplacian 21.10 as

$$\Delta f = \frac{1}{v^2}\frac{\partial^2 f}{\partial t^2}. \tag{51.3}$$

or using the **d'Alembertian**

$$\Box_v := \frac{1}{v^2}\frac{\partial^2}{\partial t^2} - \Delta \tag{51.4}$$

as

$$\Box_v f = 0. \tag{51.5}$$

Formula 51.2.2 (d'Alemberts formula). Consider the wave equation $\Box f = 0$. By applying the method from Section 19.3 it is clear that the characteristics are given by

$$\xi = x + ct \qquad \text{and} \qquad \eta = x - ct. \tag{51.6}$$

Furthermore, it follows that the wave equation is a hyperbolic equation and, accordingly, can be rewritten in the canonical form

$$\frac{\partial^2 u}{\partial\xi\partial\eta}(\xi,\eta) = 0 \tag{51.7}$$

Integration with respect to ξ and η and rewriting the solution in terms of x and t gives

$$u(x,t) = f(x+ct) + g(x-ct), \tag{51.8}$$

where f, g are arbitrary functions. This solution represents a superposition of a left-moving wave and a right-moving wave.

Now, consider the wave equation subject to the general boundary conditions

$$u(x,0) = v(x) \qquad \text{and} \qquad \frac{\partial u}{\partial t}(x,0) = q(x). \tag{51.9}$$

By inserting these conditions in the solution (51.8), it can be shown that the general solution, subject to the given boundary conditions, is given by

$$u(x,t) := \frac{1}{2}\big(v(x+ct) + v(x-ct)\big) + \frac{1}{2c}\int_{x-ct}^{x+ct} q(z)dz. \tag{51.10}$$

Definition 51.2.3 (Harmonic wave). The wave form obtained by taking

$$v(x \pm vt) = A\sin\left(\frac{2\pi}{\lambda}(x \pm vt)\right)$$

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:

$$\begin{aligned}
\xi(x,t) &= A\sin\left(\frac{2\pi}{\lambda}(x \pm vt)\right)\\
&= A\sin\left(2\pi(\frac{x}{\lambda} \pm \nu t)\right)\\
&= A\sin\left(2\pi(\frac{x}{\lambda} \pm \frac{t}{\tau})\right)\\
&= A\sin(kx \pm \omega t).
\end{aligned}$$

The constants 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:

$$v := \lambda\nu. \tag{51.11}$$

One can also introduce an additional constant such the wave form is shifted in space/time:

$$\xi(x,t) = A\sin\left(\frac{2\pi}{\lambda}(x \pm vt) + \varphi\right). \tag{51.12}$$

This constant φ is called a **phase** (shift).

Definition 51.2.4 (Group velocity). Consider a superposition of two harmonic (cosine) 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:

$$v_g := \frac{d\omega}{dk}. \tag{51.13}$$

This is the definition of the group velocity of an arbitrary wave packet

$$u(x,t) = \int_{-\infty}^{\infty} A(k)e^{i(kx-\omega t)}dk.$$

Method 51.2.5 (Huygens's 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 these secondary waves.

51.3 Electromagnetic radiation

Consider Maxwell's equations from Section 50.3. By inserting Maxwell's law in the rotor of Faraday's law, the following equation for the electric field is obtained:

$$\Delta\vec{\boldsymbol{E}} = \varepsilon_0\mu_0\frac{\partial^2\vec{\boldsymbol{E}}}{\partial t^2}. \tag{51.14}$$

This has the form of the wave equation (51.2) with speed $c^2 = \frac{1}{\varepsilon_0\mu_0}$. A similar derivation gives rise to a wave equation of the magnetic field. Moreover, by the Maxwell equations, the solutions of these wave equations are coupled. For example, if one considers harmonic solutions:

$$\begin{aligned}\vec{\boldsymbol{E}}(x,t) &= \vec{\boldsymbol{E}}_0\sin(kx-\omega t)\\ \vec{\boldsymbol{B}}(x,t) &= \vec{\boldsymbol{B}}_0\sin(kx-\omega t),\end{aligned}$$

Faraday's law gives

$$\vec{\boldsymbol{E}}_0 k\cos(kx-\omega t) = \vec{\boldsymbol{B}}_0\omega\cos(kx-\omega t). \tag{51.15}$$

This in turns gives rise to the following important identity:

$$\frac{\|\vec{\boldsymbol{E}}_0\|}{\|\vec{\boldsymbol{B}}_0\|} = \frac{\omega}{k} = c. \tag{51.16}$$

Formula 51.3.1 (Energy).

$$E = h\nu = \hbar\omega = \frac{hc}{\lambda} \tag{51.17}$$

Formula 51.3.2 (Momentum).

$$p = \frac{h}{\lambda} = \hbar k \tag{51.18}$$

Definition 51.3.3 (Polarization). The polarization of an electromagnetic wave is defined as the direction of the $\vec{\boldsymbol{E}}$-component.

Consider a superposition of two orthogonally polarized plane waves:

$$\begin{aligned}\vec{\boldsymbol{E}}_1(x,t) &= E_1 \sin(kx - \omega t)\hat{\boldsymbol{e}}_{\boldsymbol{y}}\\ \vec{\boldsymbol{E}}_2(x,t) &= E_2 \sin(kx - \omega t + \varphi)\hat{\boldsymbol{e}}_{\boldsymbol{z}}.\end{aligned}$$

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.

51.4 Refraction

Formula 51.4.1 (Refractive index). The relative refractive index between two materials is defined as follows:

$$n := \frac{v_1}{v_2}, \tag{51.19}$$

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** is obtained.

Formula 51.4.2 (Dielectric function). In the case of nonmagnetic materials ($\mu_r \approx 1$), one can write the dielectric function as follows:

$$\epsilon = \epsilon_r + i\epsilon_i =: \widetilde{n}^2 = (n + ik)^2, \tag{51.20}$$

where $\widetilde{n}$ is the **complex refractive index** and k is the **extinction coefficient**.

Theorem 51.4.3 (Malus-Dupin). *Light rays are always perpendicular to the wave fronts.*

Property 51.4.4 (Reflection and refraction laws). For plane waves reflection and refraction are characterized by the following three rules:

1. The normal 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**:

$$\frac{\sin(\theta_i)}{\sin(\theta_r)} = \frac{n_2}{n_1}. \tag{51.21}$$

The second and third item can be derived from both Huygens's principle and Malus's law or Fermat's principle.

51.5 Absorption

Theorem 51.5.1 (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*

$$I(x) = I(0)\exp\left(-\frac{4\pi\nu k}{c}x\right). \tag{51.22}$$

Proof. From Equation (51.20) it is known that the complex refractive index can be written as

$$\widetilde{n} = n + ik,$$

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:

$$c = \widetilde{n}v.$$

It readily follows that the wave number (sadly also denoted by the letter k) can be written as

$$k = \frac{\omega}{v} = \widetilde{n}\frac{\omega}{c}.$$

From electromagnetism one knows that a plane wave can be written as

$$E(x,t) = \mathrm{Re}\big\{A\exp[i(kx - \omega t + \phi)]\big\}.$$

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

$$E(x,t) = \mathrm{Re}\left\{A\exp\left(in\frac{\omega}{c}x\right)\cdot\exp\left(-k\frac{\omega}{c}x\right)\cdot\exp(-i\omega t)\cdot\exp(i\phi)\right\}.$$

It is also known that the intensity is given by the following relation:

$$I(x) = |E(x)|^2 = E^*(x)\cdot E(x).$$

This implies that only the second exponential factor will remain. Dividing the result by its value at $x = 0$ gives

$$\frac{I(x)}{I(0)} = \frac{E(x)\cdot E^*(x)}{E(0)\cdot E^*(0)} = \exp\left(-\frac{2k\omega}{c}x\right) = \exp(-\alpha x),$$

where α is the absorption coefficient as defined in Definition 51.5.2. □

Definition 51.5.2 (Absorption coefficient). The scale factor in the Lambert-Beer law is called the absorption coefficient:

$$\alpha = \frac{4\pi\nu k}{c}. \tag{51.23}$$

Definition 51.5.3 (Scattering). When scattering light off matter, two situations can occur:

- **Rayleigh scattering**: Photons are elastically scattered off the bounded electrons. The incoming and outgoing frequencies are equal.
- **Raman scattering**: Photons are inelastically scattered.

Chapter 52

Astronomy

52.1 Ellipsoidal coordinates

Consider the following equation:

$$f(\tau) = \frac{x^2}{\tau + \alpha} + \frac{y^2}{\tau + \beta} + \frac{z^2}{\tau + \gamma} - 1, \tag{52.1}$$

where $\alpha < \beta < \gamma < 0$. By multiplying by the denominators and choosing $f(\tau) = 0$, a polynomial equation of degree 3 in τ is obtained. This polynomial can be formally factorized as

$$-(\tau - \lambda)(\tau - \mu)(\tau - \nu) = 0. \tag{52.2}$$

This solutions of this equation obey the following conditions:

- $\nu \in \,]-\gamma, -\beta[\,$,
- $\mu \in \,]-\beta, -\alpha[\,$, and
- $\lambda \in \,]-\alpha, \infty[$.

From the previous two equations one can find a solution for x^2 by multiplying by $(\tau + \alpha)$ and taking the limit $\tau \to -\alpha$. Solutions for y^2 and z^2 can be found in a similar way:

$$\begin{cases} x^2 = \frac{(\lambda+\alpha)(\mu+\alpha)(\nu+\alpha)}{(\beta-\alpha)(\gamma-\alpha)} \\ y^2 = \frac{(\lambda+\beta)(\mu+\beta)(\nu+\beta)}{(\beta-\alpha)(\beta-\gamma)} \\ z^2 = \frac{(\lambda+\gamma)(\mu+\gamma)(\nu+\gamma)}{(\alpha-\gamma)(\beta-\gamma)}. \end{cases} \tag{52.3}$$

These solutions can be divided in different families depending on the value of τ.

52.1.1 Ellipsoid: $\tau = \lambda$

The first family consists of the surfaces defined by fixing $\tau = \lambda$ in Equation (52.1). By noting that all denominators are positive in this case, it can be seen that the obtained surface is an ellipsoid with the x-axis as the shortest axis. By taking the limit $\lambda \longrightarrow \infty$ the equation of a sphere with radius $\sqrt{\lambda}$ is obtained, whilst taking $\lambda \longrightarrow -\alpha$ results in an ellipse in the yz-plane. This ellipse is called the **focal ellipse**.

52.1.2 One-sheet hyperboloid: $\tau = \mu$

By fixing $\tau = \mu$ in (52.1) the equation of a one-sheet hyperboloid (also called a **hyperbolic hyperboloid**) around the x-axis is obtained. By taking the limit $\mu \longrightarrow -\alpha$ the hyperboloid collapses on the yz-plane and the surface outside the focal ellipse is obtained. If one takes $\mu \longrightarrow -\beta$, the hyperboloid becomes degenerate and one gets the surface inside the **focal hyperbola** defined by

$$\frac{x^2}{\alpha-\beta} + \frac{z^2}{\gamma-\beta} = 1. \tag{52.4}$$

This hyperbola intersects the z-plane in the foci of the focal ellipse.

52.1.3 Two-sheet hyperboloid: $\tau = \nu$

By fixing $\tau = \nu$ in (52.1) the equation of a two-sheet hyperboloid (also called an **elliptic hyperboloid**) around the z-axis is obtained. By taking the limit $\nu \longrightarrow -\beta$ the hyperboloid becomes degenerate and one obtains the surface outside the focal hyperbola (52.4). If $\nu \longrightarrow -\gamma$ the two sheets coincide in the xy-plane.

52.1.4 Hamiltonian function

When writing out the kinetic energy in ellipsoidal coordinates and noting that mixed terms of the form $\dfrac{\partial x^a}{\partial\lambda^i}\dfrac{\partial x^a}{\partial\lambda^j}$ cancel out due to (52.3), it is clear that the Hamiltonian function can be separated:

$$H = \frac{1}{2}\left(\frac{p_\lambda^2}{Q_\lambda^2} + \frac{p_\mu^2}{Q_\mu^2} + \frac{p_\nu^2}{Q_\nu^2}\right) + V, \tag{52.5}$$

where $Q_j^2 = \sum_i \left(\dfrac{\partial x^i}{\partial\lambda^j}\right)^2$ are the metric coefficients in ellipsoidal coordinates. After a straightforward calculation these can be found to be:

$$Q_\lambda^2 = \frac{1}{4}\frac{(\lambda-\mu)(\lambda-\nu)}{(\lambda+\alpha)(\lambda+\beta)(\lambda+\gamma)}, \tag{52.6}$$

which is also valid for μ and ν after cyclically permutating the coordinates.

Because of the Stäckel conditions (48.75), the potential must be of the form

$$V = \sum_i \frac{W_i(\lambda^i)}{Q_i^2} \tag{52.7}$$

if one wants to obtain a separable Hamilton-Jacobi equation. Due to the disjoint nature of λ, μ and ν one can consider W_λ, W_μ and W_ν as three components of a single function:

$$G(\tau) := -4(\tau+\beta)W_\tau(\tau). \tag{52.8}$$

The 3D potential is thus completely determined by a univariate function $G(\tau)$.

52.1.5 Hamilton-Jacobi equation

If a time-independent system is considered, one can use the Hamilton-Jacobi equation (48.73) as the starting point. By multiplying this equation by $(\lambda-\mu)(\lambda-\nu)(\mu-\nu)$ one obtains

$$\begin{aligned}(\mu-\nu)\Bigg[2(\lambda+\alpha)(\lambda+\beta)(\lambda+\gamma)\left(\frac{dS^\lambda(\lambda)^2}{d\lambda}\right)& \\ -(\lambda+\alpha)(\lambda+\gamma)G(\lambda) - \lambda^2 E\Bigg]& + \text{cyclic permutations} = 0,\end{aligned} \tag{52.9}$$

where the multiplicative factor was rewritten in the form $a\lambda^2 + b\mu^2 + c\nu^2$ before multiplying the right-hand side of (48.73). This equation can also be rewritten as

$$(\mu - \nu)U(\lambda) + (\lambda - \mu)U(\nu) + (\nu - \lambda)U(\mu) = 0. \tag{52.10}$$

Differentiating twice with respect to any λ^i gives $U''(\tau) = 0$ or, equivalently,

$$U(\tau) = I_3 - I_2\tau, \tag{52.11}$$

where I_2 and I_3 are two new integrals of motion.

Using the Hamiltonian-Jacobi equation (48.72) one can obtain the conjugate momenta. After a lengthy calculation one obtains

$$p_\tau^2 = \left(\frac{dS^\tau}{d\tau}\right)^2 = \frac{1}{2(\tau+\beta)}[E - V_{\text{eff}}(\tau)], \tag{52.12}$$

where the effective potential is given by

$$V_{\text{eff}} = \frac{J}{\tau+\alpha} + \frac{K}{\tau+\gamma} - G(\tau). \tag{52.13}$$

The two conserved quantities J and K are given by

$$J = \frac{\alpha^2 E + \alpha I_2 + I_3}{\alpha - \gamma} \qquad \text{and} \qquad K = \frac{\gamma^2 E + \gamma I_2 + I_3}{\gamma - \alpha}.$$

To be physically acceptable, p_τ^2 should be positive. This leads to following conditions on the energy:

$$\begin{cases} E & \geq V_{\text{eff}}(\lambda) \\ E & \geq V_{\text{eff}}(\mu) \\ E & \leq V_{\text{eff}}(\nu). \end{cases} \tag{52.14}$$

The generating function $G(\tau)$ should also satisfy some conditions. Note that the Stäckel potential $V(\lambda, \mu, \nu)$ can be rewritten as

$$V = -\frac{1}{\lambda - \nu}\left(\frac{F(\lambda) - F(\mu)}{\lambda - \mu} - \frac{F(\mu) - F(\nu)}{\mu - \nu}\right) \leq 0, \tag{52.15}$$

where $F(\tau) = (\tau+\alpha)(\tau+\gamma)G(\tau)$. For $\lambda \longrightarrow \infty$ (or $r^2 \longrightarrow \infty$) one obtains $V \approx -\frac{F(\lambda)}{\lambda^2} \approx -G(\lambda)$. Because $V \sim \lambda^{-1}$ it is clear that $G(\tau)$ cannot decay faster than $\lambda^{-1/2}$ at infity. Furthermore, one can interpret (52.15) as an approximation of $-F''(\tau)$. It follows that $F(\tau)$ should be convex. For $\tau \longrightarrow -\gamma$ one obtains

$$\begin{cases} \alpha + \tau < 0 \\ \tau + \gamma \longrightarrow 0. \end{cases}$$

So, if $G(\tau)$ decays faster than $\frac{1}{\tau+\gamma}$, then $F(\tau) \longrightarrow -\infty$, which is not possible for a convex function.

To fulfil these conditions assume that the generating function can be written as

$$G(\tau) = \frac{GM}{\sqrt{\gamma_0 + \tau}}, \tag{52.16}$$

where G is the gravitational constant and M is the galactic mass.

Theorem 52.1.1 (Kuzmin). *The spatial mass density function generated by a Stäckel potential is completely determined by a function of the form $\rho(z)$.*

Corollary 52.1.2. For triaxial mass models in ellipsoidal coordinates the axial ratios are inversely proportional to the axial ratios of the coordinate system.

Part VIII

Quantum Mechanics

Chapter 53

Wave and Matrix Mechanics

The main reference for this chapter is [55]. In this chapter the two basic formalisms of Quantum Mechanics are introduced: wave and matrix mechanics.

53.1 Schrödinger picture

Formula 53.1.1 (Time-independent Schrödinger equation).

$$\hat{H}\psi(x) = E\psi(x). \tag{53.1}$$

The operator $\hat{H}$ is called the **Hamiltonian** of the system and 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.

Property 53.1.2 (Orthogonality). Let $\{\psi_i\}_{i\in I} \subset L^2(\mathbb{R})$ be a collection of eigenfunctions of the TISE (the internal space is hidden for convenience). These functions can be normalized with respect to the inner product (16.36) such that they obey the following relation:

$$\int_{\mathbb{R}} \overline{\psi_i}(x)\psi_j(x)dx = \delta_{ij}. \tag{53.2}$$

The time evolution of a wave function is defined through the following equation:

Formula 53.1.3 (TDSE). The following partial differential equation is called the **(time-dependent) Schrödinger equation**:

$$i\hbar\frac{\partial}{\partial t}\psi(x,t) = \hat{H}\psi(x,t). \tag{53.3}$$

In case $\hat{H}$ is time-independent, the TISE can be obtained from this equation by separation of variables (see below).

Example 53.1.4 (Massive particle in a stationary potential).

$$i\hbar\frac{\partial}{\partial t}\psi(x,t) = \left(\frac{\hat{p}^2}{2m} + \hat{V}(x)\right)\psi(x,t) \tag{53.4}$$

Derivation of TISE from TDSE. Starting from the one-dimensional TDSE in position space with a time-independent potential

$$i\hbar\frac{\partial}{\partial t}\Psi(x,t) = \left(-\frac{\hbar^2}{2m}{\partial_x}^2 + V(x)\right)\Psi(x,t), \tag{53.5}$$

one can perform a separation of variables and assert a solution of the form $\Psi(x,t) = \psi(x)T(t)$. Inserting this in the previous equation gives

$$i\hbar\psi(x)T'(t) = -\frac{\hbar^2}{2m}T(t)\psi''(x) + V(x)\psi(x)T(t). \tag{53.6}$$

Dividing both sides by $X(x)T(t)$ and rearranging the terms gives

$$i\hbar\frac{T'(t)}{T(t)} = \left(-\frac{\hbar^2}{2m} + V(x)\right)\frac{\psi''(x)}{\psi(x)}. \tag{53.7}$$

Because the left side only depends on t and the right side only depends on x, one can conclude that they are both equal to a constant E. This leads to the following system of differential equations:

$$\begin{cases} i\hbar T'(t) - ET(t) = 0, \\ \left(-\frac{\hbar^2}{2m} + V(x)\right)\psi''(x) = E\psi(x). \end{cases} \tag{53.8}$$

The first equation immediately gives a solution for T:

$$T(t) = Ce^{-\frac{i}{\hbar}Et}. \tag{53.9}$$

Rearranging the second solution in the system gives the TISE:

$$\psi''(x) = -\frac{2m}{\hbar^2}(E - V(x))\psi(x). \tag{53.10}$$

□

Formula 53.1.5 (General solution). A general solution of the time-dependent Schrödinger equation (for time-independent Hamiltonians) is given by the following formula (cf. Formula 18.3.2):

$$\psi(x,t) = \sum_E c_E\psi_E(x)e^{-\frac{i}{\hbar}Et}, \tag{53.11}$$

where the functions $\psi_E(x)$ are the eigenfunctions of the TISE 53.1.1. The coefficients c_E can be found using the orthogonality relations

$$c_E = \left(\int_{\mathbb{R}} \overline{\psi_E}(x)\psi(x,t_0)dx\right)e^{\frac{i}{\hbar}Et_0}. \tag{53.12}$$

?? COMPLETE ??

53.1.1 Hydrogen atom

Consider the hydrigen 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):

$$\hat{H} := \frac{\hat{p}_p^2}{2m_p} + \frac{\hat{p}_e^2}{2m_e} - \frac{e^2}{4\pi\varepsilon r^2}. \tag{53.13}$$

It is not hard to see that this is the quantum mechanical version of the Kepler problem (Section 48.2.3). The special property of the Kepler problem was that it contained a “hidden” symmetry that gave rise to the conserved Laplace-Runge-Lenz vector 48.2.12. 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 follows from rotational symmetry, but the energy levels of the hydrogen atom only depend on the principal quantum number n. The degeneracy of the total angular quantum number l is that due to the "hidden" SO(4)-symmetry. It is often called an "accidental degeneracy" for this reason.

?? COMPLETE ??

53.1.2 Molecular dynamics

Consider the Hamiltonian of two interacting atoms:

$$\hat{H} := \frac{\hat{P}_1^2}{2M_1} + \frac{\hat{P}_2^2}{2M_2} + \frac{q_1q_2}{4\pi\varepsilon R^2} + \sum_i \frac{\hat{p}_i^2}{2m} - \frac{eq_1}{4\pi\varepsilon r_{i1}^2} - \frac{eq_2}{4\pi\varepsilon r_{i2}^2} + \sum_{i\neq j} \frac{e^2}{4\pi\varepsilon r_{ij}^2}, \tag{53.14}$$

where the indices i, j indicate the electrons and capital 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 mass 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:

$$\hat{H}_{\text{nucl}} = \frac{\hat{P}_1^2}{2M_1} + \frac{\hat{P}_2^2}{2M_2} + \frac{Q_1Q_2}{4\pi\varepsilon R^2} + V_{\text{eff}}(R_1, R_2). \tag{53.15}$$

The electrons generate an effective potential for the nuclei and the Schrödinger equation decouples as follows:

$$\hat{H}_{\text{nucl}}(R)\psi(R) = E\psi(R) \tag{53.16}$$

$$\hat{H}_{\text{el}}(r, R)\phi(r, R) = E_{\text{el}}\phi(r, R). \tag{53.17}$$

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.

53.2 Heisenberg picture

This is the right place to reflect on what the wave function is and how it relates to the state of a system. At every point it gives the probability of observing a particle (or whatever object is being studied). But what if one wants to express this information in terms of momenta instead of positions? The information about the state should not depend on the chosen "representation". To this end, a state vector $|\psi\rangle$ that represents the state of the system as an abstract vector in some Hilbert space is introduced.

Notation 53.2.1 (Dirac notation). This notation is often called the **braket 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 $\langle\phi|\psi\rangle$.

But then, 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 on the plane waves e^{ipx}.

Remark 53.2.2. 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 $L^2(\mathbb{R},\mathbb{C})$. In the next chapter this issue will be resolved through the concept of *rigged Hilbert spaces*.

Formula 53.2.3 (Matrix representation). The following formula gives the matrix representation of an operator $\hat{A}$ with respect to the orthonormal basis $\{|\psi_i\rangle\}_{i\leq n}$ cf. Construction 20.4.18:

$$A_{ij} := \langle\psi_i|\hat{A}|\psi_j\rangle. \tag{53.18}$$

Remark 53.2.4. The basis $\big\{|\psi_i\rangle\big\}_{i\leq n}$ need not consist out of eigenfunctions of $\hat{A}$.

?? COMPLETE ??

53.3 Uncertainty principle

Definition 53.3.1 (Compatible observables). Two observables are said to be compatible if they share a complete set of eigenvectors.

Definition 53.3.2 (Expectation value). The expectation value of an operator $\hat{A}$ in a state $|\psi\rangle$ is defined as

$$\langle\hat{A}\rangle_\psi := \langle\psi|\hat{A}|\psi\rangle. \tag{53.19}$$

The subscript ψ is often left implicit. As in ordinary statistics (45.18), the uncertainty or variance is defined as follows:

$$\Delta A := \langle\hat{A}^2\rangle - \langle\hat{A}\rangle^2. \tag{53.20}$$

Formula 53.3.3 (Uncertainty relation). Let $\hat{A},\hat{B}$ be two operator and let $\Delta A,\Delta B$ be the corresponding uncertainties. The (**Robertson**) uncertainty relation reads as follows:

$$\Delta A\Delta B \geq \frac{1}{4}\left|\left\langle\left[\hat{A},\hat{B}\right]\right\rangle\right|^2. \tag{53.21}$$

Chapter 54

Mathematical Formalism

The main reference for a mathematically rigorous treatment of quantum mechanics, in particular in the infinite-dimensional setting, is [107]. The main reference for the generalization to curved backgrounds [114]. Relevant chapters in this compendium are 17, 23 and 24.

54.1 Postulates

Axiom 54.1 (State spaces). The states of a (closed) system are represented by vectors in a (complex) Hilbert space. In the infinite-dimensional setting one often further restricts to separable spaces, i.e. the spaces are required to admit a countable Hilbert basis.

Axiom 54.2 (Observables). A self-adjoint operator. In the finite-dimensional case this is equivalent to an operator that admits a complete set of eigenfunctions.

Axiom 54.3 (Rays). The dynamics of the system do not depend on the global phase or normalization, states are represented by rays in a projective Hilbert space.

54.1.1 Observables

Formula 54.1.1. Let $|\Psi\rangle$ be a state vector representing a given system and let $\{|\psi_i\rangle\}_{i\in I}$ be a complete set of eigenvectors of some observable of the system. The state vector $|\Psi\rangle$ can be expressed as a linear combination of the eigenfunctions:

$$|\Psi\rangle = \sum_i c_i|\psi_i\rangle + \int c_a|\psi_a\rangle da, \tag{54.1}$$

where the summation ranges over the discrete spectrum and the integral over the continuous spectrum (Section 23.4.5). Note that this expression only makes sense formally, since linear combinations only consist of a finite number of terms, i.e. c_a should be a finite sum of delta functionals.

Formula 54.1.2 (Closure relation). For a complete set of discrete eigenvectors the closure relation (also called the **resolution of the identity**) is given by

$$\sum_n |\psi_n\rangle\langle\psi_n| = \mathbb{1}. \tag{54.2}$$

For a complete set of continuous eigenvectors the following counterpart holds:

$$\int |x\rangle\langle x|dx = \mathbb{1}. \tag{54.3}$$

For a mixed set of eigenvectors a similar relation is obtained by summing over the discrete part and integrating over the continuous part. For simplicity the notation of Equation (54.2) will also be used for the continuous part.

Definition 54.1.3 (Canonical commutation relations). Two observables A, B are said to obey a canonical commutation relation (CCR) if they satisfy (up to a constant factor $\hbar$)

$$[A, B] = i. \tag{54.4}$$

The prime examples are the position and momentum operators $\hat{x}, \hat{p}$. Through functional calculus one can also define the exponential operators e^{isA} and e^{itB}. The above relation then induces the so-called **Weyl form** of the CCR:

$$e^{isA}e^{itB} = e^{ist}e^{itB}e^{isA}. \tag{54.5}$$

Theorem 54.1.4 (Stone-von Neumann). *All pairs of irreducible, unitary, one-parameter subgroups satisfying the Weyl form of the CCRs are unitarily equivalent.*

Corollary 54.1.5. The Schrödinger and Heisenberg pictures are unitarily equivalent.

In fact one can generalize the Weyl relation:

Definition 54.1.6 (Weyl system). Let (A, ω) be a symplectic vector space and let $\mathcal{H}$ be a Hilbert space equipped with a continuous map $W : A \to \mathrm{U}(\mathcal{H})$. This data defines a Weyl system if the following equality is satisfied for all $v, v' \in A$:

$$W(v)w(v') = e^{i\omega(v,v')/2}W(v + v'). \tag{54.6}$$

The relation itself is called a **Weyl relation**.

For every vector $v \in A$, the map $t \mapsto W(tv)$ is a continuous unitary one-parameter subgroup, so by Stone's theorem 23.4.26 one obtains a self-adjoint generator $\phi(v)$. The map $v \mapsto \phi(v)$ is called the associated **Heisenberg system**.

Remark 54.1.7. It should be noted that the Weyl relations are more fundamental than their infinitesimal counterpart. Only the Weyl relations are well-defined on more general spaces and when passing to a relativistic setting.

54.2 Symmetries

54.2.1 Quantum symmetries

Definition 54.2.1 (State space). By the postulates of quantum mechanics, the states in a quantum theory are represented by rays in the projective Hilbert space $\mathbb{P}\mathcal{H}$. Probabilities are defined through the *Fubini-Study metric* on $\mathbb{P}\mathcal{H}$ as follows:

$$\mathcal{P}(\psi, \phi) := \cos^2\left[d_{\text{FS}}(\psi, \phi)\right] = \frac{|\langle\psi|\phi\rangle|^2}{\langle\psi|\psi\rangle\langle\phi|\phi\rangle}, \tag{54.7}$$

where $|\psi\rangle, |\phi\rangle$ are representatives of the states ψ, ϕ in $\mathbb{P}\mathcal{H}$.

Definition 54.2.2 (Symmetry). A quantum symmetry (or **quantum automorphism**) is an isometric automorphism of $P\mathcal{H}$. The group of these symmetries is denoted by $\mathrm{Aut}_{\text{QM}}(\mathbb{P}\mathcal{H})$.

The following theorem due to *Wigner* gives a (linear) characterization of quantum symmetries:[1]

[1]It is a particular case of a more general theorem in projective geometry.

Theorem 54.2.3 (Wigner). *Every quantum automorphism of $\mathbb{P}\mathcal{H}$ is induced by a unitary or anti-unitary operator on $\mathcal{H}$.*

This is equivalent to saying that the group morphism

$$\pi : \mathrm{Aut}(\mathcal{H},\mathcal{P}) := \mathrm{U}(\mathcal{H}) \times \mathrm{AU}(\mathcal{H}) \to \mathrm{Aut}_{\mathrm{QM}}(\mathbb{P}\mathcal{H})$$

is surjective. Together with the kernel U(1), given by phase shifts, this forms a short exact sequence:

$$1 \longrightarrow \mathrm{U}(1) \longrightarrow \mathrm{Aut}(\mathcal{H},\mathcal{P}) \longrightarrow \mathrm{Aut}_{\mathrm{QM}}(\mathbb{P}\mathcal{H}) \longrightarrow 1. \tag{54.8}$$

In the case of symmetry breaking (e.g. lattice systems), the full symmetry group is reduced to a subgroup $G \subset \mathrm{Aut}_{\mathrm{QM}}(\mathbb{P}\mathcal{H})$. The group of operators acting on $\mathcal{H}$ is then given by the pullback $\widetilde{G}$ of the diagram

$$\mathrm{Aut}(\mathcal{H},\mathcal{P}) \longrightarrow \mathrm{Aut}_{\mathrm{QM}}(\mathbb{P}\mathcal{H}) \longleftarrow G.$$

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 $\phi : \mathrm{Aut}(\mathcal{H},\mathcal{P}) \to \mathbb{Z}_2$ that says whether an operator is unitary or anti-unitary).

?? COMPLETE ??

54.2.2 Symmetric states

Axiom 54.4 (Symmetrization postulate). Let $\mathcal{H}$ be the single-particle Hilbert space. A system of n identical particles is described by a state $|\Psi\rangle$ belonging to either $S^n\mathcal{H}$ or $\Lambda^n\mathcal{H}$, i.e. **bosonic** and **fermionic** states are of the form

$$|\Psi_B\rangle = \sum_{\sigma\in S_n} |\psi_{\sigma(1)}\rangle \cdots |\psi_{\sigma(n)}\rangle \tag{54.9}$$

and

$$|\Psi_F\rangle = \sum_{\sigma\in S_n} \mathrm{sgn}(\sigma)|\psi_{\sigma(1)}\rangle \cdots |\psi_{\sigma(n)}\rangle, \tag{54.10}$$

respectively, where the $|\psi_i\rangle$ are single-particle states and S_n is the permutation group on n elements.

Remark 54.2.4. In ordinary quantum mechanics this is a postulate, but in quantum field theory this is a consequence of the *spin-statistics theorem*.

Definition 54.2.5 (Slater determinant). Let $\{\phi_i(\vec{\boldsymbol{q}})\}_{i\leq n}$ 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

$$\psi(\vec{\boldsymbol{q}}_1,\ldots,\vec{\boldsymbol{q}}_n) = \frac{1}{\sqrt{n!}}\det\begin{pmatrix}\phi_1(\vec{\boldsymbol{q}}_1) & \cdots & \phi_n(\vec{\boldsymbol{q}}_1)\\ \vdots & \ddots & \vdots\\ \phi_1(\vec{\boldsymbol{q}}_n) & \cdots & \phi_n(\vec{\boldsymbol{q}}_n)\end{pmatrix} \tag{54.11}$$

A similar function can be defined for bosonic systems using the concept of *permanents*.

54.3 Curved backgrounds ♣

Using the tools of distribution theory and differential geometry (Chapters 17, 31 and onwards), one can introduce quantum mechanics on curved backgrounds (in the sense of "space", not "spacetime").

Remark 54.3.1 (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 *Gelfand triples* or *rigged Hilbert spaces*, where the space of square-integrable functions is replaced by the Schwartz space 17.1.26 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 self-adjointness of the operators (e.g. the angular position operator $\hat{\varphi}$ 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 33.1.21 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:
>
> $$\langle \psi, \phi \rangle := \int d^dx \sqrt{\det(g)}\overline{\psi}(x)\phi(x). \tag{54.12}$$
>
> Because the factor $\sqrt{\det(g)}$ transforms in the inverse manner of the measure d^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 self-adjointness of the momentum operator $\hat{P}_i := -i\nabla_i$:
>
> $$\begin{aligned}\langle \psi, \hat{P}_i\phi \rangle &= \int d^dx \sqrt{\det(g)}\,\overline{\psi}(x)(-i\nabla_i)\phi(x)\\ &\overset{(33.69)}{=} \int d^dx \sqrt{\det(g)}\,\overline{\psi}(x)(-i\partial_i - i\omega_i)\phi(x)\\ &= i\int d^dx \Big(\partial_i\sqrt{\det(g)}\Big)\overline{\psi}(x)\phi(x) + \int d^dx\sqrt{\det(g)}\,\overline{(-i\partial_i\psi)}(x)\phi(x)\\ &\qquad - i\int d^dx\sqrt{\det(g)}\,\overline{\psi}(x)\omega_i\phi(x)\\ &= \langle \hat{P}_i\psi, \phi\rangle - i\int d^dx\sqrt{\det(g)}\,\overline{\psi}(x)\overline{\omega_i}\phi(x)\\ &\qquad + i\int d^dx\Big(\partial_i\sqrt{\det(g)}\Big)\overline{\psi}(x)\phi(x)\\ &\qquad - i\int d^dx\sqrt{\det(g)}\,\overline{\psi}(x)\omega_i\phi(x).\end{aligned}$$

Self-adjointness then requires that

$$\sqrt{\det(g)}(\omega_i + \overline{\omega_i}) = \partial_i \sqrt{\det(g)} \tag{54.13}$$

or

$$2\mathrm{Re}(\omega_i) = \partial_i \ln\Big(\sqrt{\det(g)}\Big). \tag{54.14}$$

?? COMPLETE ??

54.4 Topos theory ♣

Definition 54.4.1 (Bohr topos). Consider a C^*-algebra A of bounded observables on a Hilbert space $\mathcal{H}$. Denote by $\mathrm{Pos}(A)$ the poset 2.6.2 of commutative C^*-subalgebras. This set can be equipped with the **Alexandroff topology**, i.e. the topology for which the open sets are the upward closes subsets. The topological space $(\mathrm{Pos}(A), \tau_{\text{Alex}})$ is called the Bohr site of A.

The sheaf topos over the Bohr site is called the Bohr topos. It can be turned into a ringed topos, where the internal ring object (even an internal C^*-algebra) is given by the tautological functor

$$A : \mathrm{Pos}(A) \rightarrow \mathbf{Set}(A) : C \mapsto C. \tag{54.15}$$

Chapter 55

Angular Momentum

In this chapter the general angular momentum operator $\hat{J} = \left(\hat{J}_x, \hat{J}_y, \hat{J}_z\right)$ is considered.

55.1 General operator

Property 55.1.1 (Lie algebra). The angular momentum operators generate a Lie algebra 30.2.1. The Lie bracket is defined by the following commutation relation:

$$\left[\hat{J}_i, \hat{J}_j\right] = i\hbar\varepsilon_{ijk}\hat{J}_k. \tag{55.1}$$

Since rotations correspond to actions of the orthogonal group SO(3) it should not come as a surprise that the above relation is that of the Lie algebra $\mathfrak{so}(3)$ (Example 30.2.25).

Property 55.1.2. The mutual eigenbasis of $\hat{J}^2$ and $\hat{J}_z$ is defined by the following two eigenvalue equations:

$$\hat{J}^2|j,m\rangle = j(j+1)\hbar^2\,|j,m\rangle \tag{55.2}$$

$$\hat{J}_z|j,m\rangle = m\hbar\,|j,m\rangle. \tag{55.3}$$

Definition 55.1.3 (Ladder operators[1]). The raising and lowering operators $\hat{J}_+$ and $\hat{J}_-$ are defined as follows:

$$\hat{J}_+ := \hat{J}_x + i\hat{J}_y \qquad \text{and} \qquad \hat{J}_- := \hat{J}_x - i\hat{J}_y. \tag{55.4}$$

These operators only change the quantum number m_z, not the total angular momentum.

Corollary 55.1.4. From the commutation relations of the angular momentum operators, one can derive the commutation relations of the ladder operators:

$$\left[\hat{J}_+, \hat{J}_-\right] = 2\hbar\hat{J}_z. \tag{55.5}$$

Formula 55.1.5. 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 (55.1):

$$\hat{J}^2 = \hat{J}_+\hat{J}_- + \hat{J}_z^2 - \hbar\hat{J}_z. \tag{55.6}$$

Remark 55.1.6 (Casimir operator). From the definition of $\hat{J}^2$ it follows that this operator is a Casimir invariant 30.4.59 of $\mathfrak{so}(3)$.

[1] Also called the **creation** and **annihilation** operators (especially in quantum field theory).

55.2 Rotations

55.2.1 Infinitesimal rotation

Formula 55.2.1. An infinitesimal rotation $\hat{R}(\delta\vec{\varphi})$ is given by the following formula:

$$\hat{R}(\delta\vec{\varphi}) = \mathbb{1} - \frac{i}{\hbar}\vec{J}\cdot\delta\vec{\varphi}. \tag{55.7}$$

A finite rotation can be generated by applying this infinitesimal rotation repeatedly:

$$\hat{R}(\vec{\varphi}) = \left(\mathbb{1} - \frac{i}{\hbar}\vec{J}\cdot\frac{\vec{\varphi}}{n}\right)^n = \exp\left(-\frac{i}{\hbar}\vec{J}\cdot\vec{\varphi}\right). \tag{55.8}$$

Formula 55.2.2 (Matrix elements). Applying a rotation over an angle φ about the z-axis to a state $|j,m\rangle$ gives

$$\hat{R}(\varphi\vec{e}_z)|j,m\rangle = \exp\left(-\frac{i}{\hbar}\hat{J}_z\varphi\right)|j,m\rangle = \exp\left(-\frac{i}{\hbar}m\varphi\right)|j,m\rangle. \tag{55.9}$$

Multiplying these states with a bra $\langle j',m'|$ and using the orthonormality of the eigenstates gives the matrix elements of the rotation operator:

$$\hat{R}_{ij}(\varphi\vec{e}_z) = \exp\left(-\frac{i}{\hbar}m\varphi\right)\delta_{jj'}\delta_{mm'}. \tag{55.10}$$

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. 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:

$$\langle j,m'|\hat{R}(\varphi\vec{n})|j,m\rangle = \mathcal{D}^{(j)}_{m,m'}(\hat{R}), \tag{55.11}$$

where the functions $\mathcal{D}^{(j)}_{m,m'}(\hat{R})$ are called the **Wigner D-functions**. For every value of j there are $(2j+1)$ values for m. This implies that the matrix $\mathcal{D}^{(j)}(\hat{R})$ is a $(2j+1)\times(2j+1)$-matrix.

55.2.2 Spinor representation

Definition 55.2.3 (Pauli matrices).

$$\sigma_x := \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} \qquad \sigma_y := \begin{pmatrix}0 & -i\\ i & 0\end{pmatrix} \qquad \sigma_z := \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} \tag{55.12}$$

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 55.2.4. In the spinor representation ($J = \frac{1}{2}$) the Wigner-D matrix reads as follows:

$$\mathcal{D}^{(1/2)}(\varphi\vec{e}_z) = \begin{pmatrix}e^{-i/2\varphi} & 0\\ 0 & e^{i/2\varphi}\end{pmatrix}. \tag{55.13}$$

55.3 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 21.3.19). Because the angular momentum operators $\hat{J}_{l \neq i}$ do not act on the space $\mathcal{H}_i$, one can pull these operators through the tensor product:

$$\hat{J}_i |j_1\rangle \otimes \cdots \otimes |j_n\rangle = |j_1\rangle \otimes \cdots \otimes \hat{J}_i |j_i\rangle \otimes \cdots \otimes |j_n\rangle.$$

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 defined as

$$\hat{J} = \hat{J}_1 + \hat{J}_2. \tag{55.14}$$

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 55.3.1 (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:

$$|J, M\rangle = \sum_{m_1=-j_1}^{j_1} \sum_{m_2=-j_2}^{j_2} |j_1, j_2, m_1, m_2\rangle\langle j_1, j_2, m_1, m_2|J, M\rangle. \tag{55.15}$$

These coefficients are called the Clebsch-Gordan coefficients.

Property 55.3.2. By acting with the operator $\hat{J}_z$ on both sides of Equation (55.15) it is possible to prove that the Clebsch-Gordan coefficients are nonzero if and only if $M = m_1 + m_2$.

Chapter 56

Dirac Equation

References for this chapter are [47]. (Note that the authors use the mostly-pluses signature there.) For the mathematical background of Clifford algebras and Spin groups, see Chapter 25 and, in particular, Section 25.4.2. For the extension to (pseudo-)Riemannian manifolds, see Section 34.3.

56.1 Dirac matrices

Definition 56.1.1 (Dirac matrices). The Dirac (or **gamma**) matrices are defined by the following equality:

$$\{\gamma^\mu, \gamma^\nu\}_+ = 2\eta^{\mu\nu}\mathbb{1}, \tag{56.1}$$

where $\eta^{\mu\nu}$ is the Minkowski metric. This has the form of Equation (25.8), 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 matrix γ^0 is defined as

$$\gamma^0 := \begin{pmatrix} \mathbb{1}_2 & 0 \\ 0 & -\mathbb{1}_2 \end{pmatrix}. \tag{56.2}$$

The spacelike Dirac matrices γ^k $(k = 1,2,3)$ are defined using the Pauli matrices 55.2.3 σ^k:

$$\gamma^k := \begin{pmatrix} 0 & \sigma^k \\ -\sigma^k & 0 \end{pmatrix}. \tag{56.3}$$

The **Weyl** or **chiral** representation[1] is defined by replacing the timelike matrix γ^0 by

$$\gamma^0 := \begin{pmatrix} 0 & \mathbb{1}_2 \\ \mathbb{1}_2 & 0 \end{pmatrix}. \tag{56.4}$$

In signature $(3,1)$ one obtains the Weyl representation by defining $\sigma^\mu := (\mathbb{1}, \sigma_i)$ and $\overline{\sigma}^\mu := \sigma_\mu$:

$$\gamma^\mu := \begin{pmatrix} 0 & \sigma_\mu \\ \overline{\sigma}_\mu & 0 \end{pmatrix}. \tag{56.5}$$

Remark 56.1.2. In the remainder of this compendium the Weyl representation will be used.

[1] This representation is widely used in quantum field theory and supergravity.

Notation 56.1.3 (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:

$$\not{a} := a^\mu \gamma_\mu. \tag{56.6}$$

In fact this is just the embedding of Minkowski space in its Clifford algebra:

$$/ : M^4 \to C\ell(M^4, \eta) : a^\mu \mathbf{e}_\mu \mapsto a^\mu \gamma_\mu. \tag{56.7}$$

56.2 Spinors

56.2.1 Dirac equation

Formula 56.2.1 (Dirac equation). In covariant form the Dirac equation reads as

$$(i\hbar \not{\partial} - mc)\psi = 0, \tag{56.8}$$

where m denotes the mass and c denotes the speed of light.

Definition 56.2.2 (Dirac adjoint).

$$\overline{\psi} := i\psi^\dagger \gamma^0 \tag{56.9}$$

When working in the Dirac representation the factor i should be dropped.

Definition 56.2.3 (Majorana adjoint). In the context of SUSY it is often convenient to work with a different adjoint spinor. Let $\mathcal{C} := i\gamma^3\gamma^1$ denote the charge conjugation operator. The Majorana adjoint is defined by

$$\overline{\psi} := \psi^t C. \tag{56.10}$$

Formula 56.2.4 (Parity). The parity operator is defined as follows:

$$\hat{P}(\psi) = \gamma^0 \psi. \tag{56.11}$$

56.2.2 Chiral spinors

In even dimensions one can define an additional matrix that satisfies Equation (56.1):

Definition 56.2.5 (Chiral matrix). Assume that the dimension, given by $d = m + n$, is even. The chiral (helicity) matrix can be defined as follows:[2]

$$\gamma_{d+1} := \gamma_1 \gamma_2 \cdots \gamma_d. \tag{56.12}$$

In odd 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 $k\gamma_*$ 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:

$$\gamma_{d+1} \longrightarrow k\gamma_{d+1}. \tag{56.13}$$

This has the benefit that $\gamma_{d+1}^2 = \mathbb{1}$.

In $d = 3 + 1$ one generally takes the following representation for γ_5:[3]

$$\gamma_{d+1} := \begin{pmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{pmatrix}. \tag{56.14}$$

[2]Some authors add a constant to this definition.

[3]Such a block diagonal form can always be chosen by working in a *helicity-adapted basis*.

Definition 56.2.6 (Chiral projection). The chiral projections of a spinor ψ are defined as follows:

$$\psi_L := \frac{1+\gamma_{d+1}}{2}\psi \tag{56.15}$$

and

$$\psi_R := \frac{1-\gamma_{d+1}}{2}\psi. \tag{56.16}$$

Every spinor can then be written as a sum of its chiral parts:

$$\psi = \psi_L + \psi_R. \tag{56.17}$$

56.2.3 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 56.2.7 (Trace algebra).

$$\mathrm{tr}(\gamma^\mu) = \mathrm{tr}(\gamma^\mu\gamma^\nu\gamma^\rho) = 0 \tag{56.18}$$

$$\mathrm{tr}(\gamma^\mu\gamma^\nu) = 4\eta^{\mu\nu} \tag{56.19}$$

$$\mathrm{tr}(\gamma^\mu\gamma^\nu\gamma^\kappa\gamma^\lambda) = 4(\eta^{\mu\nu}\eta^{\kappa\lambda} - \eta^{\mu\kappa}\eta^{\nu\lambda} + \eta^{\mu\lambda}\eta^{\nu\kappa}) \tag{56.20}$$

$$\mathrm{tr}(\gamma^5) = \mathrm{tr}(\gamma^\mu\gamma^5) = \mathrm{tr}(\gamma^\mu\gamma^\nu\gamma^5) = \mathrm{tr}(\gamma^\mu\gamma^\nu\gamma^\rho\gamma^5) = 0 \tag{56.21}$$

$$\mathrm{tr}(\gamma^\mu\gamma^\nu\gamma^\kappa\gamma^\lambda\gamma^5) = -4i\varepsilon^{\mu\nu\kappa\lambda} \tag{56.22}$$

$$\mathrm{tr}(\gamma^{\mu_1}\cdots\gamma^{\mu_k}) = \mathrm{tr}(\gamma^{\mu_k}\cdots\gamma^{\mu_1}) \tag{56.23}$$

Formula 56.2.8 (Contraction identities).

$$\gamma^\mu\gamma_\mu = 4 \tag{56.24}$$

$$\gamma^\mu\gamma^\nu\gamma_\mu = -2\gamma^\nu \tag{56.25}$$

$$\gamma^\mu\gamma^\nu\gamma^\rho\gamma_\mu = 4\eta^{\nu\rho} \tag{56.26}$$

$$\gamma^\mu\gamma^\nu\gamma^\kappa\gamma^\lambda\gamma_\mu = -2\gamma^\lambda\gamma^\kappa\gamma^\nu \tag{56.27}$$

56.2.4 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} where $i, j = 1, \ldots, 2^D$ such that

$$(\overline{v}u)(\overline{\rho}\omega) = \sum_{i,j=1}^{2^d} a^{ij}(\overline{v}\gamma_i\omega)(\overline{\rho}\gamma_j u). \tag{56.28}$$

By using the trace orthogonality relations one can find that

$$\alpha^{ij} = \begin{cases} 0 & i \neq j \\ \frac{1}{2^{\lfloor d/2 \rfloor}} & i = j. \end{cases} \tag{56.29}$$

The above equality can then also be rewritten as follows:

$$\delta^a_b\delta^c_d = \frac{1}{2^{\lfloor d/2 \rfloor}}\sum_{i=1}^{2^d}(\gamma_i)^a_d(\gamma_i)^c_b. \tag{56.30}$$

This expression (and the techniques used to find it) allow one to rearrange almost all multilinear expressions involving spinors and cospinors.

Chapter 57

Quantization ♣

The content of Chapters 33, 32 and 35 are prerequisites for the section on *geometric quantization*.

57.1 Introduction

Given the content of Chapter 54, the general quantization procedure for a dynamical system (M, ω, H) can be axiomatized as follows:

Method 57.1.1 (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.
- $\mathcal{O}(1) = \mathbb{1}_{\mathcal{H}}$.
- **Dirac correspondence**: $[\mathcal{O}(f), \mathcal{O}(g)] = i\hbar\mathcal{O}(\{f, g\})$.

Because a Hilbert space forms an irreducible representation of any complete set of observables, it makes sense to additionally require the following axiom:

Axiom 57.1 (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 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 U(1)-central extension of the translation group $\mathbb{R}^{2n}$. Irreducible representations are characterized by the Stone-von Neumann theorem 54.1.4.

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 57.1.1, that takes the entire (Lie) algebra of classical observables to the (Lie) algebra of corresponding quantum observables.

57.2 Deformation quantization

Definition 57.2.1 (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 $a, b \in A$:

$$a * b = \sum_{i=0}^{\infty} B_i(a,b)\hbar^i \tag{57.1}$$

with $B_0(a,b) = ab$. A formal deformation quanitzation of A is a star product such that

$$\{a,b\} = B_1(a,b) - B_1(b,a). \tag{57.2}$$

Definition 57.2.2 (Moyal deformation quantization). Let V be a Poisson vector space, i.e. a vector space equipped with a Poisson bivector. Denote this bivector by π. The space of smooth functions $C^\infty(V)[[\hbar]]$ can be equipped with the following product:

$$f * g := \mu \circ \exp(\hbar\pi)(f \otimes g), \tag{57.3}$$

where π is viewed as an endomorphism on $C^\infty(V) \otimes C^\infty(V)$ and μ is the ordinary product map.

Definition 57.2.3 (Fedosov deformation quantization). Let (M, ω) be a symplectic manifold. 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 be shown that any such bundle admits at least one connection that respects the algebra structure, the **Fedosov connection**.

The space of covariantly constant sections admits the following isomorphism:

$$\ker(\nabla^{\text{Fed}}) \cong C^\infty(M)[[\hbar]]. \tag{57.4}$$

The restriction of the star product on $\Gamma(\mathcal{A})$ gives a formal deformation quantization of $C^\infty(M)$.

Remark 57.2.4 (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.

57.3 Geometric Quantization

In this section the constant $\hbar$ has been set to 1.

57.3.1 Prequantization

Definition 57.3.1 (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 57.3.2. 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

$$\forall S \in H_2(M) : \int_S \omega \in 2\pi\mathbb{Z}. \tag{57.5}$$

This condition resembles the "old" *Bohr-Sommerfeld condition* 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 33.7. 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 nonequivalent bundles and there can exist nonequivalent connections on the same bundle differing by a flat connection, where two flat connections are in turn nonequivalent 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 $H^1(M;\mathrm{U}(1))$ or, equivalently by isomorphism (32.132), $H^2(M;\mathbb{Z})$.

Corollary 57.3.3 (Dirac quantization condition). One can derive the Dirac quantization condition from Weil integrality. If one couples the system to a gauge potential, the minimal coupling procedure gives $\omega \longrightarrow \omega + eF$. Weil integrality then implies that e is an integer.

Definition 57.3.4 (Prequantum Hilbert space). Consider a symplectic manifold (M,ω) together with a prequantum line bundle L. The prequantum Hilbert space $\mathcal{H}_{\mathbb{P}}$ 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:

$$\hat{f} : \psi \to -i\nabla_{X_f}\psi + f\cdot\psi, \tag{57.6}$$

where X_f is the Hamiltonian vector field associated to f and, locally, $\nabla = d - i\theta$ with θ the Liouville one-form. 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 automorphism $\psi_t[f]$ on L that preserves both the metric and the connection. The prequantum operator $\hat{f}$ is then simply given by

$$\hat{f}s = -i\frac{d}{dt}(\psi_t[f]s)\bigg|_{t=0}. \tag{57.7}$$

Example 57.3.5 (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 $\omega = r^2 \sin\theta d\theta \wedge d\varphi$. 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 angular momentum. The reason for this is that S^2 is a homogeneous space for 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 $S^2 = \mathrm{SU}(2)/\mathrm{U}(1)$.

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 57.3.6 (Quantization). Consider a symplectic manifold (M, ω). A quantization of (M, ω) is given by a prequantum line bundle L together with a polarization P of M. The "naive" modification of the quantum state space $\mathcal{H}$ is given by the subspace of $\mathcal{H}_{\mathbb{P}}$ of those sections that are covariantly constant along P, i.e. those sections $s \in \Gamma(L)$ that satisfy $\nabla_X s = 0$ for all $X \in P$. 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

$$\omega(X, Y)s = [\nabla_X, \nabla_Y]s - \nabla_{[X,Y]}s = 0 \tag{57.8}$$

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

$$\nabla_X s = 0 \implies \nabla_X(\hat{f}s) = 0$$

for all sections $s \in \Gamma(L)$ and $X \in P$. Using the general formula for iterated covariant derivatives and the fact that the leaves of P are Lagrangian, one finds the following condition:

$$[X, X_f] = 0 \tag{57.9}$$

for all $X \in P$. So, in general, one should restrict to the subspace of $C^\infty(M)$ on those functions whose Hamiltonian flow preserves P.

There is, however, a problem with this construction. Nothing ensures that $\mathcal{H}_{\mathbb{P}}$ 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 distribution of 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 57.3.7 (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 37.1.15, P is assumed to admit a metalinear structure. 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 ψ of $L \otimes \delta^{1/2}$ such that locally $\psi = \lambda \otimes \mu$ with $\nabla_X \lambda = 0$ and $\mathcal{L}_X \mu = 0$ for all $X \in P$. By pairing two wave functions one obtains a 1-density on M which 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:

$$\hat{f}(s \otimes \mu) := (\hat{f}s) \otimes \mu - is \otimes \mathcal{L}_{X_f}\mu. \tag{57.10}$$

Example 57.3.8 (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

$$\langle\psi_1|\psi_2\rangle = \int_{\mathbb{C}^n} \overline{\psi_1}(z)\psi_2(z)\exp(-|z|^2/2)dz^n. \tag{57.11}$$

This is often called the **Bargmann**, **Segal-Bargmann** or **Bargmann-Fock** representation. The coordinates $z, \overline{z}$ are represented by the operators z and ∂_z. These correspond to creation and annihilation operators.

Remark 57.3.9. 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.

Method 57.3.10 (Bohr-Sommerfeld quantization). Here, the polarization P is again assumed to be real, but the leaves are not required to be simply-connected. The partial connection ∇ along 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

$$\exp\left(i\oint_\gamma \theta\right) = 1 \tag{57.12}$$

for all loops γ, where θ is the symplectic potential. In terms of Darboux coordinates this gives (up to a factor 2π)

$$\oint_\gamma p_i dq^i \in \mathbb{Z}. \tag{57.13}$$

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:

$$\oint_\gamma p_i dq^i = 2\pi(k_\gamma + d_\gamma), \tag{57.14}$$

where $k_\gamma \in \mathbb{Z}$.

57.3.2 Coadjoint orbits

?? COMPLETE (SEE ALSO 30.3.3) ??

57.4 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 49.

57.4.1 Dirac procedure

The first approach to the quantization of constraint 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 $\hat{G}_a$, satisfy

$$\hat{G}_a|\psi\rangle = 0 \tag{57.15}$$

for all physical states $|\psi\rangle$. The constraint algebra $\{\phi_a, \phi_b\} = C^c_{ab}\phi_c$ gives rise to a quantum algebra

$$[\hat{G}_a, \hat{G}_b] = i\hbar C^c_{ab}\hat{G}_c + \hbar^2\hat{D}_{ab}, \tag{57.16}$$

where $\hat{D}_{ab}$ 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 $\hat{D}_{ab}$ 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 empty). 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

$$\{H, \phi_a\} = V^b_a\phi_b. \tag{57.17}$$

?? COMPLETE ??

57.4.2 BRST quantization

For this approach one starts from the classical BRST construction from Section 49.3 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

$$[\Omega, \Omega] = 2\Omega^2 = 0. \tag{57.18}$$

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 57.4.1 (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

$$g = g_0 + z \in \mathbb{Z} \tag{57.19}$$

  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 57.4.2 (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:

$$\Omega|\psi\rangle = 0. \tag{57.20}$$

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 $H^0(\Omega)$.

57.4.3 BFV quantization

The extended phase space M_{ext} is defined by introducing dynamical Lagrange multipliers and their momenta

$$\{\lambda^m, \pi_n\} := \delta^m_n \tag{57.21}$$

together with two collections of (homologically) odd-degree ghosts and their momenta:

$$\{C^i, \overline{P}_j\} := \delta^i_j \tag{57.22}$$

$$\{P^i, \overline{C}_j\} := \delta^i_j. \tag{57.23}$$

Note that the Poisson brackets 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

$$\mathcal{G} := \frac{1}{2}\Big(C^i\overline{P}_i - \overline{P}^i C_i + P^i\overline{C}_i - \overline{C}_i P^i\Big). \tag{57.24}$$

This operator acts on observables as follows:

$$[\mathcal{G}, A] = i\hbar\, \mathrm{gh}(A) A. \tag{57.25}$$

BFV quantization is obtained by defining a nilpotent BRST operator $\Omega = \Omega_{\text{min}} + \Omega_{\text{aux}}$.

Chapter 58

Perturbation Theory

58.1 Interaction picture

Let $\hat{H}(t) = \hat{H}_0 + \hat{V}(t)$ be the total Hamiltonian of a system, where $\hat{V}(t)$ denotes the interaction potential. Let $|\psi(t)\rangle$ and $\hat{O}$ denote a state and operator in the Schrödinger picture.

Definition 58.1.1 (Interaction representation). In the interaction picture the state vector is defined as follows:

$$|\psi(t)\rangle_I := e^{\frac{i}{\hbar}\hat{H}_0 t}|\psi(t)\rangle. \tag{58.1}$$

From this it follows that the operators in the interaction picture are given by

$$\hat{O}_I(t) = e^{\frac{i}{\hbar}\hat{H}_0 t}\,\hat{O}\,e^{-\frac{i}{\hbar}\hat{H}_0 t}. \tag{58.2}$$

Formula 58.1.2 (Schrödinger equation). Using the previous formulas, the Schrödinger equation can be rewritten as follows:

$$i\hbar\frac{d}{dt}|\psi(t)\rangle_I = \hat{V}_I(t)|\psi(t)\rangle_I. \tag{58.3}$$

The time-evolution of operators in the interaction picture is given by

$$\frac{d}{dt}\hat{O}_I(t) = \frac{i}{\hbar}\left[\hat{H}_0, \hat{O}_I(t)\right]. \tag{58.4}$$

States evolve solely based on the interaction terms and operators evolve according to the free (time-independent) Hamiltonian.

Theorem 58.1.3 (Adiabatic theorem). *If a perturbation is acting slowly enough on a system 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*. However, later this was shown to be inessential.

Definition 58.1.4 (Møller operator). The operators relating the (free) asymptotic states to the interacting state:

$$|\psi(0)\rangle = \Omega_\pm|\text{in/out}\rangle. \tag{58.5}$$

If for $t \longrightarrow \pm\infty$ the interaction term vanishes, there should exist a wavefunction $|\phi(t)\rangle$ evolving under the free Hamiltonian such that

$$\|\psi(t) - \phi(t)\| \overset{t\to\pm\infty}{\longrightarrow} 0. \tag{58.6}$$

With $|\text{in/out}\rangle := \lim_{t\to\pm\infty} \phi(t)$ this gives:

$$\Omega_\pm = \lim_{t\to\mp\infty} \exp\Big(\tfrac{i}{\hbar}\hat{H}(t)t\Big) \exp\Big(-\tfrac{i}{\hbar}\hat{H}_0 t\Big) \tag{58.7}$$

with respect to the strong topology.

Remark 58.1.5. The limit in the definition of the Møller operators will only result in a well-defined operator for infinite-dimensional spaces.

58.2 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 $\hat{H}_0$ also form a complete set for the perturbed Hamiltonian.

Formula 58.2.1. The perturbed eigenfunctions and eigenvalues can be expanded in the following way, where λ is assumed to be a small (perturbation) parameter:

$$|\psi_n\rangle = \sum_{i=0}^{\infty} \lambda^i |\psi_n^{(i)}\rangle \tag{58.8}$$

$$E_n = \sum_{i=0}^{\infty} \lambda^i E_n^{(i)}, \tag{58.9}$$

where i is called the **order** of the perturbation.

58.3 Time-dependent perturbation theory

58.3.1 Dyson series

Formula 58.3.1 (Tomonaga-Schwinger equation). The evolution operator $\hat{U}(t)$ satisfies the following Schrödinger-type equation in the interaction picture (Section 58.1):

$$i\hbar\frac{d}{dt}\hat{U}_I|\psi(0)\rangle_I = \hat{V}_I(t)\hat{U}_I|\psi(0)\rangle_I. \tag{58.10}$$

Formula 58.3.2 (Dyson series). Together with the initial value condition $\hat{U}_I(0) = \mathbb{1}$, the Tomonaga-Schwinger equation becomes an initial value problem. A particular solution is given by

$$\hat{U}_I(t) = \mathbb{1} - \frac{i}{\hbar}\int_0^t \hat{V}_I(t')\hat{U}_I(t')dt'. \tag{58.11}$$

This solution can be iterated to obtain a series expansion of the evolution operator:

$$\hat{U}(t) = 1 - \frac{i}{\hbar}\int_0^t \hat{V}(t_1)dt_1 + \left(-\frac{i}{\hbar}\right)^2 \int_0^t dt_1 \int_0^{t_1} dt_2 \hat{V}(t_1)\hat{V}(t_2) + \cdots \tag{58.12}$$

It is clear that the integrands obey a time-ordering. By introducing the **time-ordering operator**

$$\mathcal{T}\Big(\hat{V}(t_1)\hat{V}(t_2)\Big) = \begin{cases} \hat{V}(t_1)\hat{V}(t_2) & t_1 \geq t_2 \\ \hat{V}(t_2)\hat{V}(t_1) & t_2 > t_1 \end{cases} \tag{58.13}$$

the integrals can be rewritten in a more symmetric form:

$$\hat{U}(t) = 1 - \frac{i}{\hbar}\int_0^t \hat{V}(t_1)dt_1 + \frac{1}{2!}\left(-\frac{i}{\hbar}\right)\int_0^t dt_1 \int_0^t dt_2 \mathcal{T}\left(\hat{V}(t_1)\hat{V}(t_2)\right) + \cdots . \tag{58.14}$$

By comparing this expression to the series expansion for exponential functions, the following concise formula is obtained:

$$\hat{U}(t) = \mathcal{T}\left(e^{-\frac{i}{\hbar}\int_0^t \hat{V}(t')dt'}\right). \tag{58.15}$$

This formula is called the **Dyson series**.

58.4 Variational method

Definition 58.4.1 (Energy functional).

$$E[\psi] := \frac{\langle\psi|\hat{H}|\psi\rangle}{\langle\psi|\psi\rangle} \tag{58.16}$$

Property 58.4.2 (Bounded below). The energy functional satisfies the following inequality:

$$E[\psi] \geq E_0, \tag{58.17}$$

where E_0 is the ground state energy.

Method 58.4.3. Assume that the ansatz $|\psi\rangle$ depends on a set of parameters $\{\lambda_i\}_{i\in I}$. The "optimal" wave function, i.e. the one extremizing the energy functional, is found by solving the following system of equations:

$$\frac{\partial\psi}{\partial\lambda_i} = 0 \qquad \forall i \in I. \tag{58.18}$$

58.5 Adiabatic approximation

58.5.1 Berry phase

Consider a system for which the adiabatic approximation is valid. The wave function is then of the form

$$\psi(t) = C_a(t)\psi_a(t)\exp\left(-\frac{i}{\hbar}\int_{t_0}^t E_a(t')dt'\right). \tag{58.19}$$

It follows from the orthonormality of the eigenstates $\psi_k(t)$ that the coefficient $C_a(t)$ is just a phase factor, so it can be rewritten as

$$C_a(t) = e^{i\gamma_a(t)}. \tag{58.20}$$

By substituting this ansatz in the wave function, the Schödinger equation gives a differential equation for the phase factor $\gamma_a(t)$. Integration gives

$$\gamma_a(t) = i\int_{t_0}^t \left\langle \psi_a(t') \middle| \frac{\partial\psi_a(t')}{\partial t'} \right\rangle dt'. \tag{58.21}$$

Due to time evolution, the wave function accumulates a phase over the period $t_0 - t_f$ through the coefficient $C_a(t)$. This phase is called the **Berry phase**.

Now, apply a phase transformation to remove the Berry phase:

$$\psi'_a(t) := \psi_a(t)e^{i\eta(t)}. \tag{58.22}$$

Inserting this in Equation (58.21) gives

$$\bar{\gamma}'_a(t) = \bar{\gamma}_a(t) - \eta(t_f) + \eta(t_0), \tag{58.23}$$

where the bar denotes the integration between t_0 and t_f in Equation (58.21). If the system is cyclic, then $\psi_a(t_0) = \psi_a(t_f)$. Combining this with Equation (58.22) gives

$$\eta(t_f) - \eta(t_0) = 2k\pi \tag{58.24}$$

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 58.5.1 (Berry connection). The quantity

$$\mathbf{A}(\vec{\boldsymbol{x}}) := i\langle\psi_a(\vec{\boldsymbol{x}})|\nabla_{\vec{\boldsymbol{x}}}\psi_a(\vec{\boldsymbol{x}})\rangle, \tag{58.25}$$

where $\nabla_{\vec{\boldsymbol{x}}}$ denotes the gradient in phase space, is called the Berry connection (or **Berry gauge potential**). Applying Stokes's theorem to (58.21) gives

$$\bar{\gamma}_a = \int \boldsymbol{\mathcal{B}} \cdot d\vec{\boldsymbol{S}}, \tag{58.26}$$

where $\boldsymbol{\mathcal{B}} = \nabla_{\vec{\boldsymbol{x}}} \times \mathbf{A}(\vec{\boldsymbol{x}})$ is called the **Berry curvature**. Although the Berry connection is gauge-dependent, the Berry curvature is gauge-invariant

Remark 58.5.2 (♣). Using the language of differential geometry (Chapter 33) one immediately finds that the accumulated phase $\bar{\gamma}_a$ is simply the holonomy associated with the Berry connection along the considered trajectory.

Chapter 59

Scattering Theory

59.1 Cross sections

Formula 59.1.1 (Differential cross section).

$$\frac{d\sigma}{d\Omega} = \frac{N(\theta, \varphi)}{F}, \tag{59.1}$$

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 59.1.2 (Fermi's golden rule). The transition probability from an initial state to a final state is given by

$$\Gamma_{i\to f} = \frac{2\pi}{\hbar}|\langle f|\hat{H}|i\rangle|^2 \frac{dn}{dE_f}. \tag{59.2}$$

59.2 Lippman-Schwinger equations

In this section Hamiltonians of the form $\hat{H} = \hat{H}_0 + \hat{V}$ are considered, where $\hat{H}_0$ is the free Hamiltonian and $\hat{V}$ 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 59.2.1 (Lippman-Schwinger equation).

$$|\psi^{(\pm)}\rangle = |\varphi\rangle + \frac{1}{E - \hat{H}_0 \pm i\varepsilon}\hat{V}|\psi^{(\pm)}\rangle, \tag{59.3}$$

where $|\varphi\rangle$ is an eigenstate of the free Hamiltonian with the same energy as $|\psi\rangle$, i.e. $\hat{H}_0|\varphi\rangle = E|\varphi\rangle$.

Remark 59.2.2. The term $\pm i\varepsilon$ is added to the denominator because otherwise it would be singular. The term has no real physical meaning.

Formula 59.2.3 (Born series). If the Lippman-Schwinger equation is rewritten as

$$|\psi\rangle = |\varphi\rangle + \hat{G}_0\hat{V}|\psi\rangle, \tag{59.4}$$

where $\hat{G}_0$ is the Green's operator 19.5.7 associated to $\hat{H}_0$, one can derive the following series expansion by iterating the above expression:

$$|\psi\rangle = |\varphi\rangle + \hat{G}_0\hat{V}|\varphi\rangle + \left(\hat{G}_0\hat{V}\right)^2|\varphi\rangle + \cdots. \tag{59.5}$$

Convergence issues of this series can be resolved through a Borel resummation procedure 14.7.6.

Formula 59.2.4 (Born approximation). If the Born series is truncated at the first-order term in $\hat{V}$, the Born approximation is obtained:

$$|\psi\rangle = |\varphi\rangle + \hat{G}_0\hat{V}|\varphi\rangle. \tag{59.6}$$

Chapter 60

Quantum Information Theory

60.1 Entanglement

60.1.1 Schmidt decomposition

Construction 60.1.1 (Schmidt decomposition). Consider a bipartite state $|\psi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$. For any such state there exist orthonormal sets $\big\{|e_i\rangle, |f_j\rangle\big\}_{i,j\leq\kappa}$ such that

$$|\psi\rangle = \sum_{i=1}^{\kappa} \lambda_i |e_i\rangle \otimes |f_i\rangle, \tag{60.1}$$

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 κ is called the **Schmidt rank** of $|\psi\rangle$.

Definition 60.1.2 (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.

60.1.2 Bell states

Definition 60.1.3 (Bell state). A (binary) Bell state (also called a **cat state** or **Einstein-Podolsky-Rosen pair**) is defined as the following entangled state:

$$|\Phi^+\rangle := \frac{1}{\sqrt{2}}\Big(|00\rangle + |11\rangle\Big). \tag{60.2}$$

In fact this state can be extended to a full maximally entangled basis for the 2-qubit Hilbert space:

$$\begin{aligned}
|\Phi^-\rangle &:= \frac{1}{\sqrt{2}}\Big(|00\rangle - |11\rangle\Big)\\
|\Psi^+\rangle &:= \frac{1}{\sqrt{2}}\Big(|01\rangle + |10\rangle\Big)\\
|\Psi^-\rangle &:= \frac{1}{\sqrt{2}}\Big(|01\rangle - |10\rangle\Big).
\end{aligned} \tag{60.3}$$

[1]Sometimes called **superdense coding**.

Method 60.1.4 (Dense coding[1]). Consider the Bell state $|\Phi^+\rangle$. By acting with one of the (unitary) spin-flip operators X, Y, Z one can obtain any of the other three Bell states:

$$\begin{aligned} X|\Phi^+\rangle &= |\Phi^-\rangle \\ Y|\Phi^+\rangle &= |\Psi^+\rangle \\ Z|\Phi^+\rangle &= |\Psi^-\rangle. \end{aligned} \tag{60.4}$$

In a typical Alice-and-Bob-style experiment one can ask the question if 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 ($\mathbb{1}, X, Y, Z$). 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 60.1.5 (GHZ[2] state). The GHZ state is defined as the multiparticle qudit ($d, N > 2$) version of the Bell state above and is, hence, also referenced to as a cat state:

$$|\mathrm{GHZ}\rangle = \frac{1}{\sqrt{d}} \sum_{i=0}^{d-1} |i\rangle^{\otimes N}. \tag{60.5}$$

60.2 Density operators

Definition 60.2.1 (Density operator). Consider a (finite-dimensional) Hilbert space $\mathcal{H}$. A density operator on $\mathcal{H}$ is a linear operator $\rho \in \mathrm{End}(\mathcal{H})$ satisfying the following properties:

1. **Positivity**: $\langle v|\rho v\rangle \geq 0$ for all $v \in \mathcal{H}$.
2. **Hermiticity**: $\rho^\dagger = \rho$.
3. **Unit trace**: $\mathrm{tr}(\rho) = 1$.

More concisely, density operators are the representing objects of normal states 24.1.28 on $\mathcal{B}(\mathcal{H})$.

Example 60.2.2 (Classical probability). A diagonal density matrix corresponds to a (classical) discrete probability distribution.

Definition 60.2.3 (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. A density matrix that is not of this form gives rise to a **mixed state**.

Definition 60.2.4 (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 $\hat{\rho}_A$ of A is defined as follows:

$$\hat{\rho}_A := \mathrm{tr}_B\, |\Psi\rangle\langle\Psi|. \tag{60.6}$$

Definition 60.2.5 (Purification). Let $\hat{\rho}_A$ be the density operator of a system A. A purification of $\hat{\rho}_A$ is a pure state $|\Psi\rangle$ of some composite system AB such that

$$\hat{\rho}_A = \mathrm{tr}_B\, |\Psi\rangle\langle\Psi|. \tag{60.7}$$

Property 60.2.6. Any two purifications of the same density operator $\hat{\rho}_A$ are related by a transformation $\mathbb{1}_A \otimes \hat{V}$ with $\hat{V}$ an isometry.

[2] *Greenberger-Horne-Zeilinger*

60.3 Operations

The following definition generalizes the content of Section 23.4.6 to a setting of partial information:

Definition 60.3.1 (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

$$\sum_{i=1}^{n} P_i = \mathbb{1}_{\mathcal{H}}. \tag{60.8}$$

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 $(E_n)_{n\in\mathbb{N}} \subset \Sigma$:

$$\sum_{n\in\mathbb{N}} P_{E_n} = P_{\bigcup_{n\in\mathbb{N}} E_n}. \tag{60.9}$$

The following theorem can be derived from the Stinespring theorem 24.1.31:

Theorem 60.3.2 (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*

$$P(\cdot) = V\Pi(\cdot)V^\dagger. \tag{60.10}$$

In the finite-dimensional setting V can be chosen to be an isometry.

Definition 60.3.3 (Completely positive trace-preserving). Consider a map $\Phi : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ between (trace-class) operators on two (finite-dimensional) Hilbert spaces. This map preserves density matrices if it positive 24.1.13 and if it is trace-preserving 24.1.24. 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 24.1.14.

Completely positive, trace-preserving (CPTP) maps are often called **quantum channels**.

The following property can be derived from the Stinespring theorem 24.1.31:

Property 60.3.4 (Kraus decomposition). Let $\mathcal{H}_1, \mathcal{H}_2$ be Hilbert spaces of dimensions m and n, respectively. A linear map $\Phi : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ is completely positive if and only if there exist bounded operators $\{A_i\}_{i\leq mn}$ such that

$$\Phi(B) = \sum_{i=1}^{mn} A_i^\dagger B A_i. \tag{60.11}$$

Furthermore, it is trace-preserving if and only if

$$\sum_{i=1}^{mn} A_i^\dagger A_i = \mathbb{1}. \tag{60.12}$$

A decomposition of the above form is also often called an **operator-sum decomposition**.

Part IX

Statistical Mechanics & Condensed Matter Physics

Chapter 61

Thermodynamics

61.1 General definitions

Definition 61.1.1 (System). The part of space that is of interest.

Definition 61.1.2 (Environment). The complement of the system in space. More specifically, this denotes the part of space that has a potential influence on the system.

Definition 61.1.3 (Thermodynamic coordinate). Macroscopical variable that describes the system. These are also called **state variables**.

Definition 61.1.4 (Intensive coordinate). Coordinate that does not depend on the system size. The opposite notion is called an **extensive coordinate**.

Definition 61.1.5 (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 61.1.6 (Uniformity). During thermodynamic equilibrium all intensive coordinates are uniform throughout the system.

Definition 61.1.7 (Isolated system). A system that cannot interact with its environment (e.g. due to the presence of impenetrable walls).

Definition 61.1.8 (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 61.1.9 (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 61.1.10 (Open system). A system that is allowed to interact with its environment.

Definition 61.1.11 (Quasistatic process). A sequence of equilibrium states separated by infinitesimal changes.

Definition 61.1.12 (Path). The sequence of equilibrium states in a thermodynamic process is called its path.

61.2 Postulates

Axiom 61.1 (Zeroth law). If two objects are in thermal equilibrium with a third object, they are also in thermal equilibrium with each other.

Axiom 61.2 (First law). The change in internal energy is given by

$$\Delta U = Q + W, \tag{61.1}$$

or, infinitesimally, by

$$dU = \delta Q + \delta W, \tag{61.2}$$

where W denotes the work done on the system and Q denotes the heat that was extracted from the environment.

Remark. The 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 32.4 for more information on differential forms.

Axiom 61.3 (Kelvin-Planck formulation of the second law). No machine can absorb an amount of heat and completely transform it into work.

Axiom 61.4 (Clausius formulation of the second law). Heat cannot be passed from a cooler object to a warmer object without performing work.

Formula 61.2.1 (Clausius's inequality). In differential form the inequality reads as

$$\frac{\delta Q}{T} \geq 0. \tag{61.3}$$

?? COMPLETE THIS STATEMENT (WHICH INEQUALITY?) ??

Axiom 61.5 (Third law). No process can reach absolute zero through a finite sequence of operations.

61.3 Equations of state

Property 61.3.1 (PV-system). Consider a system described by pressure P and volume V. The first law becomes

$$\delta Q = dU + PdV. \tag{61.4}$$

This formula defines a Pfaffian system 32.5.7. Since the state space is only two-dimensional, it is not too hard to see that it is integrable, i.e. there exists a function T, called an **integrating factor**, and a one-form dS such that

$$TdS = dU + PdV. \tag{61.5}$$

T and S represent the temperature and entropy of the system, respectively. If the heat exchange is zero, the entropy is conserved.

61.4 Gases

Formula 61.4.1 (Ideal gas law).

$$PV = nRT, \tag{61.6}$$

where R is the **ideal gas constant** $R \approx 8.314 \frac{\mathrm{J}}{\mathrm{K\,mol}}$.

Chapter 62

Statistical Mechanics

62.1 Axioms

Axiom 62.1 (Ergodic principle). All microstates corresponding to the same macrostate are equally probable.

Axiom 62.2 (Boltzmann formula). The entropy of a system is given by

$$S := k \ln \Omega(E, V, N, \alpha), \tag{62.1}$$

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 43.7.2.

62.2 Temperature

Definition 62.2.1. The temperature of a system in contact with a heat bath is defined as follows:

$$T := \left(\frac{\partial E}{\partial S}\right)_V. \tag{62.2}$$

62.3 Canonical ensemble

Formula 62.3.1 (Partition function). The partition function for discrete systems is defined as

$$Z(T) := \sum_i g_i e^{-\beta E_i}. \tag{62.3}$$

The analogue for continuous systems is

$$Z(T) := \int \Omega(E, V, N) e^{-\beta E} dE. \tag{62.4}$$

Formula 62.3.2. Consider a system of N indistinguishable, noninteracting particles. Let ε_i be the energy associated with the i^{th} energy level and let g_i be its degeneracy. The probability p_i of finding a particle in the i^{th} energy level is given by

$$p_i = \frac{g_i e^{-\beta E_i}}{Z}. \tag{62.5}$$

Definition 62.3.3 (Helmholtz free energy). The Helmholtz free energy is defined as follows:

$$F := -k_B T \ln Z. \tag{62.6}$$

For the canonical ensemble it can be shown that this is equal to a Legendre transformation of the energy: $F = E - TS$.

One can also obtain the Helmholtz free energy as a different Legendre transform using the ideas of information theory (Chapter 44). There it was shown that the convex potentials associated to exponential families were related to the free energy. If Equation (44.26) is compared to the above one, it can be seen that

$$\psi = -\beta F.$$

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:

$$\begin{aligned} \eta &= \beta\frac{\partial\psi}{\partial\beta} - \psi \\ \overset{\text{def.}}{\iff} \quad -S &= -\beta U + \beta F \\ \iff \quad F &= U - TS_B, \end{aligned}$$

where the relation between the Boltzmann entropy S_B and the Shannon entropy S was used.

62.4 Grand canonical ensemble

Formula 62.4.1 (Grand canonical partition function). The partition function of the i^{th} energy level is given by

$$\mathcal{Z}_i := \sum_{n_k} e^{\beta n_k(\mu-\varepsilon_i)}. \tag{62.7}$$

The grand canonical partition function is given by

$$\mathcal{Z} := \prod_i \mathcal{Z}_i = \sum_{n_k,\varepsilon_i} e^{\beta n_k(\mu-\varepsilon_i)}. \tag{62.8}$$

Remark 62.4.2. In the case of fermions, i.e. $n_i \in \{0,1\}$, this formula reduces to

$$\mathcal{Z} = e^{\beta\mu N} Z, \tag{62.9}$$

with N the total particle number.

Definition 62.4.3 (Fugacity).

$$z := e^{\mu N} \tag{62.10}$$

Formula 62.4.4 (Quantum mechanics). For quantum-mechanical systems one can rewrite the partition function as follows:

$$\mathcal{Z} = \operatorname{tr}\exp\Big(-\beta\big(\hat{H} - \mu\hat{N}\big)\Big). \tag{62.11}$$

This reduces to the above expressions when working in the single-particle eigenbasis (this is only possible for free theories).

62.5 Energy

Theorem 62.5.1 (Virial theorem).

$$\langle E_{\text{kin}}\rangle = -\frac{1}{2}\sum_k \langle \vec{\boldsymbol{r}}_k \cdot \vec{\boldsymbol{F}}_k\rangle \tag{62.12}$$

Corollary 62.5.2. For potentials of the form $V = ar^{-n}$ this becomes

$$2\langle E_{\text{kin}}\rangle = -n\langle V\rangle. \tag{62.13}$$

Theorem 62.5.3 (Equipartition theorem). *Let q be a generalized coordinate.*

$$\left\langle q^k \frac{\partial H}{\partial q^l}\right\rangle = k_B T \delta_{kl} \tag{62.14}$$

Corollary 62.5.4. For quadratic Hamiltonians this can be rewritten using Euler's theorem for homogeneous functions 14.6.13:

$$\langle T\rangle = \frac{1}{2}k_B T. \tag{62.15}$$

62.6 Lattice systems

62.6.1 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 brain signals, social networks or forest growth.

Formula 62.6.1 (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, the Ising Hamiltonian has the following form:

$$H := -\sum_{\langle i,j\rangle\in\Lambda} J_{ij}S_iS_j - h\sum_{i\in\Lambda} S_i, \tag{62.16}$$

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 situations 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. $S_iS_j = 1$, but the orientation has no influence on the energy. However, when an external field is applied (even a very small one), the second term in the Hamiltonian induces a preferential orientation and the degeneracy is lifted.

When the temperature is finite, i.e. $T > 0$, thermal fluctuations will start to randomly flip spins and the equilibrium state will not coincide with the perfectly ordered state anymore. The question now becomes if there exists a critical temperature T_c at which the system changes between a disordered state and an ordered state. For one spatial dimension the solution was found by Ising:

[1] One can generalize this Hamiltonian to also allow for site-dependent external fields. However, this will not be of relevance here.

Property 62.6.2 ($d = 1$). For a one-dimensional lattice the Ising model does not undergo a phase transition at a finite temperature. However, at $T = 0$, a second-order phase transition is obtained.

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 $J_{ij} \equiv J$:

$$\begin{aligned} Z &= \sum_{\{S_1,\dots,S_L\}} \exp\left(-\sum_{i=1}^{L-1} JS_iS_{i+1} - hS_i\right) && (62.17)\\ &= \prod_{i=1}^{L-1} \sum_{S_i,S_{i+1}} \exp(-JS_iS_{i+1} - hS_i) && (62.18)\\ &= \prod_{i=1}^{L-1} P_{i,i+1} && (62.19)\\ &\equiv \mathrm{tr}(P^L). && (62.20) \end{aligned}$$

The matrix $P_{i,i+1}$ is given by

$$P_{i,i+1} := \begin{pmatrix} e^{\beta(J+h)} & e^{-\beta J} \\ e^{-\beta J} & e^{\beta(J-h)} \end{pmatrix} \tag{62.21}$$

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:

$$\lambda_1 = e^{\beta J}\cosh(\beta h) + \sqrt{e^{2\beta J}\sinh^2(\beta h) + e^{-2\beta J}}. \tag{62.22}$$

So, for $L \gg 1$, the free energy density becomes

$$f(T, J, h) = -\frac{k_BT}{L}\ln(Z) = -k_BT\left(e^{\beta J}\cosh(\beta h) + \sqrt{e^{2\beta J}\sinh^2(\beta h) + e^{-2\beta J}}\right).$$

?? FINISH ??

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 the one-dimensional setting.

62.7 Black-body radiation

Formula 62.7.1 (Planck's law).

$$B_\nu(\nu, T) = \frac{2h\nu^3}{c^2}\frac{1}{e^{\beta h\nu} - 1} \tag{62.23}$$

Formula 62.7.2 (Wien's displacement law).

$$\lambda_{\max}T = b, \tag{62.24}$$

where the constant $b = 2.897\,772\,9(17) \times 10^{-3}$ Km is called **Wien's displacement constant**.

Chapter 63

Material Physics

63.1 Crystals

Theorem 63.1.1 (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 63.1.2 (Zone). The collection of faces parallel to a given axis. The axis itself is called the **zone axis**.

63.1.1 Analytic representation

Definition 63.1.3 (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 63.1.4. Negative numbers are often written as $\overline{a}$ instead of $-a$.

Formula 63.1.5 (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

$$u = \begin{vmatrix} k_1 & l_1 \\ k_2 & l_2 \end{vmatrix} \qquad v = \begin{vmatrix} l_1 & h_1 \\ l_2 & h_2 \end{vmatrix} \qquad w = \begin{vmatrix} h_1 & k_1 \\ h_2 & k_2 \end{vmatrix}. \tag{63.1}$$

Theorem 63.1.6 (Hauy's law of rational indices). *The Miller indices of every natural face of a crystal will always have rational proportions.*

63.2 Symmetries

Definition 63.2.1 (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 63.2.2 (Rotational symmetry). Only $1, 2, 3, 4$ and 6-fold rotational symmetries can occur.

63.3 Crystal lattice

Formula 63.3.1. For an orthogonal crystal lattice, the distance between planes of the family $(h\ k\ l)$ is given by

$$d_{hkl} = \frac{1}{\sqrt{\left(\frac{h}{a}\right)^2 + \left(\frac{k}{b}\right)^2 + \left(\frac{l}{c}\right)^2}}. \tag{63.2}$$

63.3.1 Bravais lattice

Definition 63.3.2 (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 63.3.3 (Wigner-Seitz cell). The part of space consisting of all points closer to a given lattice point than to any other.

Theorem 63.3.4 (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.*

63.3.2 Reciprocal lattice

Formula 63.3.5 (Reciprocal basis vectors). The reciprocal lattice corresponding to a Bravais lattice with primitive basis $\{\vec{a}, \vec{b}, \vec{c}\}$ is defined by the following reciprocal basis vectors:

$$\vec{a}^* := 2\pi\frac{\vec{b}\times\vec{c}}{\vec{a}\cdot(\vec{b}\times\vec{c})}. \tag{63.3}$$

The vectors $\vec{b}^*$ and $\vec{c}^*$ are obtained by cyclic permutation of (a, b, c). These vectors satisfy the relations

$$\begin{aligned} \vec{a}\cdot\vec{a}^* &= 2\pi \\ \vec{b}\cdot\vec{b}^* &= 2\pi \\ \vec{c}\cdot\vec{c}^* &= 2\pi. \end{aligned} \tag{63.4}$$

Notation 63.3.6 (Reciprocal lattice vector). The reciprocal lattice vector $\vec{r}^*_{hkl}$ is defined as follows:

$$\vec{r}^*_{hkl} := h\vec{a}^* + k\vec{b}^* + l\vec{c}^*. \tag{63.5}$$

Property 63.3.7. The reciprocal lattice vector $\vec{r}^*_{hkl}$ has the following properties:

- $\vec{r}^*_{hkl}$ is perpendicular to the family of planes $(h\ k\ l)$ of the direct lattice, and
- $\|\vec{r}^*_{hkl}\| = \frac{2\pi n}{d_{hkl}}$.

63.4 Diffraction

63.4.1 Constructive interference

Formula 63.4.1 (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:

$$\begin{aligned} a(\cos\alpha - \cos\alpha_0) &= h\lambda \\ b(\cos\beta - \cos\beta_0) &= k\lambda \\ c(\cos\gamma - \cos\gamma_0) &= l\lambda \end{aligned} \tag{63.6}$$

If these conditions have been met, a diffracted beam of order hkl will be observed.

Remark 63.4.2. 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 63.4.3 (Vectorial Laue conditions). Let $\vec{\boldsymbol{k}}_0, \vec{\boldsymbol{k}}$ denote the wave vectors of the incident and diffracted beams respectively. The Laue conditions can be reformulated in the following way:

$$\vec{\boldsymbol{k}} - \vec{\boldsymbol{k}}_0 = \vec{\boldsymbol{r}}^*_{hkl}. \tag{63.7}$$

Formula 63.4.4 (Bragg's law). Another equivalent formulation of the Laue conditions is given by the following formula:

$$2d_{hkl}\sin\theta = n\lambda, \tag{63.8}$$

where

- λ is the wavelength of the incoming beam,
- θ is the **Bragg angle**, and
- d_{hkl} is the distance between neighbouring planes.

Remark 63.4.5. The angle between the incident and diffracted beams is 2θ.

Construction 63.4.6 (Ewald sphere). A simple construction to determine if Bragg diffraction will occur is the Ewald sphere. Put the origin of the reciprocal lattice at the tip of the incident wave vector $\vec{\boldsymbol{k}}_i$. Construct a sphere with radius $\frac{2\pi}{\lambda}$ centred on the start of $\vec{\boldsymbol{k}}_i$. All points on the sphere that coincide with a reciprocal lattice point satisfy the vectorial Laue condition 63.4.3. Therefore, Bragg diffraction will occur in the direction of all the intersections of the Ewald sphere and the reciprocal lattice.

63.4.2 Intensity of diffracted beams

Property 63.4.7 (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. These absences are called systematic extinctions.

Definition 63.4.8 (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 63.4.9 (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

$$F(hkl) = \sum_j f_j \exp\left[2\pi i(hx_j + ky_j + lz_j)\right], \tag{63.9}$$

where f_j is the atomic scattering factor of the j^{th} atom in the motive.

Example 63.4.10. 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 63.4.11 (Laue indices). Higher-order diffractions can be rewritten as a first-order diffraction in the following way:

$$2d_{nhnknl}\sin\theta=\lambda \qquad \text{with} \qquad d_{nhnknl}=\frac{d_{hkl}}{n}. \tag{63.10}$$

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.

63.5 Alloys

Property 63.5.1 (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.
4. The valence is the same.

63.6 Lattice defects

Definition 63.6.1 (Interstitial). An atom placed at a position that is not a lattice point.

Definition 63.6.2 (Vacancy). A lattice point where an atom is missing. This is also called a **Schottky defect**.

Formula 63.6.3 (Concentration of Schottky defects). Let N,n denote the number of lattice points and vacancies, respectively. The following relation gives the temperature dependence of Schottky defects:

$$\frac{n}{n+N}=e^{-E_v/kT}, \tag{63.11}$$

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

$$\Omega=\frac{(N+n)!}{N!\,n!}, \tag{63.12}$$

where the fact that the removal of n particles creates n more lattice points at the surface was used. Using Boltzmann's entropy formula (62.1) and Stirling's formula (14.39) one obtains

$$S(N,n) = k\ln\Omega = k\big[(N+n)\ln(N+n) - n\ln n - N\ln N\big]. \tag{63.13}$$

Using Definition 62.2.1 of the temperature as the derivative of the energy gives

$$\frac{1}{T} = \left(\frac{\partial S}{\partial E}\right)_{N,V} = \frac{dS}{dn}\frac{dn}{dE} = \frac{k}{E_v}\ln\frac{N+n}{n} \tag{63.14}$$

or

$$\frac{n}{N+n} = \exp\left(-\frac{E_v}{kT}\right). \tag{63.15}$$

The density of Frenkel pairs can be derived analogously. □

Definition 63.6.4 (Frenkel pair). An atom displaced from a lattice point to an interstitial location, thereby creating a vacancy-interstitial pair.

Formula 63.6.5 (Concentration of Frenkel pairs). Let n_i denote the number of atoms displaced from the bulk of the lattice to any of N_i possible interstitial positions, thereby creating n_i vacancies. The following relation holds:

$$\frac{n_i}{\sqrt{NN_i}} = e^{-E_{\text{fr}}/2kT}, \tag{63.16}$$

where E_{fr} denotes the energy needed to create a Frenkel pair.

Remark 63.6.6. In compounds, the number of vacancies can be much higher than in mono-atomic lattices.

Remark 63.6.7. The existence of these defects creates the possibility of diffusion.

63.7 Electrical properties

63.7.1 Charge carriers

Formula 63.7.1 (Conductivity). Definition 50.1.11 can be modified to account for both positive and negative charge carriers:

$$\sigma := n_n q_n \mu_n + n_p q_p \mu_p. \tag{63.17}$$

Remark. The difference between the concentration of positive and negative charge carriers can differ by orders of magnitude (up to 20) across different materials.

63.7.2 Band structure

Definition 63.7.2 (Valence band). The energy band corresponding to the outermost (partially) filled atomic orbital.

Definition 63.7.3 (Conduction band). The first unfilled energy band.

Definition 63.7.4 (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 63.7.5 (Fermi level). The energy level having a 50% chance of being occupied at thermodynamic equilibrium.

[1]For a basic derivation see [55].

Formula 63.7.6 (Fermi function). The following distribution gives the probability of a state with energy E_i being occupied by an electron:

$$f(E_i) = \frac{1}{e^{\beta(E_i - E_f)} + 1}, \tag{63.18}$$

where E_f is the Fermi level as defined above.

Formula 63.7.7. Let n denote the charge carrier density as before. The following temperature dependence can be found:

$$n \sim e^{-\beta E_g/2}, \tag{63.19}$$

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 63.7.8 (Doping). Intentionally introducing impurities to modify the (electrical) properties.

Definition 63.7.9 (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 63.7.10 (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**.

63.7.3 Ferroelectricity

Some materials can exhibit phase transitions between a paraelectric and ferroelectric state. Paraelectric materials have the property that the polarization $\vec{\boldsymbol{P}}$ and the electric field $\vec{\boldsymbol{E}}$ 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 63.7.11. 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 63.7.12 (Saturation polarization). The maximum polarization obtained by a ferroelectric material. It it obtained when the *domain formation* reaches a maximum.

Definition 63.7.13 (Remanent polarization). The residual polarization of the material when the external electric field is turned off.

Definition 63.7.14 (Coercive field). The electric field needed to cancel out the *remanent* polarization.

Definition 63.7.15 (Piezoelectricity). Materials that obtain a polarization when exposed to mechanical stress are called piezoelectric materials.

Remark 63.7.16. 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 63.7.17 (Transducer). A device that converts electrical energy to mechanical energy (and vice versa).

63.8 Magnetic properties

Definition 63.8.1 (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 63.8.2. All materials exhibit diamagnetic behaviour.

Definition 63.8.3 (Paramagnetism). The susceptibility is small, positive and inversely proportional to the temperature.

Definition 63.8.4 (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.

63.8.1 Paramagnetism

Formula 63.8.5 (Curie's law). If the interactions between the particles can be neglected, the following relation is obtained:

$$\chi = \frac{C}{T}. \tag{63.20}$$

Materials that satisfy this law are called **ideal paramagnetics**.

Formula 63.8.6 (Curie-Weiss law). If the interactions between particles cannot be neglected, a correction of Curie's law is obtained:

$$\chi = \frac{C}{T-\theta}, \tag{63.21}$$

where $\theta = CN_W$ with N_W the **Weiss-constant**. This deviation of Curie's law is due to the intermolecular interactions that induce an internal magnetic field $H_m = N_W M$.

Formula 63.8.7 (Brillouin function).

$$B_J(y) := \frac{2J+1}{2J}\coth\left(\frac{2J+1}{2J}y\right) - \frac{1}{2J}\coth\left(\frac{y}{2J}\right), \tag{63.22}$$

where $y := \frac{g\mu_B JB}{k_B T}$.

Remark 63.8.8. Because $\lim_{y\to\infty}\coth(y) = 1$, one obtains:

$$T \longrightarrow 0 \implies M = \lim_{y\to\infty} Ng\mu_B JB_J(y) = Ng\mu_B J. \tag{63.23}$$

This value is called the **absolute saturation magnetization**.

63.8.2 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 63.8.9 (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 63.8.10 (Bloch walls). A wall between two magnetic domains.

Definition 63.8.11 (Ferromagnetic Curie temperature). Above this temperature the material loses its ferromagnetic properties and it becomes a paramagnetic material following the Curie-Weiss law.

Remark 63.8.12. 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.

63.8.3 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 rise.

Definition 63.8.13 (Néel temperature). At the Néel temperature, the susceptibility will reach a maximum. Above this temperature ($T > T_N$) the material will become paramagnetic, satisfying the following formula:

$$\chi = \frac{C}{T+\theta}. \tag{63.24}$$

This resembles a generalization of the Curie-Weiss law with a negative and, therefore, virtual critical temperature.

63.8.4 Ferrimagnetism

Materials that are not completely ferromagnetic nor antiferromagnetic, due to an unbalance 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 63.8.14 (Néel hyperbola). Above the Néel temperature it is possible to define a susceptibility given by

$$\frac{1}{\chi} = \frac{T}{C} - \frac{1}{\chi_0} - \frac{\sigma}{T-\theta'}. \tag{63.25}$$

[2]This can be seen in Equation (50.24): $M = \chi H$. The susceptibility should be infinite.

Chapter 64

Tensor Networks

The main references for the sections on *topological order* are [92, 93].

64.1 Matrix Product States

64.1.1 Finite-dimensional lattices

Definition 64.1.1 (Matrix product state). Let $\mathcal{H}_n$ be the local Hilbert spaces of dimension d_n where $n \in \{1, \ldots, N\}$. 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

$$|\psi\rangle = \sum_{\{i_k\}} \mathrm{tr}\left(\prod_n^N A^{i_n}(n)\right)|i_1 \ldots i_N\rangle. \tag{64.1}$$

For each lattice site n 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 64.1.2 (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 64.1.3 (MPS projector). Consider an MPS defined by tensors $\{A(n)\}_{n\leq N}$. The associated MPS projector is defined as

$$\mathcal{P}(A) := \sum_{i,\alpha,\beta} A^i_{\alpha\beta}(n)|i\rangle\langle\alpha\beta|. \tag{64.2}$$

Formula 64.1.4 (Transfer operator). Given MPS tensors $\{A(n)\}_{n\leq N}$, one can define a transfer operator:

$$\mathbb{E}(n) := \sum_{i=1}^{d_n} A^i(n) \otimes \overline{A^i}(n). \tag{64.3}$$

Formula 64.1.5 (Superoperator). More generally, one can define for every local observable $\hat{O}_n$ a superoperator in $\mathcal{L}(\mathbb{C}^{D_n} \otimes \overline{\mathbb{C}^{D_n}}, \mathbb{C}^{D_{n-1}} \otimes \overline{\mathbb{C}^{D_{n-1}}})$:

$$\mathbb{E}_{O_n}(n) := \sum_{i,i'=1}^{d_n} \langle i|\hat{O}_n|i'\rangle A^{i'}(n) \otimes \overline{A^i}(n). \tag{64.4}$$

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 by

$$\mathbb{E}_B^A(n) := \sum_{i=1}^{d_n} A^i(n) \otimes \overline{B^i}(n). \tag{64.5}$$

Example 64.1.6. Using these definitions one can rewrite the formulas for expectation values more efficiently. Given a product operator $\hat{O} = \bigotimes_i^N \hat{O}_i$ one finds that

$$\langle\psi[A]|\hat{O}|\psi[A]\rangle = \mathrm{tr}\Big(\mathbb{E}_{O_1}(1)\cdots\mathbb{E}_{O_N}(N)\Big). \tag{64.6}$$

Formula 64.1.7. Associated to the superoperator $\mathbb{E}_O(n)$ one can define a virtual superoperator:

$$\mathcal{E}_{O_n}^{(n)}(\phi) = \sum_{i,i'=1}^{d_n} \langle s|\hat{O}_n|s'\rangle A^{i'}(n)\phi A^i(n)^\dagger \tag{64.7}$$

$$\tilde{\mathcal{E}}_{O_n}^{(n)}(\phi) = \sum_{i,i'=1}^{d_n} \langle s|\hat{O}_n|s'\rangle A^i(n)^\dagger\sigma A^{i'}(n), \tag{64.8}$$

where $\phi \in \mathcal{L}(\mathbb{C}^{D_n}), \sigma \in \mathcal{L}(\mathbb{C}^{D_{n-1}})$.

Property 64.1.8. The map $\mathcal{E}_{\mathbb{1}}^{(n)}$ associated to the transfer operator is a CP map 24.1.14 and the associated Kraus operators 60.3.4 are the MPS matrices $A^i(n)$.

64.1.2 Injectivity

For translation-invariant MPS one can use an easier definition:

Alternative Definition 64.1.9 (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.

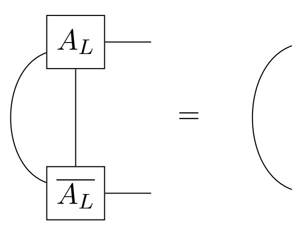
64.1.3 Gauge freedom and canonical forms

Property 64.1.10 (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 64.1.11. 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 64.1.12 (Left canonical form[1]). This form is specified by the following property:



(64.9)

[1]Also called the **left isometric form**, **left orthogonal form** or just **left gauge**.

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* 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

$$A_L(n) := L(n)A(n)L^{-1}(n+1). \tag{64.10}$$

In a similar manner one can construct the right orthogonal form A_R.

Method 64.1.13 (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* [94]. 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 $U^{[1]}SV^{[2,\ldots]}$. Recursively applying this procedure to the product of the singular values S and the right unitary V will result in the canonical form.

Construction 64.1.14 (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

$$C(n) = L(n)R(n). \tag{64.11}$$

64.1.4 Translation-invariant states

Definition 64.1.15 (Uniform MPS). By choosing all MPS tensors $A(n) = B$ for a given tensor B, one obtains a translation-invariant (TI) state, i.e. a state invariant under a shift of the index n. These MPSs form the variational class of uniform MPS.

Remark 64.1.16 (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).

64.2 Matrix product operators

Definition 64.2.1 (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:

$$\hat{O} = \sum_{\{i_k,j_l\}} \mathrm{tr}\Big(\prod_{m,n=1}^{N} O^{i_m,j_n}(n)\Big)|i_1\cdots i_N\rangle\langle j_1\cdots j_n|. \tag{64.12}$$

In terms of a basis $\{\hat{O}_i\}$ for the space of local operators this becomes:

$$\hat{O} = \sum_{\{i_k\}} \mathrm{tr}\Big(\prod_{n}^{N} A^{i_n}(n)\Big)\hat{O}_{i_n}. \tag{64.13}$$

Method 64.2.2 (Local Hamiltonian to MPO). Given a local Hamiltonian $\hat{H} = \sum_i \hat{H}^{(i)}$ one can build an MPO that generates this Hamiltonian[3]:

$$\hat{H} := \sum_{\{i_k,j_l\}} \mathrm{tr}\Big(\prod_{m,n=1}^{N} A^{i_m,j_n}(n)\Big)|i_1\cdots i_N\rangle\langle j_1\cdots j_n|. \tag{64.14}$$

[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 obtain this MPO form 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 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 64.2.3. Consider a 2-site Hamiltonian of the form

$$\hat{A}\otimes\hat{B}\otimes\mathbb{1}\otimes\cdots+\mathbb{1}\otimes\hat{A}\otimes\hat{B}\otimes\mathbb{1}\otimes\cdots+\cdots.$$

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 $\hat{A}\otimes\hat{B}$.

The transition rules for this automaton are then given by the following list:

- $1\rightarrow 2$: $\mathbb{1}$,
- $1\rightarrow 2$: $\hat{B}$,
- $2\rightarrow 3$: $\hat{A}$, and
- $3\rightarrow 3$: $\mathbb{1}$.

What is useful here is that this set of transition rules can be turned into a matrix:

$$T=\begin{pmatrix}\mathbb{1}&\hat{B}&0\\0&0&\hat{A}\\0&0&\mathbb{1}\end{pmatrix}.$$

The MPO is then obtained by setting the MPO matrix A equal to T at every site.

64.2.1 MPO-injectivity

Definition 64.2.4 (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

$$A_\Lambda:(\mathbb{C}^D)^{\otimes|\Lambda|}\rightarrow(\mathbb{C}^d)^{\otimes|\partial\Lambda|}$$

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)

$$M:\mathbb{C}^D\otimes\mathbb{C}^m\rightarrow\mathbb{C}^D\otimes\mathbb{C}^m$$

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 Figure 64.1, where the tensors M are given by crossings of black and red lines.)

Axiom 64.1 (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 Figure 64.2.

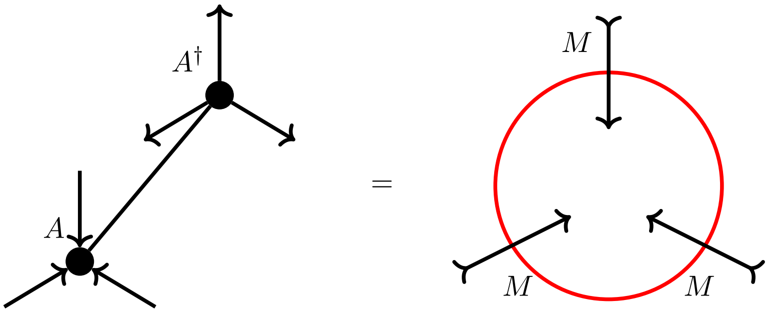


Figure 64.1: MPO-injective PEPS.

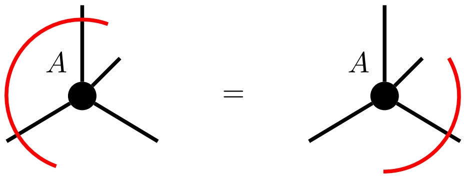


Figure 64.2: Pulling-through condition.

64.2.2 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 64.2 (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 Figure 64.3.

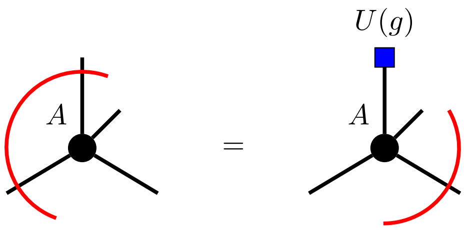


Figure 64.3: Pulling-through condition for SPT phases.

Part X

Relativity & Quantum Field Theory

Chapter 65

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 [49, 53].

65.1 Lorentz transformations

Notation 65.1.1. In the context of special relativity it is often useful to introduce the following quantities:

$$\beta := \frac{v}{c} \tag{65.1}$$

$$\gamma := \frac{1}{\sqrt{1-\beta^2}}. \tag{65.2}$$

The latter quantity is called the **Lorentz factor**.

Formula 65.1.2 (Lorentz transformations). Let $\mathbf{V}$ be a 4-vector. A Lorentz boost along the x^1-axis is given by the following transformation:

$$V'^0 = \gamma\big(V^0 - \beta V^1\big) \tag{65.3}$$

$$V'^1 = \gamma\big(V^1 - \beta V^0\big) \tag{65.4}$$

$$V'^2 = V^2 \tag{65.5}$$

$$V'^3 = V^3. \tag{65.6}$$

Remark 65.1.3. Putting $c = \infty$ in the previous formulas recovers the Galilei transformations from classical mechanics (cf. Inönü-Wigner contractions 30.5.1).

65.2 Energy and momentum

Formula 65.2.1 (4-velocity). In analogy to the definition of velocity in classical mechanics, the 4-velocity is defined as follows:

$$\mathbf{U} := \left(\frac{dx^0}{d\tau}, \frac{dx^1}{d\tau}, \frac{dx^2}{d\tau}, \frac{dx^3}{d\tau}\right). \tag{65.7}$$

By applying the formulas for proper time and time dilatation we obtain:

$$\mathbf{U} = (\gamma c, \gamma\vec{\boldsymbol{u}}). \tag{65.8}$$

Formula 65.2.2 (4-momentum). The 4-momentum is defined as follows:

$$\mathbf{p} = m_0\mathbf{U}, \tag{65.9}$$

or, after defining $E := cp^0$:

$$\mathbf{p} = \left(\frac{E}{c}, \gamma m_0\vec{\boldsymbol{u}}\right). \tag{65.10}$$

Definition 65.2.3 (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 $\vec{\boldsymbol{p}} = m\vec{\boldsymbol{u}}$ for the 3-momentum can be generalized to 4-momenta $\mathbf{p}$.

Formula 65.2.4 (Relativistic energy relation).

$$E^2 = p^2c^2 + m^2c^4 \tag{65.11}$$

This formula is often called the **Einstein relation**.

65.3 Action principle

The main guiding principles for writing down a relativistic action for a point particle are locality and geometry. 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 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:

$$S_{\text{point}} \sim \int_\gamma ds. \tag{65.12}$$

To get the units right, one should multiply by suitable Lorentz-invariant constants:

$$S_{\text{point}} := mc\int_\gamma ds. \tag{65.13}$$

By reparametrization-invariance one can choose a specific time-coordinate, e.g. $\tau = ct$. In this coordinate system the action becomes

$$S_{\text{point}} = mc\int_\gamma \sqrt{1 - \frac{v^2}{c^2}}cdt, \tag{65.14}$$

with v the speed of the particle. The Lagrangian density can be Taylor expanded as

$$L_{\text{point}} = mc^2 - \frac{1}{2}mv^2 + \cdots, \tag{65.15}$$

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 65.2.4). 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**:

$$x^\pm := x^0 \pm x^1. \tag{65.16}$$

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):

$$S_{\text{point}} := \frac{1}{2}\int_\gamma \left[\eta^{-1}\left(\frac{dx^\mu}{d\tau}\right)^2 - m^2\eta\right] d\tau. \tag{65.17}$$

The equation of motion for η is algebraic and gives

$$\eta = \frac{1}{m}\sqrt{-g_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau}}. \tag{65.18}$$

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":

$$\int_\gamma g_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\mu}{d\tau} + m^2 \longrightarrow \int_\gamma \sqrt{h}\left(hg_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\mu}{d\tau} + m^2\right). \tag{65.19}$$

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 66

General Relativity

References for this chapter are [30, 31]. See Chapter 34 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 33.8.

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).

66.1 Causal structure

Definition 66.1.1 (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:

$$g_p(v, v) = 0. \tag{66.1}$$

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.

Definition 66.1.2 (Time-orientability). A Lorentzian manifold is said to be time-orientable if there exists a nowhere-vanishing, timelike vector field. It should be noted 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 66.1.3 (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 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 66.1.4 (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 which is equal to its causal closure.

[1]For a mostly-minuses signature one should interchange these definitions.

Definition 66.1.5 (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 $p, q \in M$.

Definition 66.1.6 (Stationary spacetime). A spacetime (M, g) is called 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, a spacetime is stationary if one can find a coordinate system for which the metric coefficients are time-independent.

66.2 Einstein field equations

Formula 66.2.1 (Einstein field equations). The Einstein field equations without a cosmological constant Λ read as follows (all fundamental constants are shown for completeness):

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}, \tag{66.2}$$

where $G_{\mu\nu}$ is the Einstein tensor (34.37) and $T_{\mu\nu}$ is the stress-energy tensor (67.13).

By taking the trace of both sides one obtains $T = -R$ and, hence, the Einstein field equations can be rewritten as

$$R_{\mu\nu} = \widehat{T}_{\mu\nu}, \tag{66.3}$$

where $\widehat{T}_{\mu\nu} := T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T$ is the **reduced stress-energy tensor**.

Formula 66.2.2 (Einstein-Hilbert action). The (vacuum) field equations can be obtained by applying the variational principle to the following action:

$$S_{\mathrm{EH}}[g_{\mu\nu}] := \int_M \sqrt{-g} R. \tag{66.4}$$

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]

$$S_{\mathrm{GHY}}[g_{\mu\nu}] := \oint_{\partial M} \epsilon \sqrt{h} K, \tag{66.5}$$

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.

66.3 Black holes

Formula 66.3.1 (Schwarzschild metric).

$$ds^2 := \left(1 - \frac{R_s}{r}\right) c^2 dt^2 - \left(1 - \frac{R_s}{r}\right)^{-1} dr^2 - r^2 d\Omega^2, \tag{66.6}$$

where R_s is the Schwarzschild radius

$$R_s := \frac{2GM}{c^2}. \tag{66.7}$$

[2] Einstein had in fact already introduced a variant, the $\Gamma\Gamma$-*Lagrangian*.

Theorem 66.3.2 (Birkhoff). *The Schwarzschild metric is the unique solution of the vacuum field equation under the additional constraints of asymptotic flatness and staticity.*

Formula 66.3.3 (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:

$$ds^2 := \left(1 - \frac{2GM}{r} + \frac{GQ^2}{4\pi r^2}\right)c^2dt^2 - \left(1 - \frac{2GM}{r} + \frac{GQ^2}{4\pi r^2}\right)^{-1}dr^2 - r^2d\Omega^2. \tag{66.8}$$

Remark 66.3.4. The electric field generated by a Reissner-Nordström black hole is given by

$$E^r = \frac{Q}{4\pi r^2}. \tag{66.9}$$

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 ??

66.4 Conserved quantities

Before raising any hope that conserved quantities are to be found everywhere in GR, the following result is given:

Property 66.4.1. By Noether's third theorem 38.2.9, there exist no proper (local) conservation laws such as those of momentum and energy in classical mechanics, because the translation group is a subgroup of the infinite-dimensional symmetry group $\mathrm{Diff}(M)$.

66.5 Tetradic formulation

We will start 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. By the Erlangen program it is known that the Minkowski spacetime M^4 can be described as the coset space $\mathrm{ISO}(3,1)/\mathrm{SO}(3,1)$. The natural generalization is given by a Cartan geometry with model geometry $(\mathfrak{iso}(3,1), \mathfrak{so}(3,1))$.

Property 66.5.1 (Cartan connection). This way a $\mathrm{SO}(3,1)$-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 $\nabla + \mathbf{e}$ where

- ∇ defines a $\mathfrak{so}(3,1)$-valued principal connection, and
- $\mathbf{e}$ defines a $\mathcal{M}^4$-valued solder form.

The principal connection ∇ is called the **spin connection** and $\mathbf{e}$ is called the **vierbein** or **tetrad**. These objects are well-known in general relativity. The connection ∇ is the ordinary Levi-Civita connection associated to the Lorentzian manifold M (in case of vanishing torsion) and $\mathbf{e}$ gives the isometry between local (flat) Minkowski coordinates and "global" coordinates:

$$g := \mathbf{e}^*\eta \tag{66.10}$$

or, locally,

$$g_{\mu\nu} = e^i_\mu e^j_\nu \eta_{ij}. \tag{66.11}$$

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:

$$F^i{}_{j\mu\nu} := e^i_\rho e^\sigma_j R^\rho{}_{\sigma\mu\nu}, \tag{66.12}$$

where $R^\rho{}_{\sigma\mu\nu}$ is the ordinary Riemann curvature tensor. The Einstein-Hilbert action is then equivalent[3] to the following Yang-Mills-like action:

Formula 66.5.2 (Palatini action).

$$S[e,\nabla] := \int_M \mathbf{e}\wedge\mathbf{e}\wedge *F. \tag{66.13}$$

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 same action but only as a functional of the tetrad field, one obtains the **second order formulation** of gravity.[4]

Variation of the Palatini action gives the following EOM:

- $\delta\nabla$: $T(\mathbf{e}) = 0$ or, equivalently, $\nabla(\mathbf{e}) \equiv \nabla$, i.e. the torsion vanishes and the connection ∇ is on-shell equal to the Levi-Civita connection on M.
- $\delta\mathbf{e}$: 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 66.5.3 (Plebanski form).

$$\Sigma := \mathbf{e}\wedge\mathbf{e} \tag{66.14}$$

Because of its internal antisymmetric Lorentz indices, one can interpret this object as a $\mathfrak{so}(3,1)$-valued two-form.

As was the case for 4D Yang-Mills theory, one can introduce a topological term that leaves the EOM invariant (up to boundary terms):

Definition 66.5.4 (Holst action[5]).

$$\begin{aligned} S[\mathbf{e},\nabla] :&= \int_M \mathbf{e}\wedge\mathbf{e}\wedge *F + \frac{1}{\gamma}\int_M \mathbf{e}\wedge\mathbf{e}\wedge F \\ &= \int_M \left(*\mathbf{e}\wedge\mathbf{e} + \frac{1}{\gamma}\mathbf{e}\wedge\mathbf{e}\right)\wedge F. \end{aligned} \tag{66.15}$$

The coupling constant γ is called the **Barbero-Immirzi** constant.

?? COMPLETE ??

[3] At least in the case of pure gravity [31].

[4] 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.

[5] Holst was actually the second author to include this term.

66.6 Spinning particle

In this section, the action of spinning particle is written down in a supersymmetric way [69] and it is explained how the existence of spin structures on a manifold (Section 34.3) is related to a global anomaly in the quantization of the theory [68].

Consider the following action:

$$S_{\text{spin}}[x,\psi] := \frac{1}{2}\int d\tau g_{\mu\nu}(\dot{x}^\mu\dot{x}^\nu + i\psi^\mu\nabla_{\dot{x}}\psi^\nu) = \frac{1}{2}\int d\tau g_{\mu\nu}\Big(\dot{x}^\mu\dot{x}^\nu + i\psi^\mu\dot{\psi}^\nu\Big), \tag{66.16}$$

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, first-order in time derivatives, in the Grassmann variables that can be written down.

This action possesses a particular symmetry:

$$\delta x^\mu := i\psi^\mu\varepsilon \tag{66.17}$$

$$\delta\psi^\mu := (\dot{x}^\mu - i\omega_\kappa{}^\mu{}_\nu\psi^\kappa\psi^\nu)\varepsilon, \tag{66.18}$$

where ε is a Grassmann-odd constant and ω is the (local) spin-connection one-form. Although this theory is phsyically not supersymmetric, it does possess a "worldline supersymmetry". The associated Noether charge is

$$Q := \psi_\mu\dot{x}^\mu. \tag{66.19}$$

After canonical quantization, the Grassmann variables ψ^μ satisfy:

$$\{\psi_\mu,\psi_\nu\}_+ = g_{\mu\nu}, \tag{66.20}$$

i.e. they can be interpreted as dynamical gamma matrices (56.1). 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

$$p_\mu = g_{\mu\nu}\dot{x}^\nu + \frac{i}{4}\omega_{\mu\nu\kappa}[\psi^\nu,\psi^\kappa]. \tag{66.21}$$

In the typical representations of quantum mechanics the relation $p_\mu = -i\partial_\mu$ holds, so the above relation can be rewritten as

$$g_{\mu\nu}\dot{x}^\nu = -i\nabla_\mu := \partial_\mu + \frac{1}{4}\omega_{\mu\nu\kappa}[\psi^\nu,\psi^\kappa], \tag{66.22}$$

i.e. the spinorial covariant derivative (up to a factor $-i$). The supersymmetry Noether charge is thus given by the Dirac operator.

?? FINISH (not clear right now) ??

Chapter 67

Classical Field Theory ♣

Rigorous definitions and statements about the mathematical concepts used in this chapter can be found in Chapters 31, 32, 34 and 38.

67.1 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 fibre bundle $E \to M$ (often the tangent bundle TM). In general, a Lagrangian (function) or **action** is a function $L : \Gamma(E) \to \mathbb{R}$ from the space of (compactly supported) sections of a vector bundle to the real numbers. This is often given by local functionals for a local Lagrangian 38.4.2:

$$L : \Gamma_c(E) \to \mathbb{R} : \phi \mapsto \int_M (j^\infty\phi)^*\mathcal{L}\,\mathrm{Vol}\,. \tag{67.1}$$

In the remainder of this chapter the functional L and the density $\mathcal{L}$ will both be denoted by L for notational simplicity.

Associated to this manifold 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, while on the other hand one has a differential δ along the jet fibres, induced by the variation of fields. 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 (38.49) is defined as follows (partial derivatives are denoted by subscripted commas, e.g. $\partial_\mu\partial_\nu\phi \equiv \phi_{,\mu\nu}$):

$$\frac{\delta L}{\delta\phi} := \frac{\partial L}{\partial\phi} - \partial_\mu\left(\frac{\partial L}{\partial\phi_{,\mu}}\right) + \partial_\mu\partial_\nu\left(\frac{\partial L}{\partial\phi_{,\mu\nu}}\right) - \cdots\,. \tag{67.2}$$

By comparing this formula to the formula for the variation of the Lagrangian density, one obtains the first variational formula (38.83):

$$\delta L = \frac{\delta L}{\delta\phi^I}\delta\phi^I - d\Theta[\phi]. \tag{67.3}$$

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:

$$\omega[\phi] = \delta\Theta[\phi]. \tag{67.4}$$

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 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 a $(n-2,1)$-form $k_\xi[\phi]$ such that $\omega \approx dk[\phi]$. (More details can be found in [41].)

67.1.1 Noether's theorem for fields

In the context of field theory the Lagrangian density is often denoted by $\mathcal{L}$ instead of L. This convention will be adopted here as well.

Theorem 67.1.1 (Noether's first theorem). *Consider a general field transformation*

$$\phi \longrightarrow \phi + \alpha\delta\phi, \tag{67.5}$$

where α is an infinitesimal quantity and $\delta\phi$ is a small variation. In case of a symmetry one obtains a conservation law of the following form:

$$\partial_\mu\left(\frac{\partial\mathcal{L}}{\partial\phi_{,\mu}}\delta\phi - \mathcal{J}^\mu\right) = 0. \tag{67.6}$$

The factor between parentheses can be interpreted as a conserved current $j^\mu(x)$.

Proof. The general transformation rule for the Lagrangian is

$$\mathcal{L} \longrightarrow \mathcal{L} + \alpha\delta\mathcal{L}. \tag{67.7}$$

To have a symmetry, i.e. to keep the action invariant, the deformation factor has to be a 4-divergence:

$$\mathcal{L} \longrightarrow \mathcal{L} + \alpha\partial_\mu\mathcal{J}^\mu. \tag{67.8}$$

To obtain the conservation law (67.6), the Lagrangian is varied explicitly:

$$\begin{aligned}
\delta\mathcal{L} &= \frac{\partial\mathcal{L}}{\partial\phi}\delta\phi + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta(\partial_\mu\phi)\\
&= \frac{\partial\mathcal{L}}{\partial\phi}\delta\phi + \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta\phi\right) - \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\right)\delta\phi\\
&= \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta\phi\right) + \left[\frac{\partial\mathcal{L}}{\partial\phi} - \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\right]\delta\phi.
\end{aligned}$$

The second term vanishes due to the Euler-Lagrange equation (48.54). Combining these formulas gives

$$\partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta\phi\right) - \partial_\mu\mathcal{J}^\mu(x) = 0. \tag{67.9}$$

From this equation one can conclude that the current

$$j^\mu(x) = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta\phi - \mathcal{J}^\mu(x) \tag{67.10}$$

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**):

$$Q[\Sigma] := \int_\Sigma j^0 d^3x, \tag{67.11}$$

where Σ is a spacelike hypersurface. The conservation law can then simply be restated as

$$\frac{dQ}{dt} = 0.$$

Definition 67.1.2 (Stress-energy tensor). Consider the translation of a scalar field:

$$\phi(x) \longrightarrow \phi(x+a) = \phi(x) + a^\mu \partial_\mu \phi(x).$$

Because the Lagrangian is a scalar quantity, it transforms in the same way as the fields:

$$\mathcal{L} \longrightarrow \mathcal{L} + a^\mu \partial_\mu \mathcal{L} = \mathcal{L} + a^\nu \partial_\mu (\mathcal{L} \delta^\mu{}_\nu). \tag{67.12}$$

On a D-dimensional manifold this leads to the existence of D conserved currents. These can be used to define the stress-energy tensor:

$$T^\mu_\nu = \frac{\partial \mathcal{L}}{\partial \phi_{,\mu}} \partial_\nu \phi - \mathcal{L} \delta^\mu{}_\nu. \tag{67.13}$$

67.1.2 Gauge algebra

In this section the gauge symmetries of a local action S, i.e. an action

$$S[\phi] := \int L(\phi^I, \partial_\mu \phi^I, \dots, x^\mu) dx$$

where $L : J^\infty(E) \to \mathbb{R}$ is the Lagrangian, is considered. A **gauge transformation** of this action is a coordinate transformation that depends arbitrarily on M but leaves the action invariant. The most general form of such a transformation is

$$\delta_\varepsilon \phi^I = \overline{R}^I_{(0),\alpha} \varepsilon^\alpha + \overline{R}^{I,\mu}_{(1),\alpha} \partial_\mu \varepsilon^\alpha + \cdots + \overline{R}^{I,\mu_1\dots\mu_s}_{(s),\alpha} \partial_{\mu_1\dots\mu_s} \varepsilon^\alpha \equiv \overline{R}^I_\alpha \varepsilon^\alpha, \tag{67.14}$$

where the coefficients $\overline{R}_{(i)}$ are arbitrary functions of the coordinates and in the last step a new shorthand was introduced where the summation over j also includes an integral over x (the **DeWitt convention**):

$$\begin{aligned} R^I_\alpha \varepsilon^\alpha &:= \int R^I_\alpha(x, x') \varepsilon^\alpha(x') dx' \qquad (67.15) \\ &= \int \sum_j \Big(\overline{R}^i_{(0),j}(x) \delta(x - x') + \overline{R}^i_{(1),j}(x) \delta'(x - x') + \overline{R}^i_{(2),j}(x) \delta^{(2)}(x - x') + \cdots \Big) \varepsilon^j(x') dx'. \end{aligned}$$

Invariance of the action implies that

$$\delta_\varepsilon S = \frac{\delta S}{\delta \phi^I} \delta_\varepsilon \phi^I = \frac{\delta S}{\delta \phi^I} R^I_\alpha \varepsilon^\alpha = 0. \tag{67.16}$$

Beause this should hold for every value of the transformation parameters ε^α, one immediately obtains the variational Noether identities:

Property 67.1.3 (Noether identities). If a local action is invariant under the transformation (67.14), then

$$\frac{\delta S}{\delta \phi^I} R^I_\alpha = 0 \tag{67.17}$$

for all "indices" α. In contrast to Noether's theorem 48.2.9, these identities do not imply conserved quantities. Instead they show that the equations of motion are not independent.

The structure of the infinitesimal gauge transformations is easily seen to be that of a (real) Lie algebra, whilst that of finite (exponentiated) transformations is a Lie group. However, the gauge algebra $\overline{\mathcal{G}}$ 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 *Freedman* and *Van Proeyen* [47]:

Definition 67.1.4 (Trivial gauge symmetry). All transformations of the form

$$\delta_\varepsilon \phi^I = \varepsilon^{IJ} \frac{\delta S}{\delta \phi^J}, \tag{67.18}$$

where ε^{IJ} is antisymmetric, are physically irrelevant since they are not generated by constraints. 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.

In fact one can show that any gauge transformation, satisfying suitable conditions, that vanishes on-shell is equal to some trivial transformation. ?? EXPLAIN (see HENNEAUX and TEITELBOIM) ??

A further problem with the gauge algebra is that independent transformations might lead to dependent Noether identities wich implies that there is still some redundancy. To fix this one defines the following minimal set:

Definition 67.1.5 (Generating set). A generating set[1] of the gauge algebra is a set of transformations $\delta_\varepsilon \phi^I = R^I_\alpha \varepsilon^\alpha$ such that every gauge transformation can be written as follows:

$$\delta \phi^I = R^I_\alpha \mu^\alpha + M^{IJ} \frac{\delta S}{\delta \phi^J}, \tag{67.19}$$

where $M^{IJ} = -M^{JI}$. 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

$$R^J_\alpha \frac{\delta R^I_\beta}{\delta \phi^J} - R^I_\beta \frac{\delta R^J_\alpha}{\delta \phi^I} = C^\gamma_{\alpha\beta} R^I_\gamma + M^{IJ}_{\alpha\beta} \frac{\delta S}{\delta \phi^J}, \tag{67.20}$$

where $M^{IJ}_{\alpha\beta} = -M^{JI}_{\alpha\beta}$. 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) and otherwise it is said to be **open**. A generating set is said to be **irreducible** if there exist no nontrivial combinations of elements:

$$R^I_\alpha \varepsilon^\alpha = M^{IJ} \frac{\delta S}{\delta \phi^J} \implies \varepsilon^\alpha = N^{\alpha I} \frac{\delta S}{\delta \phi^I}. \tag{67.21}$$

[1] Sometimes called a **complete set** of gauge symmetries.

The following remark is the Lagrangian counterpart of Remark 49.3.18 in the Hamiltonian treatment of constrained systems:

Remark 67.1.6 (Lie algebroids). If one restricts to closed gauge algebras, i.e. ignores zilch symmetries, Equation 67.20 is exactly the closure condition for a Lie algebroid 42.3.3. Higher relations between the generators, i.e. a reducible theory, turns the gauge algebra into a Lie n-algebroid or even a L_∞-algebroid.

67.2 Covariant phase space

First, the formalism for mechanical systems will be introduced, i.e. no fields will be considered. Then, the formalism originally developed by *Peierls* will be introduced in the absence of local gauge symmetries.

67.2.1 Zero Hamiltonian

In Chapter 49 dynamical systems with constraints were considered. Using the tools from that chapter 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 called a **generally covariant system**.

If one starts from the action

$$S[q,p] = \int \big(p_i\dot{q}^i - H_0\big)dt, \tag{67.22}$$

one can introduce time as a generalized coordinate with momentum p_0 by modifying the action as follows:

$$S[q,p,t,p_0,u] = \int \big[p_0t' + p_iq'^i - u(p_0 + H_0)\big]d\tau, \tag{67.23}$$

where the quotes indicate derivatives with respect to the parameter τ. 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 constraint and, therefore, is given by a gauge transformation.

Remark 67.2.1 (Nonholonomic constraints). In Chapter 49 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 67.2.2 (Vanishing Hamiltonian). If the canonical coordinates (q,p) transform as scalars under τ-reparametrizations, the Hamiltonian is weakly zero.

67.2.2 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) one has to embrace a new approach:

Definition 67.2.3 (Covariant phase space). Let $S[\phi]$ be a local action functional. The covariant phase space $\mathcal{P}$ associated to S is the set of solutions of the equations of motion

$$\frac{\delta S}{\delta \phi} = 0. \tag{67.24}$$

The physical observables are defined as the smooth functions on this new phase space $\mathcal{P}$. These can be described more generally. Let M be the set of all field histories/configurations (the covariant phase space is a submanifold of this space). The ring of physical observables $C^\infty(\mathcal{P})$ is obtained as the quotient of $C^\infty(M)$ by the ideal of functions vanishing on-shell (similar to Definition 49.1.14).

?? COMPLETE ??

67.3 Batalin-Vilkovisky formalism

When considering constrained systems (Section 49.3) and their quantization (Section 57.4), one enlarged the phase space by the both ghosts and antighosts. The former corresponded to differential forms along the gauge orbits and antighosts corresponded to Koszul-Tate generators characterizing the zero locus of the field equations. Now, what about a field theory, where the solutions of the field equations are not functions in $C^\infty(M)$ but sections of a vector bundle $E \to M$. As seen in the first section of this chapter, field theories admit symmetries generated by the field equations themselves, the zilch symmetries. One can then play the same game as in Chapter 49 with the ordinary phase space replaced by the covariant phase space $\mathcal{I}$ and the symmetry group replaced by the zilch symmetries (and other gauge symmetries).

Definition 67.3.1 (Antifield). In the BRST formalism one extends the set of fields ϕ^I by a pair of ghosts and antighosts $(\eta^a, \mathcal{P}_a)$ for every gauge symmetry $\hat{G}_a$. In the antifield formalism one also introduces a set of antifields $\mathcal{P}_I$. These carry the following cohomological grading:

$$\varepsilon(\mathcal{P}_I) = \varepsilon_I + 1 \tag{67.25}$$

$$\mathrm{antigh}(\mathcal{P}_I) = \mathrm{antigh}(\phi^I) + 1 = 1. \tag{67.26}$$

They are the Koszul generators related to zilch symmetries:

$$\delta \mathcal{P}_I = -\frac{\delta S}{\delta \phi^I}. \tag{67.27}$$

The (anti)ghost fields then arise as the Koszul-Tate generators induced by the Noether identities 67.1.3 (which introduce some reducibility in the symmetries). $\delta(R^I_a \mathcal{P}_I) = 0$ by the above relation, so all Noether identities induce elements of $H^1(\delta)$. According to the Koszul-Tate construction one needs to introduce generators $\mathcal{P}_a$ and so on.

If one extends the above grading properties to the antighost fields, one obtains

$$\mathrm{antigh}(\mathcal{P}_a) = 2, \tag{67.28}$$

so the antifields are shifted in degree by 1 when compared to the classical setting. Also the Grassmann parity is shifted compared to the classical case:

$$\varepsilon(\mathcal{P}_I) = \varepsilon_I + 1 \qquad\qquad \varepsilon(\mathcal{P}_a) = \varepsilon_a + 1 \qquad\qquad \ldots \tag{67.29}$$

Definition 67.3.2 (Antifield bracket). The antifield bracket (**antibracket**) of two functionals on $C^\infty(\mathcal{I}) \otimes \mathbb{C}[\eta^a] \otimes \mathbb{C}[\mathcal{P}_I] \otimes \mathbb{C}[\mathcal{P}_a]$ is defined as follows:

$$\{f, g\} := \left(\frac{\partial^R f}{\partial \phi^I}\frac{\partial^L g}{\partial \mathcal{P}_I} - \frac{\partial^R f}{\partial \mathcal{P}_I}\frac{\partial^L g}{\partial \phi^I}\right) + \left(\frac{\partial^R f}{\partial \eta^a}\frac{\partial^L g}{\partial \mathcal{P}_a} - \frac{\partial^R f}{\partial \mathcal{P}_a}\frac{\partial^L g}{\partial \eta^a}\right), \tag{67.30}$$

where the index a denotes (anti)ghost fields of arbitrary degree.

Property 67.3.3. The BV-antibracket has the following algebraic properties:

- It is BRST-odd: $\mathrm{gh}(\{f,g\}) = \mathrm{gh}(f) + \mathrm{gh}(g) + 1$.
- It induces the structure of a Gerstenhaber algebra 27.7.4, where the degree is the Grassmann parity.

As before, of one considers the extended state space with "coordinates" the fields and "momenta" the antifields, the antibracket gives the structure of an odd symplectic manifold and, in particular, that of a "BV manifold" as defined below:

$$\{f,g\} = \frac{\partial^R f}{\partial z^\mu}\omega^{\mu\nu}\frac{\partial^L g}{z^\nu}, \tag{67.31}$$

where $(z^\mu) \equiv (\phi^I, \eta^a, \mathcal{P}_I, \mathcal{P}_a)$ and

$$(\omega^{\mu\nu}) := \begin{pmatrix} 0 & \delta^I_J + \delta^a_b \\ -\delta^I_J + \delta^a_b & 0 \end{pmatrix}. \tag{67.32}$$

Definition 67.3.4 (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 (42.12) is satisfied:

$$\{S,S\} = 0, \tag{67.33}$$

where $\{\cdot,\cdot\}$ is the Poisson bracket induced by ω (Definition 42.2.29).

The most straightforward example of a BV manifold is the BV-BRST complex associated to a field theory. For this reason the Poisson bracket is often called the **antibracket**, while the grading is often called the **ghost number** and denoted by gh. The function S is for field theories given by the action (functional).

Definition 67.3.5 (BV Laplacian). Consider a BV manifold (M, ω, S). A BV Laplacian is an operator Δ on the space of half-densities $|\Omega|^{1/2}(M)$ that satisfies the following conditions:

1. **Nilpotency**: $\Delta^2 = 0$, and
2. **Product rule**: $\Delta(fg) = (\Delta f)g + (-1)^{\varepsilon(f)}f(\Delta g) + (-1)^{\varepsilon(f)}\{f,g\}$.

This implies that the BV Laplacian acts as a graded derivation:

$$\Delta\{f,g\} = \{\Delta f, g\} + (-1)^{\varepsilon(f)+1}\{f, \Delta g\}. \tag{67.34}$$

The canonical BV Laplacian in field-antifield coordinates is given by

$$\Delta_{\mathrm{BV}} := \sum_I (-1)^{\varepsilon_I+1}\frac{\partial^R}{\partial\phi^I}\frac{\partial^R}{\partial\mathcal{P}_I} + \sum_a (-1)^{\varepsilon_a+1}\frac{\partial^R}{\partial\eta^a}\frac{\partial^R}{\partial\mathcal{P}_a}. \tag{67.35}$$

This operator satisfies

$$\Delta_{\mathrm{BV}} f = -\frac{1}{2}\mathrm{div}X_f, \tag{67.36}$$

for all $f \in C^\infty(M)$, where div denotes the divergence 34.1.21 with respect to the standard Berezinian volume form.

Definition 67.3.6 (BV integral). Consider an odd symplectic (M, ω) of superdimension $m|m$. Let ψ, the **gauge fixing fermion**, be an odd function of the even coordinates of a Darboux basis (denote these by q). This function determines a projectable Lagrangian submanifold

$L_\psi \subset \Pi T^*\mathbb{R}^m$ by the Maslow-Hörmander theorem 35.3.7. The Batalin-Vilkovisky integral of a function $f \in C^\infty(M)$ with respect to ψ is defined as

$$\int_{L_\psi} f := \int f|_{p_i=\partial_{q^i}\psi}, \tag{67.37}$$

The BV integral and BV Laplacian interact in the following way (for BV integrable f):

1. If $f = \Delta_{\text{BV}} g$, then $\int_{L_\psi} f = 0$ for all gauge fixing fermions ψ.
2. If $\Delta_{\text{BV}} f = 0$, $\frac{d}{dt}\int_{L_{\psi_t}} f = 0$, where $\{\psi_t\}_{t\in\mathbb{R}}$ 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.

Formula 67.3.7 (Quantum master equation). Consider the function $f := e^{iS/\hbar}$ on a BV manifold (M, ω, S). Because

$$\Delta f = \frac{i}{\hbar}\Delta S e^{iS/\hbar} + \left(\frac{i}{\hbar}\right)^2 \frac{1}{2}\{S, S\}e^{iS/\hbar}, \tag{67.38}$$

the condition that f is BV-harmonic is equivalent to S satisfying

$$\frac{1}{2}\{S, S\} - i\hbar\Delta S = 0. \tag{67.39}$$

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 (42.12).

Example 67.3.8 (AKSZ model). The Alexandrov-Kontsevich-Schwarz-Zabronsky model considers the mapping space between a dg-manifold (M, Q) and 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 $Q := d$. A symplectic form on $C^\infty(M, N)$ is then given by

$$\Omega := \int_{\Pi T\Sigma} \omega_{\mu\nu}\delta\Phi^\mu\delta\Phi^\nu \text{Vol}. \tag{67.40}$$

The BV action is defined as follows:

$$S := \int_{\Pi T\Sigma} (\alpha_\mu d\Phi^\mu + \Theta)\text{Vol}, \tag{67.41}$$

where α is a symplectic potential for ω, which necessarily exists globally by Property 42.2.31 if $\text{gh}(\omega) \neq 0$.

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 $\text{ev} : C^\infty(M, N) \times M \to N$ and then one pushes it forward along the projection on the first factor (cf. fibre integration 32.8.22).

Chapter 68

Canonical Quantization

The main reference for this chapter is [49].

68.1 Klein-Gordon field

68.1.1 Lagrangian and Hamiltonian

The "simplest" Lagrangian (density) is given by

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2. \tag{68.1}$$

Using the principle of least action, the following Euler-Lagrange equation is obtained:

$$\left(\partial^\mu\partial_\mu + m^2\right)\phi = 0. \tag{68.2}$$

This can be rewritten using the **d'Alembertian** $\Box = \partial_\mu\partial^\mu$:

$$(\Box + m^2)\phi = 0. \tag{68.3}$$

This equation is called the **Klein-Gordon equation**. In the limit $m \to 0$ this equation reduces to the well-known wave equation.

From the Lagrangian (68.1) one can also derive a Hamiltonian function using relations 48.2.3 and 48.3.2:

$$H = \frac{1}{2}\int d^3x\left[\pi^2(x) + (\nabla\phi(x))^2 + m^2\phi^2(x)\right]. \tag{68.4}$$

68.1.2 Raising and lowering operators

Fourier transforming the scalar field $\phi(\vec{\boldsymbol{x}}, t)$ in momentum space and inserting it into the Klein-Gordon equation gives

$$\left(\partial_t^2 + p^2 + m^2\right)\phi(\vec{\boldsymbol{p}}, t) = 0. \tag{68.5}$$

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 $a^\dagger_{\vec{\boldsymbol{p}}}$ and $a_{\vec{\boldsymbol{p}}}$ can be defined such that

$$\phi(\vec{\boldsymbol{x}}) = \iiint \frac{d^3p}{(2\pi)^{3/2}}\frac{1}{\sqrt{2\omega_{\vec{\boldsymbol{p}}}}}\left(a_{\vec{\boldsymbol{p}}}e^{i\vec{\boldsymbol{p}}\cdot\vec{\boldsymbol{x}}} + a^\dagger_{\vec{\boldsymbol{p}}}e^{-i\vec{\boldsymbol{p}}\cdot\vec{\boldsymbol{x}}}\right) \tag{68.6}$$

$$\pi(\vec{\boldsymbol{x}}) = \iiint \frac{d^3p}{(2\pi)^{3/2}}(-i)\sqrt{\frac{\omega_{\vec{\boldsymbol{p}}}}{2}}\left(a_{\vec{\boldsymbol{p}}}e^{i\vec{\boldsymbol{p}}\cdot\vec{\boldsymbol{x}}} - a^\dagger_{\vec{\boldsymbol{p}}}e^{-i\vec{\boldsymbol{p}}\cdot\vec{\boldsymbol{x}}}\right). \tag{68.7}$$

An equivalent definition is obtained by performing the transformation $\vec{\boldsymbol{p}} \longrightarrow -\vec{\boldsymbol{p}}$ in the second term of $\phi(\vec{\boldsymbol{x}})$ and $\pi(\vec{\boldsymbol{x}})$:

$$\phi(\vec{\boldsymbol{x}}) = \iiint \frac{d^3p}{(2\pi)^{3/2}} \frac{1}{\sqrt{2\omega_{\vec{\boldsymbol{p}}}}} \left(a_{\vec{\boldsymbol{p}}} + a^\dagger_{-\vec{\boldsymbol{p}}}\right) e^{i\vec{\boldsymbol{p}}\cdot\vec{\boldsymbol{x}}} \tag{68.8}$$

$$\pi(\vec{\boldsymbol{x}}) = \iiint \frac{d^3p}{(2\pi)^{3/2}} (-i) \sqrt{\frac{\omega_{\vec{\boldsymbol{p}}}}{2}} \left(a_{\vec{\boldsymbol{p}}} - a^\dagger_{-\vec{\boldsymbol{p}}}\right) e^{i\vec{\boldsymbol{p}}\cdot\vec{\boldsymbol{x}}}. \tag{68.9}$$

When the commutation relation

$$[a_{\vec{\boldsymbol{p}}}, a^\dagger_{\vec{\boldsymbol{q}}}] := \delta(\vec{\boldsymbol{p}} - \vec{\boldsymbol{q}}) \tag{68.10}$$

is imposed, the following commutation relation for the scalar field and its conjugate momentum is obtained:

$$[\phi(\vec{\boldsymbol{x}}), \pi(\vec{\boldsymbol{y}})] = i\delta(\vec{\boldsymbol{x}} - \vec{\boldsymbol{y}}). \tag{68.11}$$

Now, the Hamiltonian can be calculated explicitly:

$$H = \int \frac{d^3p}{(2\pi)^3} \omega_{\vec{\boldsymbol{p}}} \left(a^\dagger_{\vec{\boldsymbol{p}}} a_{\vec{\boldsymbol{p}}} + \frac{1}{2}[a_{\vec{\boldsymbol{p}}}, a^\dagger_{\vec{\boldsymbol{p}}}]\right). \tag{68.12}$$

However, it is clear from Equation (68.10) 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 (68.4) 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 Equation (68.12) together with the canonical commutation relations is

$$[H, a^\dagger_{\vec{\boldsymbol{p}}}] = \omega_p a^\dagger_{\vec{\boldsymbol{p}}} \tag{68.13}$$

$$[H, a_{\vec{\boldsymbol{p}}}] = -\omega_p a_{\vec{\boldsymbol{p}}}. \tag{68.14}$$

As was the case for the quantum harmonic oscillator, the creation and annihilation operators deserve their names and one can write:

$$|\vec{\boldsymbol{k}}_1, \ldots, \vec{\boldsymbol{k}}_n\rangle = a^\dagger(\vec{\boldsymbol{k}}_1) \cdots a^\dagger(\vec{\boldsymbol{k}}_n)|0\rangle. \tag{68.15}$$

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 (in fact Fock) state space is axiomatized by the GNS construction 24.1.40, where the vacuum plays the role of cyclic vector.

68.1.3 Scalar propagator

Formula 68.1.1 (Pauli-Jordan function).

$$i\Delta(x-y) := i[\phi(x), \phi(y)] = \int \frac{d^3p}{(2\pi)^3} \frac{1}{2\omega_p} \left(e^{-i\mathbf{p}\cdot(\mathbf{x}-\mathbf{y})} - e^{i\mathbf{p}\cdot(\mathbf{x}-\mathbf{y})}\right) \tag{68.16}$$

In the case that $x^0 = y^0$ (ETCR) or $(x-y)^2 < 0$ (spacelike curves), the Pauli-Jordan function is identically 0. (See also the *axiom of microcausality* 72.1).

68.1.4 Normalization constant

By Property 17.1.23 the delta function $\delta^{(3)}(\vec{\boldsymbol{p}}-\vec{\boldsymbol{q}})$ transforms as $\delta^{(3)}(\Lambda\vec{\boldsymbol{p}}-\Lambda\vec{\boldsymbol{q}})\frac{\Lambda E}{E}$ under a general Lorentz boost Λ. Although this is clearly not Lorentz invariant, the quantity $E_p\delta^{(3)}(\vec{\boldsymbol{p}}-\vec{\boldsymbol{q}})$ can be seen to be invariant. From this observation it follows that the correct normalization in the momentum representation becomes

$$|\mathbf{p}\rangle = \sqrt{2E_p}a^\dagger_{\mathbf{p}}|0\rangle \tag{68.17}$$

and, hence,

$$\langle\mathbf{p}|\mathbf{q}\rangle = 2E_p(2\pi)^3\delta^{(3)}(\vec{\boldsymbol{p}}-\vec{\boldsymbol{q}}), \tag{68.18}$$

where the constants are a matter of convention to cancel the constants in expression (68.6).

68.1.5 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:

$$\int\frac{d^3p}{2E_p} = \int\delta(p^2-m^2)d^4p\bigg|_{p^0>0}. \tag{68.19}$$

By using this measure it is ensured that the integral of any Lorentz invariant function is again Lorentz invariant.

Example 68.1.2 (One-particle identity operator).

$$\mathbb{1} := \int\frac{d^3p}{2E_p}|\mathbf{p}\rangle\langle\mathbf{p}| \tag{68.20}$$

68.2 Contractions and Wick's theorem

68.2.1 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 the classic Fourier integral. By looking at Equation (68.6) and remembering that the $(1,3)$-signature is adopted,the annihilators can be seen to always occur together with a positive frequency exponential.

Definition 68.2.1 (Contraction for neutral bosonic fields).

$$\overbrace{\phi(x)\phi}(y) := \begin{cases}[\phi(x)^{(+)},\phi(y)^{(-)}] & x^0>y^0\\ [\phi(y)^{(+)},\phi(x)^{(-)}] & y^0>x^0\end{cases} \tag{68.21}$$

Formula 68.2.2 (Feynman propagator).

$$i\Delta_F(x-y) := \overbrace{\phi(x)\phi}(y) := i\lim_{\varepsilon\to0^+}\int\frac{d^4p}{(2\pi)^4}\frac{e^{-ip\cdot(x-y)}}{p^2-m^2+i\varepsilon} \tag{68.22}$$

Definition 68.2.3 (Contraction for charged bosonic fields).

$$\overbracket{\phi(x)\overline{\phi}}(y) := \begin{cases} [\phi(x)^{(+)}, \overline{\phi}(y)^{(-)}] & x^0 > y^0 \\ [\phi(y)^{(+)}, \overline{\phi}(x)^{(-)}] & y^0 > x^0 \end{cases} \tag{68.23}$$

Definition 68.2.4 (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.:

$$\mathcal{N}\Big(\phi(x)\phi^\dagger(y)\phi(z)\Big) = \phi^\dagger(y)\phi(x)\phi(z).$$

In fact the normal ordering operator is not a morphism between CCR-algebras since this would lead to a contradiction:

$$b_i b_j^\dagger = b_j^\dagger b_i + \delta_{ij} \implies \mathcal{N}(b_i b_j^\dagger) = \mathcal{N}(b_j^\dagger b_i + \delta_{ij}) = \mathcal{N}(b_j^\dagger b_i) + \delta_{ij} \implies \delta_{ij} \equiv 0.$$

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 $\mathcal{N} : \mathrm{Sym}^\bullet A \to A$, where A is the CCR-algebra of the theory (see [25]).

Property 68.2.5. From this definition it immediately follows that the vacuum expectation value of a normal ordered sequence is 0.

Formula 68.2.6 (Wick's theorem for bosonic fields).

$$\mathcal{T}\Big(\phi(x_1)\phi(x_2)\cdots\phi(x_n)\Big) = \mathcal{N}\Big(\phi(x_1)\cdots\phi(x_n) + \text{all possible contractions}\Big) \tag{68.24}$$

When acting with such a time-ordered product on the vacuum, this theorem implies that only fully contracted terms will remain. Moreover, by Formula 68.2.2, every such action can be expressed solely in terms of propagators.

Remark 68.2.7. In the case of charged bosons, only contractions of the form $\overbracket{\phi(x)\overline{\phi}}(y)$ will remain because $[a, b^+] = 0$.

Corollary 68.2.8.

$$\overbracket{\phi(x)\phi}(y) = \mathcal{T}\Big(\phi(x)\phi(y)\Big) - \mathcal{N}\Big(\phi(x)\phi(y)\Big) \tag{68.25}$$

Wick's theorem has an analogue in probability theory:

Formula 68.2.9 (Isserlis' theorem). Let $X_1, \ldots, X_n$ be a set of random variables following a multinormal distribution with mean 0.

$$\mathrm{E}[X_1 \cdots X_n] = \sum_{\sigma\in P_{n,2}} \prod_{\{i,j\}\in\sigma} \mathrm{E}[X_i X_j], \tag{68.26}$$

where $P_{n,2}$ denotes the set of binary partitions of $\{1, \ldots, n\}$. Because the mean of the distribution is zero, this expression is equal to

$$\mathrm{E}[X_1 \cdots X_n] = \sum_{\sigma\in P_{n,2}} \prod_{\{i,j\}\in\sigma} \mathrm{cov}[X_i X_j]. \tag{68.27}$$

68.2.2 Fermionic fields

Definition 68.2.10 (Contraction).

$$\overbracket{\psi(x)\overline{\psi}}(y) := \begin{cases}\{\psi(x)^{(+)}, \overline{\psi}(y)^{(-)}\}_+ & x^0 > y^0\\ -\{\psi(y)^{(+)}, \overline{\psi}(x)^{(-)}\}_+ & y^0 > x^0\end{cases} \tag{68.28}$$

Remark 68.2.11. Only contractions of the form $\overbracket{\psi(x)\overline{\psi}}(y)$ will remain because $\{a, b^\dagger\}_+ = 0$.

Formula 68.2.12 (Feynman propagator).

$$i\Delta_F(x-y) := \overbracket{\psi(x)\overline{\psi}}(y) := i\lim_{\varepsilon\to 0^+}\int \frac{d^4p}{(2\pi)^4}\frac{\not{p}+m}{p^2-m^2+i\varepsilon}e^{-ip\cdot(x-y)} \tag{68.29}$$

Remark 68.2.13 (Normal ordering). One should take into account the Fermi-Dirac statistics when permuting fermionic field operators under a normal ordering. A general factor $\mathrm{sgn}(\sigma)$, where σ is the permutation of the operators, will arise in every term, e.g.:

$$\mathcal{N}\Big(\psi(x)\overline{\psi}(y)\psi(z)\Big) = -\overline{\psi}(y)\psi(x)\psi(z)$$

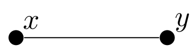
A similar remark should be made for the time-ordering operator $\mathcal{T}$. As was the case for bosonic theories one should pay attention to the nature of the normal ordering. It is not a morphism between CAR-algebras, but instead it is an algebra morphism between a free (odd) algebra and a CAR-algebra.

68.3 Feynman rules

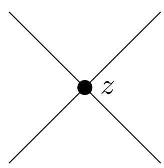
68.3.1 Scalar theory

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 (for these rules 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 $\Delta_F(x-y)$:



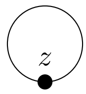
- Interaction vertex[1] $-i\lambda\int d^4z$:



The main idea of 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:

[1]Four legs were drawn as an example but this can be generalized to any order of interaction term.

Remark 68.3.1. 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 $-\frac{i\lambda}{2}\int d^4z\,\Delta_F(z-z)$.

68.4 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.

68.4.1 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 62.6.1 on a lattice Λ:

$$\hat{H} := -\sum_{\langle i,j\rangle\in\Lambda} J_{ij}\hat{S}_i\hat{S}_j - h\sum_{i\in\Lambda}\hat{S}_i. \tag{68.30}$$

68.4.2 Dimensional regularization

?? COMPLETE ??

68.4.3 Wilsonian renormalization

?? COMPLETE ??

68.5 Quantum Chromodynamics

Property 68.5.1 (OZI rule[2]). 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.

68.6 Entanglement in QFT

The main reference is [29, 85]. This section should be seen as a generalization of the content of Chapter 60 to the continuum setting (in particular the characterization and computation of entanglement).

68.6.1 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.

[2] *Okubo, Zweig* and *Iizuka*

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:[3]

$$S(\rho_A) := -\operatorname{tr}(\rho_A \ln \rho_A) = -\sum_i \rho_i \ln \rho_i. \tag{68.31}$$

In the same way one can also introduce the Rényi q-entropy:

$$S_q(\rho_A) := \frac{1}{1-q} \ln\left(\sum_i \rho_i^q\right). \tag{68.32}$$

Property 68.6.1 (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 ??

[3]Certain assumption 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 [84]).

Chapter 69

Gauge Theory

References for this chapter are [60, 64, 100, 114]. The section on the *Higgs mechanism* is mainly based on [10]. Using the tools of differential geometry, as presented in Chapter 31 and onwards, one can introduce a general formulation of gauge theories and, in particular, Yang-Mills theories.

69.1 Gauge invariance

Consider a general Lie group G, often called the **gauge group**, acting on a vector bundle with typical fibre $\mathcal{H}$ over a base manifold M. This bundle is in general obtained as an associated bundle of the frame bundle FM. A general gauge transformation has the form

$$\psi'(x) = U(x)\psi(x), \tag{69.1}$$

where $\psi, \psi' : M \to \mathcal{H}$ are sections of $\mathcal{H}$ 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 $\mathcal{H}$. As such, a gauge transformation constitutes a vertical automorphism of the vector bundle.

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 field $B_\mu(x)$, 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

$$\mathcal{D}_\mu = \partial_\mu + igB_\mu(x), \tag{69.3}$$

where $B_\mu : M \to \mathfrak{g}$ is a new field with values in the Lie algebra of the gauge group. This procedure is called **minimal coupling**. It should be noted that the explicit action of the covariant derivative depends on the chosen representation of $\mathfrak{g}$ on $\mathcal{H}$. Furthermore, one should pay attention to the fact that the physics convention was used where one multiplies the gauge field B 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. $\mathcal{D}'\psi' = U\mathcal{D}\psi$ (from here the coordinate-dependence of all fields will be supressed):

$$\begin{aligned}U^{-1}\left(\frac{\partial}{\partial x^\mu} + igB'_\mu\right)\psi' &= U^{-1}\left(\frac{\partial}{\partial x^\mu} + igB'_\mu\right)U\psi \\ &= U^{-1}\frac{\partial U}{\partial x^\mu}\psi + \frac{\partial\psi}{\partial x^\mu} + igU^{-1}B'_\mu U\psi. \qquad (69.4)\end{aligned}$$

This expression can only be equal to $\mathcal{D}\psi$ if

$$igB_\mu = U^{-1}\frac{\partial U}{\partial x^\mu} + igU^{-1}B'_\mu U, \qquad (69.5)$$

which can be rewritten as

$$B'_\mu = UB_\mu U^{-1} - \frac{1}{ig}(\partial_\mu U)U^{-1} \qquad (69.6)$$

or, in coordinate-independent form, as

$$\mathbf{B}' = U\mathbf{B}U^{-1} - \frac{1}{ig}dUU^{-1}. \qquad (69.7)$$

Up to conventions this is exactly the content of Equations (33.38) and (33.40) appearing in the study of connections on principal bundles. 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 physical manifold to a principal bundle.

Example 69.1.2 (QED). For quantum electrodynamics, which has $\mathrm{U}(1) \cong S^1$ as its gauge group, the parametrization $U(x) = e^{ie\chi(x)}$ is used with $\chi : \mathbb{R}^n \rightarrow \mathbb{R}$. Minimal coupling leads to

$$\partial_\mu \longrightarrow \mathcal{D}_\mu = \partial_\mu + ieA_\mu \qquad (69.8)$$

$$A_\mu \longrightarrow A'_\mu = A_\mu - \partial_\mu\chi, \qquad (69.9)$$

where A_μ is the classic electromagnetic potential. These are the formulas that introduced in Chapter 50.

69.2 Spontaneous symmetry breaking

Theorem 69.2.1 (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 destroy 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**.

69.2.1 Higgs mechanism

In Property 33.1.21 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:

$$V(\phi) = 0, \tag{69.10}$$
$$\nabla\phi = 0, \tag{69.11}$$

where ∇ is the covariant derivative and ϕ is a section of some associated (finite-rank) vector bundle $P \times_\rho E$. If the space of solutions $\mathcal{M}$ to the first equation admits a transitive G-action, i.e. it is a homogeneous space, by Property 3.3.13 one can write

$$\mathcal{M} \cong G/H, \tag{69.12}$$

where H is the isotropy group of any 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 of 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 $\phi : P \to \mathcal{M}_0 \cong G/H$, 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 a smaller symmetry group H, induces by Property 33.5.7 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:

$$\omega_{\mathfrak{g}} = \omega_{\mathfrak{h}} + \gamma, \tag{69.13}$$

where the left-hand side denotes the pullback of the G-connection along the reduction $\iota : P_H \hookrightarrow P$ and γ is a tensorial $(\mathrm{Ad}_H, \mathfrak{m})$-form with $\mathfrak{m}$ the complement of $\mathfrak{h}$ in $\mathfrak{g}$.[2]

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

$$\nabla_X\phi = (\rho_{e,*} \circ \gamma)(X)m_0. \tag{69.14}$$

The generators $\rho_{e,*}(\gamma^i_\mu)m_0$, where $i = 1, \ldots, \dim(\mathfrak{m})$, 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 33.5.10 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.

[2] 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 33.8.4.

Chapter 70

Particle Physics

This chapter summarizes the content of the foregoing chapters and tries to relate it to a more phenomenological approach to modern physics.

70.1 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 70.1.1 (Leptons). The leptons are fermionic particles, i.e. characterized by the Dirac equation 56.2.1. 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 70.1.2 (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 70.1.3 (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 70.1.4. 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 70.1.5 (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 70.1.6 (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

$$\mathcal{L}_{\text{Yukawa}}(\phi, \psi) \sim \overline{\psi}\phi\psi \tag{70.1}$$

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.

70.2 Standard model

70.2.1 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: U(1),
- Weak interaction: SU(2), and
- Strong interaction: SU(3).

The fact that all particles can be organised in terms of representations of these groups has some important consequences.

Property 70.2.1 (Electromagnetism). All (electromagnetically interacting) fundamental particles transform under the fundamental representation of U(1). Since this representation is one-dimensional, interactions with a photon do not change particle types.

Property 70.2.2 (Weak interactions). All (weakly interacting) fundamental particles can be organised into fundamental representations of 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.:

$$\begin{pmatrix} e^- \\ \nu_e \end{pmatrix}_L \qquad\qquad e^-_R \qquad\qquad \nu_{e,R}.$$

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 70.2.3 (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 $r\overline{r}+g\overline{g}+b\overline{b}$ would exist, it could couple colourless states to colourful states, which is not allowed.

70.2.2 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

$$\mathrm{U}(1)\times\mathrm{SU}(2)\times\mathrm{SU}(3)/\mathbb{Z}_6. \tag{70.2}$$

The center $\mathbb{Z}_6$ acts trivially on all particle types. It is generated by the element

$$(e^{\pi i/3}, -1, e^{2\pi i/3}). \tag{70.3}$$

As noted in the previous section the second factor only acts on left-handed particles, so a better notation would be $\mathrm{SU}(2)_L$. However, the first factor is also not what it seems. The U(1) gauge group is not simply that of the electromagnetic interaction. If 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**:

$$Q = I_3 + \tfrac{1}{2}Y, \tag{70.4}$$

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 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:

$$\gamma = W^0 + \tfrac{1}{2}B \tag{70.5}$$

$$Z^0 = W^0 - \tfrac{1}{2}B. \tag{70.6}$$

This recovers the Gell-Mann-Nishijima formula.

The reason for electroweak symmetry breaking from $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ to $\mathrm{U}(1)_{\mathrm{em}}$ is given by the Higgs mechanism (see Section 69.2.1). Before symmetry breaking, the Higgs field is a $\mathrm{SU}(2)_L$-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 $\mathrm{SU}(2)_L$-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 ??

70.3 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.

70.3.1 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 73 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.

70.3.2 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:

$$\vec{\boldsymbol{B}} = g\frac{e}{2m}\vec{\boldsymbol{S}}. \tag{70.7}$$

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 71

Conformal Field Theory

References for this chapter are [38] and lecture notes by *Schellekens*. For an introduction to Riemannian and conformal geometry, see Chapter 34 and, in particular, Section 34.4.

Property 71.0.1 (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:

$$T^\mu_\mu = 0. \tag{71.1}$$

71.1 In dimension $d = 2$

In dimension 2 (in Euclidean signature) something special happens. By inserting $d = 2$ in the conformal Killing equation (34.56), the Cauchy-Riemann equations 15.2.4 are obtained. The scale factor can thus be written as

$$\kappa(z) = \left|\frac{\partial f}{\partial z}\right|^2 \tag{71.2}$$

for some analytic function $f(z)$. 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:

$$g_{zz} = g_{\overline{z}\overline{z}} = 0 \qquad g_{z\overline{z}} = \frac{1}{2} \tag{71.3}$$

$$\partial_z T_{\overline{z}\overline{z}} = \partial_{\overline{z}} T_{zz} = 0 \qquad T_{z\overline{z}} = 0. \tag{71.4}$$

The stress-energy tensor thus contains a meromorphic[1] component $T_{zz} \equiv T(z)$ and an antimeromorphic component $T_{\overline{z}\overline{z}} \equiv \overline{T}(\overline{z})$.

Definition 71.1.1 (Witt algebra). Infinitesimally this gives an infinite-dimensional algebra. The generators can be chosen to be

$$l_n(z) := -z^{n+1}\partial_z \tag{71.5}$$

$$\bar{l}_n(\overline{z}) := -\overline{z}^{n+1}\partial_{\overline{z}}. \tag{71.6}$$

These generate the transformation $z \mapsto z - z^{n+1}$ and $\overline{z} \mapsto \overline{z} - \overline{z}^{n+1}$ respectively. They also give rise to isomorphic Lie algebras with the following commutation relation:

$$[l_m, l_n] = (m-n)l_{m+n}. \tag{71.7}$$

This Lie algebra is called the **Witt algebra**.

[1] The literature often just calls this holomorphic.

Remark 71.1.2 (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 $\mathrm{PSL}(2,\mathbb{C})$. 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 2D Minkowski space is infinite-dimensional. It can be shown that $\mathrm{Conf}(\mathbb{R}^{1,1})$ is isomorphic to $\mathrm{Diff}(S^1)_+ \times \mathrm{Diff}(S^1)_+$, where the orientation-preserving diffeomorphism group $\mathrm{Diff}(S^1)_+$ 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 71.1.3 (Primary field). A field that transforms tensorially under global conformal transformations:

$$\phi'(z',\overline{z}') = \left(\frac{\partial f}{\partial z}\right)^h \left(\frac{\partial f}{\partial \overline{z}}\right)^{\overline{h}} \phi\big(f(z),\overline{f}(\overline{z})\big). \tag{71.8}$$

Fields that satisfy this relation for all conformal transformations are called primary fields. The tuple $(h,\overline{h})$ is called the **conformal weight** of the field.

71.1.1 Intermezzo: 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.

?? MOVE THIS TO DIFFERENT CHAPTER ??

Definition 71.1.4 (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).

Using this definition one can start to analyze the group $\mathrm{Diff}(S^1)_+$. First of all, look at the space of all smooth maps $S^1 \to S^1$. 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]:

$$\mathfrak{X}(S^1) = \left\{ \xi(\theta)\frac{\partial}{\partial\theta} \,\middle|\, \theta \in C^\infty(S^1) \right\}. \tag{71.9}$$

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.

[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 function ξ.

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

$$\left[\xi_k(\theta)\frac{\partial}{\partial\theta}, \xi_l(\theta)\frac{\partial}{\partial\theta}\right] = -i(l-k)\xi_{k+l}(\theta)\frac{\partial}{\partial\theta}. \tag{71.10}$$

These are exactly the relations for the Witt algebra.

71.1.2 Minkowski space

By Property 34.4.5 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, note that with the next section in mind, it is important to look at the interaction between representation theory and quantum mechanics.

When treating spin quantum mechanically we were forced to pass from the rotation group $\mathrm{SO}(3)$ to its double cover $\mathrm{SU}(2)$. Then, when passing to special relativity, we had to replace the Lorentz group $\mathrm{SO}^{\uparrow}(3,1)$ by its double cover $\mathrm{SL}(2,\mathbb{C})$. The same story repeats itself here, we should replace the conformal group by its double cover $\mathrm{SU}(2,2)$. This is also how *twistor theory* comes into the picture.

71.2 Quantization in $d = 2$

71.2.1 Radial quantization

The charge of a conserved current $\mathcal{J}^\mu$ is generally given by Equation (67.11):

$$Q = \int_\Sigma \mathcal{J}^0(x,t), \tag{71.11}$$

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:

$$Q_\varepsilon = \frac{1}{2\pi i}\oint dz\varepsilon(z)T(z) + \frac{1}{2\pi i}\oint d\overline{z}\varepsilon(\overline{z})\overline{T}(\overline{z}). \tag{71.12}$$

As usual an infinitesimal transformation of a field ϕ is given by the commutator $[Q_\varepsilon, \phi(w,\overline{w})]$ or, in integral form, by[3]

$$\frac{1}{2\pi i}\oint dz\varepsilon(z)[T(z)\phi(w,\overline{w}) - \phi(w,\overline{w})T(z)].$$

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**:

$$\mathcal{R}(A(z,\overline{z})B(w,\overline{w})) := \begin{cases} A(z,\overline{z})B(w,\overline{w}) & |z| > |w| \\ B(w,\overline{w})A(z,\overline{z}) & |w| > |z|. \end{cases} \tag{71.13}$$

[3] A holomorphic split is implicitly assumed such that antiholomorphic contributions $\overline{T}(\overline{z})$ can be ignored.

After a deformation of the integration contour the following general formula is obtained:

$$[Q_\varepsilon, \phi(w,\overline{w})] = \frac{1}{2\pi i}\oint dz\varepsilon(z)\mathcal{R}(T(z)\phi(w,\overline{w})), \tag{71.14}$$

where the contour is a circle around the point w. For primary fields with conformal weight h, one can also write an infinitesimal transformation as

$$\delta_\varepsilon\phi(z,\overline{z}) = h(\partial_z\varepsilon(z))\phi(z,\overline{z}) + \partial_z\phi(z,\overline{z}).$$

Comparing these two expressions leads to the following form of the operator product:

$$\mathcal{R}(T(z)\phi(w,\overline{w})) = \frac{h}{(z-w)^2}\phi(w,\overline{w}) + \frac{1}{z-w}\partial_w\phi(w,\overline{w}) + \text{higher order in } (z-w). \tag{71.15}$$

An expression of this form is called an **operator product expansion** (OPE).

71.2.2 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 30.2.6) 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 30.2.41, one obtains the **Virasoro algebra** as a (universal) central extension of the Witt algebra by[4] $\mathbb{C}$ associated to the cocycle

$$\Theta : (L_m, L_n) \mapsto \frac{c}{12}m(m^2-1)\delta_{m+n,0}. \tag{71.16}$$

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:

$$L_n := \frac{1}{2\pi i}\oint dz z^{n+1}T(z). \tag{71.17}$$

This relation can be inverted using the residue theorem 15.5.20 to obtain the following expression for (the holomorphic component of) the stress-energy tensor:

$$T(z) = \sum_{n=-\infty}^{\infty} z^{-n-2}L_n. \tag{71.18}$$

Using the above expression and the product operator expansion of $T(z)T(w)$, one obtains exactly the commutation relations of the Virasoro algebra:

$$[L_m, L_n] = (m-n)L_{m+n} + \frac{c}{12}m(m^2-1)\delta_{m+n,0}. \tag{71.19}$$

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.

[4]The reason to extend by $\mathbb{C}$ instead of $\mathbb{R}$ is that all Lie algebras are complexified in this section.

71.2.3 Representation theory

Definition 71.2.1 (Highest weight state). A state with minimal eigenvalue for L_0. Equivalently, a state that is annihilated by all generators L_n for $n \geq 1$:

$$L_n|h\rangle = 0. \tag{71.20}$$

Definition 71.2.2 (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

$$L_n|0\rangle = 0$$

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 71.2.3 (Descendants). By acting with the generators L_n, where $n \leq -1$, 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 30.4.62 associated to $|h\rangle$.

71.3 Extensions ♣

71.3.1 TCFT

By the *FRS theorem*, rational 2D CFTs are classified by pairs of modular tensor categories and internal (special symmetric) Frobenius algebras. By categorifying the latter, Calabi-Yau categories were obtained 27.2.15. 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*).

Chapter 72

Axiomatic QFT ♣

For the sections on the Haag-Kastler framework and its extensions hte reader is referred to the work of *Brunetti, Fredenhagen et al.* A reference for the remaining sections on algebraic QFT is [33]. The sections on cohomological quantization and quantization from a sheaf point of view are based on [128]. For the sections on TQFTs and *open-closed* TQFTs the reader is referred to the original papers [83] and [95], respectively.

72.1 Algebraic QFT

72.1.1 Haag-Kastler axioms

Axiom 72.1 (Local net of observables). To every causally closed set 66.1.4 one associates a C^*-algebra. This assignment is required to satisfy the following conditions:

1. **Isotony**: If $O_1 \subset O_2$, then $\mathcal{A}(O_1) \hookrightarrow \mathcal{A}(O_2)$.
2. **(Causal) locality**[1]: If O_1 and O_2 are spacelike separated, then $[\mathcal{A}(O_1), \mathcal{A}(O_2)] = 0$ (as a graded commutator) within a larger algebra $\mathcal{A}(O)$ such that $O_1, O_2 \subset O$.

Remark 72.1.1. The isotony condition implies that local nets of observables are modelled by copresheaves $\mathbf{Mink} \rightarrow \mathbf{C^*Alg}$ that map (mono)morphisms to monomorphisms.

Axiom 72.2 (Poincaré covariance). For all causally closed sets O and Poincaré transformations Λ there exists an isomorphism $\alpha^O_\Lambda : \mathcal{A}(O) \rightarrow \mathcal{A}(\Lambda O)$ 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 72.3 (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 24.1.40. 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:

$$U(a)\rho_O(c)U(a)^{-1} = \rho_{O+a}(\alpha^O_a(c)) \tag{72.1}$$

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.

[1] Also called **microcausality** or **Einstein causality**.

The following axiom is not part of the standard Haag-Kastler framework but can be added to introduce dynamics:

Axiom 72.4 (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 72.5 (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[2]

$$\mathcal{A}(\overline{O})' = \mathcal{A}(O) \tag{72.2}$$

for all causally closed sets O.

Remark 72.1.2. 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 [109].

To generalize the above axiom system to globally hyperbolic space times, one must enter the realm of category theory. The notation[3] of [58] (?? AND OTHERS ??) 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 theories) together with suitable algebra morphisms. The assignment of algebras is then given by a functor $\mathfrak{U} : \mathbf{Loc} \rightarrow \mathbf{Obs}$. The Haag-Kastler framework is recovered when $\mathfrak{U}$ is restricted to the subcategory on globally hyperbolic subsets of some manifold (with inclusions as morphisms).

72.1.2 Weyl systems

Definition 72.1.3 (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

$$W(z)W(z') = e^{\frac{i}{2}\omega(z,z')}W(z + z') \tag{72.3}$$

for all $z, z' \in L$.

Remark 72.1.4. This is a generalization of the Weyl form of the canonical commutation relations 54.1.3.

Definition 72.1.5 (Heisenberg system). The generators $\phi(z)$ of the maps $t \mapsto W(tz)$, which exist by Stone's theorem 23.4.26, are said to form a Heisenberg system. These operators satisfy the following properties:

- $\lambda\phi(z) = \phi(\lambda z)$ for all $\lambda > 0$,
- $[\phi(z), \phi(z')] = -i\omega(z, z')$, and
- $\phi(z + z')$ is the closure 23.4.18 of $\phi(z) + \phi(z')$.

72.2 Perturbative AQFT

Perturbative algebraic quantum field theory (pAQFT) tries to formalize the notion of perturbation theory in QFT as is is nowadays used in particle physics and high energy physics by

[2] 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 sets in $\overline{O}$.

[3] This could potentially cause confusion with other notations used in this text. **Loc** here has nothing to do with the category of locales from Chapter 7.

combining the formalism of AQFT with the causal pertubartion theory as introduced by *Epstein* and *Glaser*.

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 17.1.1. Here one can use the fact that every manifold is σ-compact and admits a connection. Moreover, as before, the spacetime manifolds M will always be assumed to be globally hyperbolic, i.e. admit a Cauchy surface.

72.2.1 Functional analysis

To handle the dynamics, one first has to generalize the notion of Lagrangians and actions from Chapter 67. Instead of simply considering a local action functional $S : \mathcal{E} \to \mathbb{R}$, one uses a generalization:

$$S : C_c^\infty(M) \to \mathcal{F}_{\text{loc}}(\mathcal{E}), \tag{72.4}$$

i.e. one uses local functional-valued distributions on M. Because local functionals are of the form

$$F : \phi \mapsto \int_M (j^\infty\phi)^* L \,\text{Vol}, \tag{72.5}$$

the standard example of such generalized action functionals are simply given by taking a product:

$$S[f] : \phi \mapsto \int_M f(j^\infty\phi)^* L \,\text{Vol}\,. \tag{72.6}$$

This approach has to benefits. First of all, one does not have to restrict to compact sets to have a well-defined integration by parts formula, since all objects under the integral 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 72.2.1 (Functional approach vs variational bicomplex). In Chapter 67, 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 nonlocal functionals. For completeness, algebraically, a local functional is characterized by the following conditions:

1. For all ϕ, ψ and $\chi \in \mathcal{E}$:

$$F(\phi + \psi + \chi) = F(\phi + \psi) - F(\psi) + F(\psi + \chi) \tag{72.7}$$

   whenever $\text{supp}(\phi) \cap \text{supp}(\chi) = \emptyset$.

2. The wave front set 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 72.2.2 (Graded functional). One can show that 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 the spaces of sections of the individual bundles:

$$\Gamma(E^{\boxtimes k})^* \cong \big(\Gamma(E)^*\big)^{\widehat{\otimes}_\pi k}. \tag{72.8}$$

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:

$$\mathcal{F}(\mathcal{E}) := C^\infty\left(V_0, \bigoplus_{\substack{k,l=0\\ i_1+\cdots i_l=k}} \Gamma_{i_1|\underline{i_2}|i_3|\ldots}\Big(\boxtimes_{j=1}^l E[-j]_0^{\boxtimes i_j}\Big)^*\right), \tag{72.9}$$

where the sequence $i_1|\underline{i_2}|i_3|\ldots$ denotes symmetric and antisymmetric tensor powers (underlined indices denote antisymmetric tensors). Note that because of the desuspension operator 27.1.7, 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:

$$\mathcal{F}(\mathcal{E}) = C^\infty\left(V_0, \bigoplus_{\substack{k,l=0\\ i_1+\cdots i_l=k}} \Gamma_{i_1|i_2|i_3|\ldots}\Big(\boxtimes_{j=1}^l E_j^{\boxtimes i_j}\Big)^*\right), \tag{72.10}$$

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 23.3.12, one can now also define a "graded derivative":

Formula 72.2.3 (Graded derivative). Denote the functional target space

$$\bigoplus_{\substack{l=0\\ i_1+\cdots i_l=k}} \Gamma_{i_1|\underline{i_2}|i_3|\ldots}\Big(\boxtimes_{j=1}^l E[-j]_0^{\boxtimes i_j}\Big)^*$$

by $\mathcal{A}^k(\mathcal{E})$ and consider a homogeneous element $F \in \mathcal{A}^k(\mathcal{E})$. The left (graded) derivative is defined as follows:

$$\begin{cases} \left\langle \dfrac{\delta^l F}{\delta\phi}(v), w\right\rangle := F(w\wedge v) & k>0 \\[2ex] \dfrac{\delta^l F}{\delta\phi} := 0 & k=0. \end{cases} \tag{72.11}$$

Similarly, the right (graded) derivative is defined by

$$\begin{cases} \left\langle \dfrac{\delta^r F}{\delta\phi}(v), w\right\rangle := F(v\wedge w) & k>0 \\[2ex] \dfrac{\delta^r F}{\delta\phi} := 0 & k=0. \end{cases} \tag{72.12}$$

72.2.2 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:

$$\langle \delta_{\mathrm{EL}} L(\phi), \psi \rangle := \langle L'[f](\phi), \psi \rangle, \tag{72.13}$$

for any cutoff function f on $\mathrm{supp}(\psi)$.

Remark 72.2.4 (Functoriality). A more categorical formulation is that a generalized Lagrangian is a natural transformation

$$L : C_c^\infty \Rightarrow \mathcal{F}_{\mathrm{loc}} \tag{72.14}$$

that satisfies its own locality conditions:

1. **Additivity**: $L[f+g+h] = L[f+g] - L[g] + L[g+h]$ if $\mathrm{supp}(f) \cap \mathrm{supp}(h) = \emptyset$, and
2. **Support**: $\mathrm{supp}(L[f]) \subseteq \mathrm{supp}(f)$.

To obtain an action functional S, one quotients out by the relation

$$\mathrm{supp}(L_1[f] - L_2[f]) \subset \mathrm{supp}(df). \tag{72.15}$$

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 second variational derivative of the action then satisfies:

$$\langle \delta^2 S(\phi), \psi_1 \otimes \psi_2 \rangle = \langle L''[f](\phi), \psi_1 \otimes \psi_2 \rangle, \tag{72.16}$$

where f is now a cutoff function on $\mathrm{supp}(\psi_1) \cup \mathrm{supp}(\psi_2)$. 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 17.2.6 this gives rise to a linear operator $P_S(\phi) : \mathcal{E} \to \mathcal{E}^*$.

Axiom 72.6 (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 $\Delta_{S,\phi}^{R/A} : \mathcal{E}_c^* \to \mathcal{E}$:

$$P_S(\phi) \circ \Delta_{S,\phi}^{R/A} = \mathbb{1}_{\mathcal{E}_c^*} \tag{72.17}$$

$$\Delta_{S,\phi}^{R/A} \circ P_S(\phi) = \mathbb{1}_{\mathcal{E}_c} \tag{72.18}$$

$$\mathrm{supp}(\Delta_{S,\phi}^{R}(\psi)) \subset J^-(\mathrm{supp}(\psi)) \tag{72.19}$$

$$\mathrm{supp}(\Delta_{S,\phi}^{A}(\psi)) \subset J^+(\mathrm{supp}(\psi)). \tag{72.20}$$

These operator also define integral kernels through the kernel theorem:

$$\Delta_{S,\phi}^R(x,y) = \Delta_{S,\phi}^A(y,x). \tag{72.21}$$

Their difference gives the causal propagator:

$$\Delta_{S,\phi} := \Delta_{S,\phi}^R - \Delta_{S,\phi}^A. \tag{72.22}$$

Property 72.2.5. If the action S is a quadratic functional, the operator $P_S(\phi)$ is constant:

$$\forall\phi\in\mathcal{E}: P_S(\phi)=P. \tag{72.23}$$

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 67 this will be the Peierls bracket:

Definition 72.2.6 (Peierls bracket). Consider two multilocal functionals $F,G\in\mathcal{F}_{\text{ml}}(\mathcal{E})$ together with an action functional S.

$$\{F,G\}_S(\phi):=\big\langle F'(\phi),\Delta_{S,\phi}G'(\phi)\big\rangle. \tag{72.24}$$

The main problem with the Peierls bracket is that it does not define a Poisson algebra structure on $\mathcal{F}_{\text{ml}}(\mathcal{E})$ 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 72.2.7. The integral kernel of the causal propagator has the following wave front set:

$$\text{WF}(\Delta_{(S,\phi)})=\{(x,k,x',-k')\in T^*M^2\mid(x,k)\sim(x',k')\}, \tag{72.25}$$

where $\sim$ indicates that the points are connected by the lift of a null geodesic whose fibrewise projection is given by the metric dual:

$$(x(t),k(t))\equiv(\gamma(t),g(\dot\gamma(t,\cdot))) \tag{72.26}$$

for some null geodesic γ on M.

Definition 72.2.8 (Microcausal functionals). A functional $F\in\mathcal{E}^*$ that is compactly supported and such that

$$\text{WF}(F^{(n)}(\phi))\subset\Xi_n \tag{72.27}$$

for all $n\in\mathbb{N}$ and $\phi\in\mathcal{E}$, where

$$\Xi_n:=T^*M^n\backslash\{(x,k)\in T^*M^n\mid k\in\overline{V}^+(x)\cup\overline{V}^-(x)\}. \tag{72.28}$$

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 72.2.9. 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 72.2.10 (Wightmann propagator). Consider the causal propagator Δ_S. By Property 72.2.7 its wave front set consists of two part lying in the closed past and closed future lightcones. This corresponds to the following decomposition:

$$\frac{i}{2}\Delta_S=\Delta_S^H-H, \tag{72.29}$$

where

$$\text{WF}(\Delta_S^H)=\{(x,k,x',-k')\in T^*M^2\mid(x,k)\sim(x',k')\wedge k\in\overline{V}^+(x)\} \tag{72.30}$$

and

$$H(x,y)=H(y,x). \tag{72.31}$$

The Wightmann or **Hadamard** propagator represents the positive-frequency part of the causal propagator.

72.2.3 Symmetries

Now, to include symmetries, one has to introduce vector fields. On general infinite-dimensional manifolds (Section 42.1) 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 then simply $C^\infty(\mathcal{E},\mathcal{E})$.

Definition 72.2.11 (Multilocal multivector fields). The space of multivector fields is now simply defined as the space of functionals on the odd cotangent bundle:

$$\mathfrak{X}_{\text{loc}}(\mathcal{E}) := \mathcal{F}_{\text{loc}}\big(\mathcal{E}\oplus\mathcal{E}^*[1]\big). \tag{72.32}$$

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 $\mathcal{F}_{\text{ml}}(\mathcal{E})$-module of the local vector fields, are then de 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

$$\partial_X F(\phi) := \langle F'(\phi), X(\phi)\rangle. \tag{72.33}$$

In physics literature this is often formally written as

$$\partial_X F(\phi) = \int_M X_x(\phi)\frac{\delta F}{\delta\phi(x)}. \tag{72.34}$$

Note that for any multilocal vector field, the following functional is trivial on the zero locus $\mathcal{E}_S$ of $\delta_{\text{EL}}S$:

$$\phi\mapsto\delta_S(X)(\phi) := \partial_X S(\phi) = \langle\delta_{\text{EL}}S(\phi), X(\phi)\rangle. \tag{72.35}$$

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:

$$\mathcal{F}_S := \mathcal{F}_{\text{ml}}/\delta_S(\mathfrak{X}^1_{\text{ml}}). \tag{72.36}$$

As in Chapters 49 and 67 one can now characterize this space in homological terms. For this one constructs the Koszul complex

$$\cdots\xrightarrow{\delta_S}\Lambda^2\mathcal{E}\xrightarrow{\delta_S}\mathcal{E}\xrightarrow{\delta_S}\mathcal{F}_{\text{ml}}\longrightarrow 0, \tag{72.37}$$

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 finish 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. Expression 67.3.2 for the antifield (ignoring the ghosts and antighosts for now) then corresponds to the expression

$$\{X,Y\}(\phi) := \left\langle\frac{\delta^r X}{\delta\phi},\frac{\delta^l Y}{\delta\phi^\dagger}\right\rangle - \left\langle\frac{\delta^r X}{\delta\phi^\dagger},\frac{\delta^l Y}{\delta\phi}\right\rangle, \tag{72.38}$$

where the derivatives with respect to ϕ and $\phi^\dagger$ should be interpreted as functional (Gateaux) and graded derivatives 72.2.3 with respect to the base and fibre coordinates of the odd cotangent bundle $\mathcal{E}\oplus\mathcal{E}^*[1]$.

?? ADD SCHOUTEN BRACKET TO CHAPTER ON VECTOR BUNDLES ??

72.2.4 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 57.2.2 is applicable:

$$F * G := \mu \circ \exp(\tfrac{i\hbar}{2}\pi_\Delta)(F \otimes G), \tag{72.39}$$

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:

$$F *_H G := \mu \circ \exp(\hbar\pi_{\Delta^H})(F \otimes G), \tag{72.40}$$

Property 72.2.12. On the space of regular functionals, the causal and Wightmann star products are isomorphic:

$$F *_H G = \alpha_H(\alpha_H^{-1}F * \alpha_H^{-1}G), \tag{72.41}$$

where

$$\alpha_H := \exp\left(\frac{\hbar}{2}\left\langle H, \frac{\delta^2}{\delta\phi^2}\right\rangle\right). \tag{72.42}$$

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 72.2.13 (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

$$F^* := \alpha_H^{-1}\left(\overline{\alpha_H F}\right). \tag{72.43}$$

The support of an element $F \in \mathfrak{U}^H$ is given by

$$\text{supp}(F) := \text{supp}(\alpha_H F). \tag{72.44}$$

Remark 72.2.14. 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:

$$P_x(H - H')(x, y) = P_y(H - H')(x, y) = 0. \tag{72.45}$$

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 72.2.15. 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

$$\mathfrak{U}_{S_0} := \mathfrak{U}/\mathfrak{J}_{S_0}, \tag{72.46}$$

where $\mathfrak{J}_{S_0}$ is generated by the functionals $\langle \delta_{\text{EL}} S, X\rangle$ for $X \in \mathfrak{X}_{\text{loc}}$. The Moyal product is extended to microcausal multivector fields by replacing the pointwise multiplication in (72.40) by the wedge product:

$$\begin{aligned}\langle X *_H Y(\phi), v_1 \otimes \cdots \otimes v_{p+q}\rangle :=& \qquad\qquad (72.47)\\ \sum_{i=0}^{\infty} \frac{\hbar^i}{i!p!q!} \sum_{\sigma\in S_{p+q}} \left\langle\!\!\left\langle \frac{\delta^i X}{\delta\phi^i}(\phi), v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(p)}\right\rangle, (\Delta^H_{S_0})^{\otimes i}\left\langle \frac{\delta^i Y}{\delta\phi^i}(\phi), v_{\sigma(p+1)} \otimes \cdots \otimes v_{\sigma(p+q)}\right\rangle\!\!\right\rangle.\end{aligned}$$

?? COMPLETE ??

72.3 Topological QFT

For convenience the reader is reminded of some of the notations that are used. **FinVect** will denote the category of finite-dimensional vector spaces over $\mathbb{C}$ and $\mathbf{Bord}^d_{d-1}$ denotes the category of d-dimensional (framed[4]) cobordisms 29.4.6.

72.3.1 Atiyah-Segal axioms

Axiom 72.7 (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. **Disjoint union**: $F(M \sqcup M') = F(M) \otimes F(M')$.
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 $F \circ f = F$.

In the above conditions M, M' are d-dimensional cobordisms between $(d-1)$-dimensional (closed) smooth manifolds.

Example 72.3.1 (1D). In $d = 1$ a TQFT functor gives 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 72.3.2 (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 20.7.6.

[4] A framing in this context means a framing of the d-stabilized tangent bundles $TM \oplus \mathbb{R}^{d-\dim(M)}$.

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 72.3.3 (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.

Remark 72.3.4. One can replace the category **FinVect** by any other symmertric monoidal category $\mathbf{C}$.

The following conjecture by *Baez* and *Dolan*, later proven by *Lurie*, 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.

Theorem 72.3.5 (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* $F(\mathrm{pt}) \in \mathrm{ob}(\mathbf{C})$.

72.3.2 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 29.4). For simplicity only the case of 2D TQFTs (as in the original definition) will be considered.

?? 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 4.4.54 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 by *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:

$$\mathrm{Fields}(M) \cong [\textstyle\int M, \mathbf{Fields}], \tag{72.48}$$

where **Fields** is the classifying stack and $\int$ sends a manifold to its homotopy type, i.e. it is the shape modality 13.8.7.

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 $\mathbf{H} \to \mathrm{U}(1)$, to its boundary components it should assign a gauge field, i.e. a map $\mathbf{H} \to \mathbf{B}\mathrm{U}(1)$, etc. Combining this with the pQFT, one obtains a functor $\mathbf{Bord}^d \to \mathbf{Corr}_d(\mathbf{H}/_{\mathbf{B}^d\mathrm{U}(1)})$ 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 $\mathbf{Bord}^d \rightarrow \mathbf{Corr}_d(\mathbf{H})$ is a lift (of monoidal functors) to a symmetric monoidal (∞, n)-functor $\mathbf{Bord}^d \rightarrow \mathbf{Corr}_d(\mathbf{H}/_{\mathbf{B}^d\mathrm{U}(1)})$.

By the cobordism hypothesis, such a functional is uniquely defined by a morphism $\mathbf{Fields} \rightarrow \mathbf{B}^d\mathrm{U}(1)$, 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 $U \rightarrow [\int M, \mathbf{B}^d\mathrm{U}(1)]$ 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 $[\int M, \mathbf{B}^d\mathrm{U}(1)] \rightarrow \mathbf{B}^{d-k}\mathrm{U}(1)$, where k is the dimension of U. Composition with the classifying functor $\mathbf{Fields} \rightarrow \mathbf{B}^d\mathrm{U}(1)$ 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

Extensions and Unification

The main reference for supersymmetric quantum mechanics is the seminal paper by *Witten* [78]. For an introduction to algebraic superstructures, see Section 27.1.

73.1 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 while studying supersymmetry.

73.1.1 Supersymmetric quantum mechanics

In this section a general graded Hilbert space $\mathcal{H}$ will be considered. This space is equipped with an algebra of bounded operator $A \subset \mathcal{B}(\mathcal{H})$ together with a set of N odd self-adjoint operators $\{D^i\}_{i\leq N}$. More precisely, a spectral triple $(\mathcal{H}, A, \{D^i\}_{i\leq N})$ is considered (Definition 41.2.1).

This data defines an SQM system if the Hamiltonian H satisfies the following condition:

$$\{D^i, D^j\}_+ = 2\delta^{ij}H. \tag{73.1}$$

For $N = 2$ the whole theory can be rephrased in terms of a nilpotent operator $d \sim D^1 + iD^2$ and its adjoint:

$$\{d, d^\dagger\}_+ \sim H. \tag{73.2}$$

Example 73.1.1 (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

$$\langle\alpha|\beta\rangle = \int_M \alpha \wedge *\beta. \tag{73.3}$$

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:

$$d^W := e^{-W(t)} \circ d \circ e^{W(t)} \tag{73.4}$$

$$(d^W)^\dagger := e^{W(t)^\dagger} \circ d^\dagger \circ e^{-W(t)^\dagger}. \tag{73.5}$$

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, [79] considers a loop space ΩM (a theory of closed strings), where the the deformation operator at a point $\gamma \in \Omega M$ is given by

$$W(\gamma)\omega := \int_\gamma dt\ B_{\mu\nu}(\gamma(t))dx^\mu(t) \wedge dx^\nu(t) \wedge \omega(t), \tag{73.6}$$

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.

73.1.2 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 $T_p\Omega M$ is the space of vector fields along the path p. As such they 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:

$$\partial_{\mu,\sigma} := \frac{\delta}{\delta X^\mu(\sigma)}, \tag{73.7}$$

where $X^\mu(\sigma)$ is the μ^{th} coordinate of the loop at the parameter $\sigma \in [0, 2\pi[$.

73.1.3 Extensions of the Standard Model

Theorem 73.1.2 (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.*

Remark. In other words, it is impossible to combine the Poincaré group in a nontrivial way with the internal symmetry 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 71. CFTs do not admit an S-matrix and, hence, the above theorem is clearly not 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 27.7.2. 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*.

[1] Technically its universal cover should be considered.

[2] See the original paper [105] for why exactly the algebra plays an essential role.

73.2 Chern-Simons theory

73.2.1 Holomorphic Chern-Simons theory

Consider a complex manifold of dimension 3 equipped with a holomorphic volume form Vol, i.e. a Calabi-Yau three-fold 37.3.13. For every complex vector bundle $E \to M$ and connection ∇ on E, one can consider the antiholomorphic connection one-form B. This form induces the following holomorphic Chern-Simons action:

$$S_{\text{CS}}[B] := \int_M \left(\langle B, \overline{\partial} B \rangle + \frac{2}{3} \langle B, [B \wedge B] \rangle \right) \wedge \text{Vol}_M. \tag{73.8}$$

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 37.2.7 these correspond exactly to the holomorphic structures on E.

73.3 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:

$$Q([\omega], [\nu]) := \int_M \omega \wedge \nu. \tag{73.9}$$

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: $\text{rk}(Q) = q_+ + q_- = b_2(M)$. 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 corresponds 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

$$p_1(M) = 2\sigma(M). \tag{73.10}$$

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 Equation (33.91):

$$c_1^2(M) = 2\chi(M) + 3\sigma(M). \tag{73.11}$$

73.3.1 Donaldson-Witten theory

In four dimensions the classification of manifolds is considerably more intricate than in higher dimensions (cf. Remark 29.1.11). To find invariants *Donaldson* studied instanton configurations of non-Abelian gauge theories on 4-manifolds.

[3]For topological manifolds and odd intersection forms, two classes exist.

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 $\mathrm{SU}(n)$-bundle $P \to M$. In this case the second Chern number reads (by Chern-Weil theory)

$$c_2(P)[M] = \int_M \frac{\mathrm{tr}(F^2) - \mathrm{tr}(F)^2}{8\pi^2}. \tag{73.12}$$

Because elements of $\mathfrak{su}(n)$ are traceless, this reduces to

$$c_2(P)[M] = \int_M \frac{\mathrm{tr}(F^2)}{8\pi^2} = -\frac{1}{8\pi^2}\int_M \left(\|F^+\|^2 - \|F^-\|^2\right)\mathrm{Vol}_M, \tag{73.13}$$

where the curvature was decomposed in a selfdual and anti-selfdual part $F^\pm$ and the Killing metric $K(X,Y) = -\mathrm{tr}(XY)$ is used. Now, consider the Yang-Mills action

$$\begin{aligned} S_{\mathrm{YM}}[A] &= -\int_M \mathrm{tr}(F \wedge *F) \\ &= \int_M \|F\|^2\,\mathrm{Vol}_M \\ &= \int_M \left(\|F^+\|^2 + \|F^-\|^2\right)\mathrm{Vol}_M. \end{aligned}$$

gather So, 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 minimzation 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 constitutes 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 73.3.1. 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:

$$\mathcal{M}^\pm(P) := \{A \in \mathcal{A}(P) \mid *F_A = \pm F_A\} \,/\, \mathrm{Aut}_V(P). \tag{73.14}$$

One should, however, 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. The stabilizer of a connection is given by the covariantly constant sections of $\mathrm{Aut}_V(P)$. 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 $\mathcal{M}^\pm$.

According to Equation (33.64) one can characterize the stabilizer of A as those sections of $\mathrm{Aut}_V(P)$ that are covariantly constant (with respect to A). The Lie algebra of this group is

[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.

given by $\Gamma(\mathrm{ad}(P))$, hence the reducible connections are those that correspond to a nonzero kernel of

$$\nabla^A : \Gamma(\mathrm{ad}(P)) \to \Omega^1(M; \mathrm{ad}(P)), \tag{73.15}$$

where the codomain is exactly the model space of the affine space of connections $\mathcal{A}(P)$ and, accordingly, also the (model) tangent space $T_A\mathcal{A}(P)$.

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:

$$\Omega^1(M; \mathrm{Ad}(P)) = \mathrm{im}(\nabla) \oplus \ker(\nabla^\dagger). \tag{73.16}$$

This shows that the neighbourhood of A (consisting of irreducible connections) is modelled on $\ker(\nabla^\dagger)$. Moreover, by linearizing the condition for F_{A+a} to be (anti-)selfdual, where $a \in \Omega^1(M; \mathrm{ad}(P))$, one obtains the condition

$$\mathrm{pr}^\mp(\nabla^A a) = 0. \tag{73.17}$$

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 $\mathrm{pr}^\mp\nabla^A$ since $\nabla^2 = F$. This gives rise to the **Atiyah-Hitchin-Singer complex** or **instanton deformation complex** $C^\bullet_{\mathrm{AHS}}$:

$$0 \longrightarrow \Gamma(\mathrm{ad}(P)) \xrightarrow{\nabla^A} \Omega^1(M; \mathrm{ad}(P)) \xrightarrow{\mathrm{pr}^\mp\nabla^A} \Omega^{2,\pm}(M; \mathrm{ad}(P)) \longrightarrow 0. \tag{73.18}$$

By the above arguments, one obtains the following model:

$$T_{[A]}\mathcal{M}^\pm \cong H^1(C^\bullet_{\mathrm{AHS}}). \tag{73.19}$$

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...) ??

73.3.2 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) ASD 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 29.5.3, 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** $\mathrm{sf}(A)$ 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:

$$\|F_A\|_2^2 := \int_{\mathbb{R}\times M} \|F_A\|^2 \,\mathrm{Vol}_M < +\infty. \tag{73.20}$$

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 $\mathrm{sf}(A_-, A_+) - 1$. The boundary operator is defined as follows:

$$\partial A_- := \sum_{\substack{A_+ \in \mathcal{M}^-_{\mathrm{flat}}(P) \\ \mathrm{sf}(A_-, A_+)=1}} \left|\overline{\mathcal{M}}(A_-, A_+)\right| \langle A_+ \rangle. \tag{73.21}$$

The resulting homology theory is called **instanton Floer homology**.

?? COMPLETE ??

73.3.3 Seiberg-Witten theory

In this section M will denote an oriented, Riemannian 4-manifold equipped with a $\mathrm{Spin}^{\mathbb{C}}$-structure $P \to M$. It can be shown that $\mathrm{Spin}^{\mathbb{C}}(4)$ is isomorphic to $\mathrm{SU}(2) \times \mathrm{SU}(2) \times \mathrm{U}(1)/\mathbb{Z}_2$. This group admits two two-dimensional representations

$$s_+ : \mathrm{Spin}^{\mathbb{C}}(4) \to \mathrm{U}(2) : [s^+, s^-, s] \mapsto [s_+, s] \tag{73.22}$$
$$s_- : \mathrm{Spin}^{\mathbb{C}}(4) \to \mathrm{U}(2) : [s^+, s^-, s] \mapsto [s_-, s] \tag{73.23}$$

and the one-dimensional determinant representation

$$\det : \mathrm{Spin}^{\mathbb{C}}(4) \to \mathrm{U}(1) : [s^+, s^-, s] \mapsto s^2. \tag{73.24}$$

These representations in turn induce associated $\mathrm{Spin}^{\mathbb{C}}(4)$-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 some representation-theoretic arguments): $\Lambda^2 S^\pm \cong \det(P)$. Moreover, because the representations are unitary, these bundles admit a (Hermitian) metric.

Any connection on $\det(P)$ induces a connection on an associated $\mathrm{Spin}^{\mathbb{C}}(4)$-bundle after addition by the Levi-Civita connection, since

$$\mathrm{Lie}(\mathrm{Spin}^{\mathbb{C}}(4)) \cong \mathrm{Lie}(\mathrm{SO}(4)) \oplus \mathrm{Lie}(\mathrm{U}(1)).$$

So choose such a connection A and choose a self-dual 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

$$D\psi = 0 \tag{73.25}$$
$$F_A^+ = q(\psi) + i\mu, \tag{73.26}$$

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 73.3.2 (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:

$$\psi \mapsto \psi \otimes \psi^\dagger - \frac{1}{2}|\psi|^2. \tag{73.27}$$

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]

$$\iota_{C\ell}(F_A^+ - i\mu) = \psi\otimes\psi^\dagger - \frac{1}{2}|\psi|^2. \tag{73.26b}$$

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]

$$q(\psi)\sim\frac{1}{2}\sum_{i\neq j}\langle\psi\mid\gamma_i\gamma_j\cdot\psi\rangle dx^i\wedge dx^j. \tag{73.28}$$

A third approach uses the isomorphism $\mathrm{SU}(2)\cong\mathrm{Sp}(1)$ and identifies the $\mathbb{C}^2$-bundles with $\mathbb{H}$-bundles. Imaginary two-forms can then be identified with sections of an $\mathrm{Im}(\mathbb{H})$-subbundle, where $\mathrm{Im}(\mathbb{H}) := \{x\in\mathbb{H}\mid\overline{x}=-x\}$. The squaring map is induced by the following morphism:

$$q:\mathbb{H}\to\mathrm{Im}(\mathbb{H}): x\mapsto\frac{1}{2}\overline{x}ix. \tag{73.29}$$

Remark 73.3.3 (Hyper-Kähler manifolds). The study of Seiberg-Witten equations can be widely generalized exactly due to the relation between $\mathrm{Spin}^{\mathbb{C}}(4)$ and the quaternions $\mathbb{H}$ 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 $\mathrm{U}(1)$ acts on $\mathbb{H}$ by scalar multiplication and preserves the hyper-Kähler structure. The associated moment map is exactly the morphism inducing the squaring map:

$$\mu:\mathbb{H}\to\mathrm{Sp}(1): x\mapsto\frac{1}{2}\overline{x}ix. \tag{73.30}$$

Property 73.3.4 (Weitzenböck identity). The Lichnerowicz formula 34.3.25 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]

$$D^2 = \nabla^*\nabla + \frac{1}{4}R + \frac{1}{4}F. \tag{73.31}$$

In the case of a proper $\mathrm{Spin}(4)$-structure on M, the half-spinor bundles $S^\pm$ can be identified with the tensor product $S\otimes\det^{1/2}(P)$. In this case the Dirac operator on these bundles satisfy the above Weitzenböck identity. However, if only a $\mathrm{Spin}^{\mathbb{C}}(4)$-structure exists, the Dirac operator cannot be obtained as a twisted operator, but it still satisfies this identity.

[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.

[7] In general one used the term **Weitzenböck identity** 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.

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 $\mathrm{Spin}^{\mathbb{C}}$-structure). As usual, this space admits some gauge symmetries. Any function $g : M \to \mathrm{U}(1)$ acts on the connection ∇^L by ordinary gauge transformations:

$$A \longrightarrow A - 2g^{-1}dg \qquad\qquad \psi \longrightarrow g\psi. \tag{73.32}$$

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, expect at elements with $\psi = 0$, which have a $\mathrm{U}(1)$ stabilizer. To remove these gauge symmetries, one should pass to a suitable quotient. Either by all of $C^\infty(M, \mathrm{U}(1))$ 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 73.3.5 (Moduli space). The moduli spaces $\mathcal{M}$ and $\mathcal{M}_0$ satisfy the following properties:

- $\mathcal{M}$ 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 $\mathrm{U}(1)$-bundle.
- For almost all μ, $\mathcal{M}$ is orientable and its dimension is given by

  $$\dim(\mathcal{M}) = \dim(H^1(M)) - 1 - q_+ + \frac{Q(c_1(\det(P)), c_1(\det(P))) - q_+ + q_-}{4}, \tag{73.33}$$

  where (q_+, q_-) is the signature of Q.

73.4 String theory

73.4.1 Worldsheets

Similar to how a relativistic particle is characterized by an action that is proportional to the arc length of its worldline (65.13), i.e.

$$S_{\text{point}} = -mc\int_\gamma ds, \tag{73.34}$$

the action of the relativistic string is proportional to the area of its worldsheet:

Formula 73.4.1 (Nambu-Goto action).

$$S_{\text{NG}} := -T_0\int_\Sigma d\Sigma = -\frac{1}{2\pi\alpha'}\int_\Sigma d\Sigma, \tag{73.35}$$

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 73.4.2 (Polyakov action).

$$S_{\text{Polyakov}} := \frac{T_0}{2}\int_\Sigma \sqrt{-h}h^{ab}g_{\mu\nu}(x)\partial_a x^\mu \partial_b x^\nu d^2\tau. \tag{73.36}$$

Property 73.4.3 (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 on the worldsheet $h \longrightarrow \Omega h$. It follows that the stress-energy tensor

$$T_{ab} := \frac{-2}{\sqrt{-h}}\frac{\delta S}{\delta h_{ab}} \tag{73.37}$$

is traceless.

73.4.2 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 27.4.6 and an internal special symmetric Frobenius algebra 26.1.7. 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, ...) ??

73.4.3 Duality

?? ADD (S, T, Montonen-Olive, ...) ??

The first type of duality that one can consider is so-called T-duality. In its simplest form this the string spectrum on two torus bundles. Consider two spacetime manifolds (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 by differential cohomology (Section 33.7) 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:

$$\pi_*\alpha = c_1(P') \qquad\qquad \pi'_*\alpha' = c_1(P), \tag{73.38}$$

where c_1 denotes the first Chern class. If this condition is satisfied, there exists an isomorphism between twisted K-theories:

$$K^\bullet_\alpha(P) \cong K^{\bullet-1}_{\alpha'}(P'). \tag{73.39}$$

73.4.4 Superstrings

?? ADD (GSO, 5 theories, ...) ??

73.5 M-theory

?? ADD ??

Chapter 74

Epilogue

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 should also add 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
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
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

Part XI

Appendices

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	$\mathrm{SL}(n,\mathbb{R})$	$\mathrm{GL}^+(n,\mathbb{R})$ is sufficient for orientability. The special linear group gives rise to a volume form.
Riemannian metric	$\mathrm{O}(n)$	
Almost-symplectic structure*	$\mathrm{Sp}(n,\mathbb{R})$	Integrability (in the form of a closed form) gives a symplectic manifold.
Almost-complex structure*	$\mathrm{GL}(k,\mathbb{C})$	Integrability (in the sense of Newlander-Nirenberg) gives a complex manifold.
Almost-Hermitian structure*	$\mathrm{U}(k)$	Integrability gives a Kähler manifold.
Calabi-Yau*	$\mathrm{SU}(k)$	
Hyperkähler**	$\mathrm{Sp}(k)$	Hyperkähler implies Calabi-Yau.
Almost quaternionic**	$(\mathrm{GL}(k,\mathbb{H})\times\mathbb{H}^\times)/\mathbb{R}^\times$	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**	$(\mathrm{Sp}(k)\times\mathrm{Sp}(1))/\mathbb{Z}_2$	These manifolds are not strictly Kähler since the structure group is not a subgroup of $\mathrm{U}(2k)$.

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 *nonsymmetric*) Riemannian manifolds by *Berger*. (The $\mathrm{SL}(n,\mathbb{R})$-structure is technically not part of the original classification since it is not a subgroup of $\mathrm{O}(n)$ 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 [37].

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 $FM \rightarrow \Lambda^2 V^* \otimes \mathfrak{g}$ 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 hyperkähler manifolds are a specific class of quaternionic Kähler manifolds since $\mathrm{Sp}(k)$ can be embedded in $\mathrm{Sp}(k) \cdot \mathrm{Sp}(1)$. To exclude this class, one can just require the holonomy group to be all of $\mathrm{Sp}(k) \cdot \mathrm{Sp}(1)$. 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 Einstein 34.2.11. The hyperkähler 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 the Schwarz theorem 14.6.10, 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

$$T^{\mu}_{\ \nu} := F^{\mu\lambda}F_{\nu\lambda} - \frac{1}{4}\delta^{\mu}_{\nu}F^{\lambda\kappa}F_{\lambda\kappa} + \mathrm{eom} \tag{B.4}$$

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

[1]Pseudotensor here just means an object that does not transform tensorially.

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.
- 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 quasi-local 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:

$$\delta S = -\int_M d^4x E^{\mu\nu}\delta g_{\mu\nu} + \frac{1}{2}\int_{\Sigma_\pm} d^3x\sqrt{h}P_{ij}\delta h_{ij} + \frac{1}{2}\int_N d^3x\sqrt{\gamma}\tau^{ij}_{\text{BY}}\delta\gamma_{ij} + \text{matter terms}. \tag{B.10}$$

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:

$$\tau^{ij}_{\text{BY}} = -\frac{1}{\kappa\sqrt{\gamma}}(\Theta\gamma^{ij} - \Theta^{ij}) + \text{regularizing terms}. \tag{B.11}$$

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 symbols are used throughout the summary:

Abbreviations

AIC	Akaike information criterion
ARMA	autoregressive moving-average model
BCH	Baker-Campbell-Hausdorff
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
EPR	Einstein-Podolsky-Rosen
ETCS	Elementary Theory of the Category of Sets
FWHM	full width at half maximum
GA	geometric algebra
GHZ	Greenberger-Horne-Zeilinger
GNS	Gel'fand-Naimark-Segal
HoTT	Homotopy Type Theory
KKT	Karush-Kuhn-Tucker
LIVF	left-invariant vector field
MPO	matrix product operator
MPS	matrix product state
MTC	modular tensor category
NDR	neighbourhood deformation retract
OPE	operator product expansion
OZI	Okubo-Zweig-Iizuka
PAC	probably approximately correct
PL manifold	piecewise-linear manifold
PVM	projection-valued measure

RKHS	reproducing kernel Hilbert space
SVM	support-vector machine
TQFT	topological quantum field theory
VIF	variance inflation factor
ZFC	Zermelo-Frenkel set theory with the axiom of choice
TVS	topological vector space
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
$\deg(f)$	degree of the polynomial f
e	identity element of a group
$\Gamma(E)$	set of global sections of a fibre bundle E
Im	imaginary part of a complex number
$\mathrm{Ind}_f(z)$	index of a point $z \in \mathbb{C}$ with respect to a function f
$\hookrightarrow$	injective function
$\cong$	is isomorphic to
Par_t^γ	parallel transport map with respect to the curve γ
Re	real part of a complex number
Res	residue of a complex function
$\twoheadrightarrow$	surjective function
$\{\cdot,\cdot\}$	Poisson bracket
∂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 the sets X and Y
$X + Y$	sum of the vector spaces X and Y
$X \oplus Y$	direct sum of the vector spaces X and Y
$V \otimes W$	tensor product of the vector spaces V and W
$\mathbb{1}_X$	identity morphism on the object X
$\approx$	is approximately equal to
$\hookrightarrow$	is included in
$\cong$	is isomorphic to
$\mapsto$	mapsto
Collections	
Ab	category of Abelian groups
$\mathrm{Aut}(X)$	automorphism group of an object X
$\mathcal{B}_0(V,W)$	space of compact bounded operators between the Banach spaces V and W

$\mathcal{B}(V, W)$	space of bounded linear maps from the space V to the space W
CartSp	the category of Euclidean spaces and "suitable" homomorphisms (e.g. linear maps, smooth maps, ...)
$C_\bullet$	chain complex
$\mathbf{Ch}(\mathbf{A})$	category of chain complexes with objects in the additive category $\mathbf{A}$
$\mathbf{C}^\infty$	category of smooth spaces
$C_p^\infty(M)$	ring of smooth functions $f : M \to \mathbb{R}$ on a neighbourhood of $p \in M$
$C^\omega(V)$	the set of all analytic functions defined on the set V
$\mathrm{Conf}(M)$	conformal group of (pseudo-)Riemannian manifold M
$C(X, Y)$	set of continuous functions between two topological spaces X and Y
$\mathbf{C}^\infty\mathbf{Ring}, \mathbf{C}^\infty\mathbf{Alg}$	category of smooth algebras
Diff	category of smooth manifolds
DiffSp	category of diffeological spaces and smooth maps
D^n	standard n-disk
$\mathrm{dom}(f)$	domain of a function f
$\mathrm{End}(X)$	endomorphism monoid of a an object X
$\mathcal{E}\mathrm{nd}$	endomorphism operad
$\mathbf{FormalCartSp_{diff}}$	category of infinitesimally thickened Euclidean spaces
$\mathrm{GL}(V)$	general linear group, the group of automorphisms of a vector space V
$\mathrm{GL}(n, K)$	general linear group: the group of all invertible $n \times n$-matrices over the field K
Grp	category of groups and group homomorphisms
Grpd	category of groupoids
$\mathrm{Hol}_p(\omega)$	holonomy group at the point p with respect to the principal connection ω
$\mathrm{Hom}_{\mathbf{C}}(V, W)$	set of homomorphisms from an object V to an object W in a category $\mathbf{C}$
hTop	homotopy category
$\mathrm{im}(f)$	image of a function f
$K^0(X)$	K-theory over a (compact Hausdorff) space X
Kan	category of Kan complexes
$\mathcal{K}_n(A, v)$	Krylov subspace of dimension n generated by the matrix A and the vector v
L^1	space of integrable functions
Law	category of Lawvere theories
Lie	category of Lie groups
$\mathfrak{Lie}$	category of Lie algebras
$\mathfrak{X}^L$	space of left-invariant vector fields on a Lie group
LX	free loop space on X
$\mathbf{Man}^p$	category of C^p-manifolds
Meas	category of measure spaces and measure-preserving functions
$N\mathbf{C}$	the simplicial nerve of a small category $\mathbf{C}$
$\mathbf{Open}(X)$	category of open subsets of a topological space X
$\mathrm{O}(n, K)$	group of $n \times n$ orthogonal matrices over a field K
$P(S), 2^S$	power set of S

$\mathrm{Pin}(V)$	pin group of the Clifford algebra $C\ell(V,Q)$
$\mathbf{Psh}(\mathbf{C}), \widehat{\mathbf{C}}$	category of presheaves on a (small) category $\mathbf{C}$
$\mathbf{Sh}(X)$	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
$\mathrm{SL}_n(K)$	special linear group: group of all invertible n-dimensional matrices with unit determinant over the field K
S^n	standard n-sphere
$S^n(V)$	space of symmetric rank n tensors over a vector space V
$W^{m,p}(U)$	the Sobolov space in L^p of order m
$\mathbf{Span}(\mathbf{C})$	span category over $\mathbf{C}$
$\mathrm{Spec}(R)$	spectrum of a commutative ring R
$\mathrm{supp}(f)$	support of a function f
$\mathrm{Syl}_p(G)$	set of Sylow p-subgroups of a finite group G
S_n	symmetric group of degree n
$\mathrm{Sym}(X)$	symmetric group on the set X
$\mathrm{Sp}(n, K)$	group of matrices preserving a canonical symplectic form over the field K
$\mathrm{Sp}(n)$	compact symplectic group
$\mathrm{TL}_n(\delta)$	Temperley-Lieb algebra with $n-1$ generators and parameter δ.
T^n	standard n-torus (the n-fold Cartesian product of S^1)
Top	category of topological spaces
Topos	the 2-category of (elementary) topoi and geometric morphisms
$U(\mathfrak{g})$	universal enveloping algebra of a Lie algebra $\mathfrak{g}$
$\mathrm{U}(n, K)$	group of $n \times n$ unitary matrices over a field K
$\mathbf{Vect}(X)$	category of vector bundles over a manifold X
$\mathbf{Vect}_K$	category of vector spaces and linear maps over a field K
Y^X	set of functions from a set X to a set Y
$\emptyset$	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 X
$\Omega^k(M)$	$C^\infty(M)$-module of differential k-forms on the manifold M
$\rho(A)$	resolvent set of a bounded linear operator A
$\mathfrak{X}(M)$	$C^\infty(M)$-module of vector fields on the manifold M
Units	
C	coulomb
T	tesla

Bibliography

[1] Saunders Mac Lane. *Categories for the Working Mathematician*, volume 5. Springer Science & Business Media, 2013.

[2] Fosco Loregian. Coend calculus. arXiv:1501.02503.

[3] Gregory M. Kelly. *Basic Concepts of Enriched Category Theory*, volume 64. CUP Archive, 1982.

[4] John C. Baez and Peter May. *Towards Higher Categories*, volume 152 of *IMA Volumes in Mathematics and its Applications*. Springer, 2009.

[5] Tom Leinster. Basic bicategories. 1998. arXiv:math/9810017.

[6] Peter T. Johnstone. *Topos Theory*. Dover Publications, 2014.

[7] Sjoerd E. Crans. Localizations of transfors. 1998.

[8] Marek Capinski and Peter E. Kopp. *Measure, Integral and Probability*. Springer Science & Business Media, 2013.

[9] Yvonne Choquet-Bruhat, Cecile DeWitt-Morette, and Margaret Dillard-Bleick. *Analysis, Manifolds and Physics, Part 1: Basics*. North-Holland.

[10] Yvonne Choquet-Bruhat and Cecile DeWitt-Morette. *Analysis, Manifolds and Physics, Part 2*. North-Holland.

[11] Georgiev Svetlin. *Theory of Distributions*. Springer, 2015.

[12] The nlab. `https://ncatlab.org/nlab`.

[13] Thomas Augustin, Frank PA Coolen, Gert De Cooman, and Matthias CM Troffaes. *Introduction to imprecise probabilities*. John Wiley & Sons, 2014.

[14] Jonathan R. Shewchuk. An introduction to the conjugate gradient method without the agonizing pain. Technical report, 1994.

[15] Arindam Banerjee, Srujana Merugu, Inderjit S Dhillon, and Joydeep Ghosh. Clustering with Bregman divergences. *Journal of Machine Learning Research*, 6(Oct):1705--1749, 2005.

[16] Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *J. Mach. Learn. Res.*, 9:371-–421, 2008.

[17] Vladimir Vovk, Alex Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer Science & Business Media, 2005.

[18] Jean-Daniel Boissonnat, Frank Nielsen, and Richard Nock. Bregman Voronoi diagrams. *Discrete & Computational Geometry*, 44(2):281--307, 2010.

[19] Shun-ichi Amari. *Information Geometry and Its Applications.* Springer Publishing Company, Incorporated, 2016.

[20] Niclas Sandgren and Petre Stoica. On moving average parameter estimation. Technical Report 2006-022, Department of Information Technology, Uppsala University, 2006.

[21] Victor Chernozhukov, Kaspar Wüthrich, and Zhu Yinchu. Exact and robust conformal inference methods for predictive machine learning with dependent data. In *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 732--749. PMLR, 2018.

[22] Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338--353, 1965.

[23] Vladimir I. Arnol'd. *Mathematical Methods of Classical Mechanics*, volume 60. Springer Science & Business Media, 2013.

[24] Richard Palais. The symmetries of solitons. *Bulletin of the American Mathematical Society*, 34(4):339--403, 1997.

[25] Tetsuji Miwa, Michio Jimbo, Michio Jimbo, and E Date. *Solitons: Differential Equations, Symmetries and Infinite-dimensional Algebras*, volume 135. Cambridge University Press, 2000.

[26] Edwin J. Beggs and Shahn Majid. *Quantum Riemannian Geometry.* Springer, 2020.

[27] Marc Henneaux and Claudio Teitelboim. *Quantization of Gauge Systems.* Princeton university press, 1992.

[28] Mark Hovey. *Model Categories.* Number 63. American Mathematical Soc., 2007.

[29] Mukund Rangamani and Tadashi Takayanagi. *Holographic Entanglement Entropy.* Springer, 2017.

[30] Charles W. Misner, Kip S. Thorne, and John A. Wheeler. *Gravitation.* Princeton University Press, 2017.

[31] Carlo Rovelli and Francesca Vidotto. *Covariant Loop Quantum Gravity: An Elementary Introduction to Quantum Gravity and Spinfoam Theory.* Cambridge University Press, 2014.

[32] Richard W. Sharpe. *Differential Geometry: Cartan's Generalization of Klein's Erlangen Program*, volume 166. Springer Science & Business Media, 2000.

[33] John C. Baez, Irving E. Segal, and Zhengfang Zhou. *Introduction to Algebraic and Constructive Quantum Field Theory.* Princeton University Press, 2014.

[34] Raoul Bott and Loring W. Tu. *Differential Forms in Algebraic Topology.* Graduate Texts in Mathematics. Springer New York, 1995.

[35] The Univalent Foundations Program. *Homotopy Type Theory: Univalent Foundations of Mathematics.* `https://homotopytypetheory.org/book`, Institute for Advanced Study, 2013.

[36] Bruce Blackadar. *Operator Algebras: Theory of C∗-Algebras and von Neumann Algebras.* Springer, 2013.

[37] Gerd Rudolph and Matthias Schmidt. *Differential Geometry and Mathematical Physics: Part II. Fibre Bundles, Topology and Gauge Fields.* Springer, 2017.

[38] Martin Schottenloher. *A Mathematical Introduction to Conformal Field Theory*, volume 759. 2008.

[39] Dusa McDuff and Deitmar Salamon. *Introduction to Symplectic Topology.* Oxford Graduate Texts in Mathematics. Oxford University Press, 2017.

[40] Mikhail. M. Kapranov and Vladimir A. Voevodsky. *2-categories and Zamolodchikov Tetrahedra Equations*, volume 56 of *Proc. Sympos. Pure Math.* Amer. Math. Soc., Providence, RI, 1994.

[41] Geoffrey Compère. *Advanced Lectures on General Relativity*, volume 952. Springer, 2019.

[42] Pavel Etingof, Shlomo Gelaki, Dmitri Nikshych, and Victor Ostrik. *Tensor Categories*, volume 205. American Mathematical Soc., 2016.

[43] David Mumford. *The Red Book of Varieties and Schemes: Includes the Michigan Lectures (1974) on Curves and Their Jacobians*, volume 1358. Springer Science & Business Media, 1999.

[44] Charles A. Weibel. *An Introduction to Homological Algebra.* Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1994.

[45] Peter J. Hilton and Urs Stammbach. *A Course in Homological Algebra.* Springer.

[46] Jean-Luc Brylinski. *Loop Spaces, Characteristic Classes and Geometric Quantization.* Birkhauser.

[47] Antoine Van Proeyen and Daniel Freedman. *Supergravity.* Cambridge University Press.

[48] William S. Massey. *A Basic Course in Algebraic Topology.* Springer.

[49] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to Quantum Field Theory.* Westview Press.

[50] Nadir Jeevanjee. *An Introduction to Tensors and Group Theory for Physicists.* Birkhauser.

[51] Herbet Goldstein, John L. Safko, and Charles P. Poole. *Classical Mechanics.* Pearson.

[52] Franco Cardin. *Elementary Symplectic Topology and Mechanics.* Springer.

[53] Walter Greiner and Joachim Reinhardt. *Field Quantization.* Springer.

[54] Walter Greiner. *Quantum Mechanics.* Springer.

[55] B. H. Bransden and Charles J. Joachain. *Quantum Mechanics.* Prentice Hall.

[56] Heydar Radjavi and Peter Rosenthal. *Invariant Subspaces.* Dover Publications.

[57] Max Karoubi. *K-Theory: An Introduction.* Springer.

[58] Damien Calaque and Thomas Strobl. *Mathematical Aspects of Quantum Field Theories.* Springer, 2015.

[59] Ivan Kolar, Peter W. Michor, and Jan Slovak. *Normal Operations in Differential Geometry.* Springer.

[60] Stephen B. Sontz. *Principal Bundles: The Classical Case.* Springer.

[61] Stephen B. Sontz. *Principal Bundles: The Quantum Case.* Springer.

[62] William Fulton and Joe Harris. *Representation Theory: A First Course.* Springer.

[63] Peter Petersen. *Riemannian Geometry.* Springer.

[64] Charles Nash and Siddharta Sen. *Topology and Geometry for Physicists.* Dover Publications.

[65] Ian M. Anderson. *The Variational Bicomplex.*

[66] Joel Robbin and Dietmar Salamon. The maslov index for paths. *Topology*, 32(4):827--844, 1993.

[67] Nima Moshayedi. 4-manifold topology, donaldson-witten theory, floer homology and higher gauge theory methods in the BV-BFV formalism. 2021.

[68] Edward Witten. Global anomalies in string theory. In *Symposium on Anomalies, Geometry, Topology*, 6 1985.

[69] F.A. Berezin and M.S. Marinov.

[70] Paul A. M. Dirac. Generalized Hamiltonian dynamics. *Canadian Journal of Mathematics*, 2:129–148, 1950.

[71] Angelo Vistoli. Notes on Grothendieck topologies, fibered categories and descent theory. *arXiv:math/0412512*, 2004.

[72] Emily Riehl and Dominic Verity. The theory and practice of Reedy categories. *Theory and Applications of Categories*, 29, 2013.

[73] Emily Riehl. Homotopical categories: From model categories to $(\infty, 1)$-categories. 2019. arXiv:1904.00886.

[74] Floris Takens. A global version of the inverse problem of the calculus of variations. *Journal of Differential Geometry*, 14(4):543--562, 1979.

[75] John Baez and Alexander Hoffnung. Convenient categories of smooth spaces. *Transactions of the American Mathematical Society*, 363(11):5789--5825, 2011.

[76] John C. Baez and Alissa S. Crans. Higher-dimensional algebra vi: Lie 2-algebras. 2003. arXiv:math/0307263.

[77] John C. Baez and Aaron D. Lauda. Higher-dimensional algebra v: 2-groups. 2003. arXiv:math/0307200.

[78] Edward Witten. Supersymmetry and Morse theory. *J. Diff. Geom*, 17(4):661--692, 1982.

[79] Urs Schreiber. *From Loop Space Mechanics to Nonabelian Strings.* PhD thesis, 2005.

[80] John C Baez and Urs Schreiber. Higher gauge theory. 2005. arXiv:math/0511710.

[81] Jade Master. Why is homology so powerful? 2020. arXiv:2001.00314.

[82] Marcus Berg, Cécile DeWitt-Morette, Shangjr Gwo, and Eric Kramer. The Pin groups in physics: C, P and T. *Reviews in Mathematical Physics*, 13(08):953--1034, 2001.

[83] Michael F. Atiyah. Topological quantum field theory. *Publications Mathématiques de l'IHÉS*, 68:175--186, 1988.

[84] Jens Eisert, Christoph Simon, and Martin B Plenio. On the quantification of entanglement in infinite-dimensional quantum systems. *Journal of Physics A: Mathematical and General*, 35(17):3911--3923, 2002.

[85] Benoît Tuybens. Entanglement entropy of gauge theories. 2017.

[86] John Baez, Alexander Hoffnung, and Christopher Rogers. Categorified symplectic geometry and the classical string. *Communications in Mathematical Physics*, 293:701--725, 2010.

[87] Charles Rezk. A model for the homotopy theory of homotopy theory. *Transactions of the American Mathematical Society*, 353(3):973--1007, 2001.

[88] Peter May. A note on the splitting principle. *Topology and Its Applications*, 153(4):605--609, 2005.

[89] Irina Markina. Group of diffeomorphisms of the unit circle as a principal $U(1)$-bundle.

[90] Alexander E. Hoffnung. Spans in 2-categories: A monoidal tricategory. 2011. arXiv:1112.0560.

[91] Eugenia Cheng and Nick Gurski. The periodic table of n-categories for low dimensions ii: Degenerate tricategories. 2007. arXiv:0706.2307.

[92] Mehmet B. Şahinoğlu, Dominic Williamson, Nick Bultinck, Michael Mariën, Jutho Haegeman, Norbert Schuch, and Frank Verstraete. Characterizing topological order with matrix product operators. 2014. arXiv:1409.2150.

[93] Dominic J. Williamson, Nick Bultinck, Michael Mariën, Mehmet B. Şahinoğlu, Jutho Haegeman, and Frank Verstraete. Matrix product operators for symmetry-protected topological phases: Gauging and edge theories. *Phys. Rev. B*, 94, 2016.

[94] Guifré Vidal. Efficient classical simulation of slightly entangled quantum computations. *Phys. Rev. Lett.*, 91, 2003.

[95] Aaron D. Lauda and Hendryk Pfeiffer. Open–closed strings: Two-dimensional extended TQFTs and Frobenius algebras. *Topology and its Applications*, 155(7):623--666, 2008.

[96] Domenico Fiorenza. An introduction to the Batalin-Vilkovisky formalism. 2004. arXiv:math/0402057v2.

[97] Stefan Cordes, Gregory Moore, and Sanjaye Ramgoolam. Lectures on 2d Yang-Mills theory, equivariant cohomology and topological field theories. arXiv:hep-th/9411210v2.

[98] Donald C. Ferguson. A theorem of Looman-Menchoff. `http://digitool.library.mcgill.ca/thesisfile111406.pdf`.

[99] Holger Lyre. Berry phase and quantum structure. arXiv:1408.6867.

[100] Florin Belgun. Gauge theory. `http://www.math.uni-hamburg.de/home/belgun/Gauge4.pdf`.

[101] Vladimir Itskov, Peter J. Olver, and Francis Valiquette. Lie completion of pseudogroups. *Transformation Groups*, 16:161--173, 2011.

[102] Richard Borcherds. Lie groups. `https://math.berkeley.edu/~reb/courses/261/`.

[103] Andrei Losev. From Berezin integral to Batalin-Vilkovisky formalism: A mathematical physicist's point of view. 2007.

[104] Edward Witten. Coadjoint orbits of the Virasoro group. *Comm. Math. Phys.*, 114(1):1--53, 1988.

[105] Sidney R. Coleman and Jeffrey E. Mandula. All possible symmetries of the S-matrix. *Phys. Rev.*, 159, 1967.

[106] Emily Riehl. Monoidal algebraic model structures. *Journal of Pure and Applied Algebra*, 217(6):1069--1104, 2013.

[107] Valter Moretti. Mathematical foundations of quantum mechanics: An advanced short course. *International Journal of Geometric Methods in Modern Physics*, 13, 2016.

[108] Antonio Michele Miti. Homotopy comomentum maps in multisymplectic geometry, 2021.

[109] John E. Roberts. Spontaneously broken gauge symmetries and superselection rules. 1974.

[110] Richard Sanders. Commutative spectral triples & the spectral reconstruction theorem.

[111] Jean Gallier. Clifford algebras, Clifford groups, and a generalization of the quaternions, 2008. arXiv:0805.0311.

[112] Bozhidar Z. Iliev. Normal frames for general connections on differentiable fibre bundles. arXiv:math/0405004.

[113] Piotr Stachura. Short and biased introduction to groupoids. arXiv:1311.3866.

[114] Frederic Schuller. Lectures on the geometric anatomy of theoretical physics. `https://www.youtube.com/channel/UC6SaWe7xeOp31Vo8cQG1oXw`.

[115] Nima Amini. Infinite-dimensional Lie algebras. `https://people.kth.se/~namini/PartIIIEssay.pdf`.

[116] Peter Selinger. Lecture notes on lambda calculus.

[117] Nigel Hitchin. Lectures on special Lagrangian submanifolds. `https://arxiv.org/abs/math/9907034v1`, 1999.

[118] Olivia Caramello. Lectures on topos theory at the university of Insubria. `https://www.oliviacaramello.com/Teaching/Teaching.htm`.

[119] Derek Sorensen. An introduction to characteristic classes. `http://derekhsorensen.com/docs/sorensen-characteristic-classes.pdf`, 2017.

[120] Arun Debray. Characteristic classes. `https://web.ma.utexas.edu/users/a.debray/lecture_notes/u17_characteristic_classes.pdf`.

[121] Pascal Lambrechts.

[122] Chris Tiee. Contravariance, covariance, densities, and all that: An informal discussion on tensor calculus. `https://ccom.ucsd.edu/~ctiee/notes/tensors.pdf`, 2006.

[123] Emily Riehl. Homotopy (limits and) colimits. `http://www.math.jhu.edu/~eriehl/hocolimits.pdf`.

[124] Andreas Gathmann. Algebraic geometry. `https://www.mathematik.uni-kl.de/~gathmann/class/alggeom-2019/alggeom-2019.pdf`.

[125] Will J. Merry. Algebraic topology. `https://www.merry.io/algebraic-topology`.

[126] Stacks project. `https://stacks.math.columbia.edu/`.

[127] Wikipedia. `https://www.wikipedia.org/`.

[128] JJ Nuiten. Cohomological quantization of local prequantum boundary field theory. Master's thesis, 2013.

Index

B

E

F

G

M

Q

R

S