Summary: Mathematics & Physics

Nicolas Dewolf

November 11, 2017

Contents

II Topology 33

III Calculus 53

V Differential Geometry 157

IX   Relativity Theory   276

40 Special relativity   277

41 General Relativity   279

X   Quantum Mechanics   280

42 Schrödinger equation   281

43 Mathematical formalism   283

44 Angular Momentum   286

45 Dirac equation   290

46 Perturbation theory   292

XII Solid State Physics 310

XIII Appendices 321

Chapter 1

Introduction

Definitions, properties and formulas marked by a dagger symbol † are explained and/or derived in one of the appendices. This has been done such that the 'summary' itself contains only core notions and theorems.

Definitions of words in the middle of a text will be indicated by **bold text**. Terminology that has been defined in the past but that receives a new meaning/nuance will be indicated by *italic text*. Notions that have not been defined in this 'summary' but that are relevant/crucial are also indicated by *italic text*.

Vectors in Euclidean space will be denoted by a bold font letter with an arrow above: $\vec{\boldsymbol{a}}$. Vectors in Minkowski space (4-vectors) will be written without the arrow: $\mathbf{a}$.

Part I

Set Theory

Chapter 2

Set theory

Definition 2.0.1 (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.

Corollary 2.0.2. $S \subset P(S)$

Definition 2.0.3 (Partition). A partition of X is a family of disjoint subsets $(A_i)_{i \in I} \subset X$ such that $\bigcup_{i \in I} A_i = X$.

Definition 2.0.4 (Refinement). Let P be a partition of X. A refinement P' of P is a finite collection of subsets such that every $A \in P$ can be written as a disjoint union of elements in P'. Hence P' is also a partition.

Definition 2.0.5 (Cover). A cover of S is a collection of sets $\mathcal{F} \subseteq 2^S$ such that

$$\bigcup_{V \in \mathcal{F}} V = S \tag{2.1}$$

2.1 Set operations

2.1.1 Difference

Definition 2.1.1 (Symmetric difference).

$$A \Delta B = (A \backslash B) \cup (B \backslash A) \tag{2.2}$$

2.1.2 Complement

Definition 2.1.2 (Complement). Let Ω be the universal set . Let $E \subseteq \Omega$. The complement of E is defined as:

$$E^c = \Omega \backslash E \tag{2.3}$$

Formula 2.1.3 (de Morgan's laws).

$$\left(\bigcup_i A_i \right)^c = \bigcap_i A_i^c \tag{2.4}$$

$$\left(\bigcap_i A_i \right)^c = \bigcup_i A_i^c \tag{2.5}$$

2.2 Ordered sets

2.2.1 Posets

Definition 2.2.1 (Preordered set). A preordered set is a set equipped with a reflexive and transitive binary relation.

Definition 2.2.2 (Partially ordered set). A set P equipped with a binary relation $\leq$ is called a partially ordered set (**poset**) if the following 3 axioms are fulfilled for all elements $a, b, c \in P$:

1. Reflexivity: $a \leq a$

2. Antisymmetry: $a \leq b \wedge b \leq a \implies a = b$

3. Transitivity: $a \leq b \wedge b \leq c \implies a \leq c$

It is a preordered set for which the binary relation is also anti-symmetric.

Definition 2.2.3 (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 (non-strict) totally ordered set. This property is called **'totality'**.

Definition 2.2.4 (Strict total order). A non-strict order $\leq$ has an associated strict order $<$ that satisfies $a < b \iff a \leq b \wedge a \neq b$.

Definition 2.2.5 (Maximal element). An element m of a poset P is maximal if for every $p \in P$, $m \leq p$ implies that $m = p$.

Definition 2.2.6 (Chain). A totally ordered subset of a poset is called a chain.

Theorem 2.2.7 (Zorn's lemma[1]). *Let $(P, \leq)$ be a poset. If every chain in P has an upper bound in P, then P has a maximal element.*

---

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

2.2.2   Lattices

Definition 2.2.8 (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 2.2.9. The join of $\{a, b\}$ is denoted by $a \wedge b$. The meet of $\{a, b\}$ is denoted by $a \vee b$.

Definition 2.2.10 (Lattice). A poset $(P, \leq)$ is called a lattice if it is both a join- and a meet-semilattice.

Definition 2.2.11 (Directed[2] set). A directed set is a set X equipped with a preorder $\leq$ and 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 2.2.12 (Net). A net on a topological space X is a subset of X indexed by a directed set I.

2.2.3   Bounded sets

Definition 2.2.13 (Supremum). The supremum $\sup(X)$ of a set X is the smallest upper bound of X.

Definition 2.2.14 (Infimum). The infimum $\inf(X)$ of a set X is the greatest lower bound of X.

Definition 2.2.15 (Maximum). If $\sup(X) \in X$ the supremum is called the maximum of X. This is denoted by $\max(X)$.

Definition 2.2.16 (Minimum). If $\inf(X) \in X$ the supremum is called the minimum of X. This is denoted by $\min(X)$.

2.2.4   Ordered field

2.2.5   Real numbers

Property 2.2.17 (First axiom). The set of real numbers is an ordered field $(\mathbb{R}, +, \cdot, <)$

Property 2.2.18 (Completeness axiom[3]). Every non-empty subset of $\mathbb{R}$ that is bounded above has a supremum.

---

[2]Sometimes called an *upward* directed set. Downward directed sets are analogously defined with a lower bound for every two elements. Directed sets are also sometimes called **filtered sets**.

[3]This form of the completeness axiom is also called the supremum property or the Dedekind completeness.

Property 2.2.19. $\mathbb{Q} \subset \mathbb{R}$

Remark. There is only one way to extend the field of rational numbers to the field of reals such that it satisfies the two previous axioms. This means that for every possible construction, their exists a bijection (isomorphism) between the two.

Definition 2.2.20 (Extended real line).

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\} = [-\infty, \infty] \tag{2.6}$$

2.2.6   Filter

Definition 2.2.21 (Filter). Let X be a partially ordered set. A family $\mathcal{F} \subseteq 2^X$ is a filter on X if it satisfies following conditions:

1. $\emptyset \notin \mathcal{F}$

2. $\forall A, B \in \mathcal{F} : A \cap B \in \mathcal{F}$

3. If $A \in \mathcal{F}$ and $A \subseteq B$ then $B \in \mathcal{F}$

2.3   Algebra of sets

Definition 2.3.1 (Algebra of sets). A collection $\mathcal{F}$ of subsets of 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 called a field of sets.

2.3.1 σ-algebra

Definition 2.3.2 (σ-algebra). A family of sets Σ is a σ-algebra over a set X if it satisfies the following 3 axioms:

1. $X \in \Sigma$

2. Closed under complements: $\forall E \in \Sigma : E^c \in \Sigma$

3. Closed under countable unions: $\{E_i\}_{i=1}^n \subset \Sigma : \bigcup_{i=1}^n E_i \in \Sigma$

Remark 2.3.3. Due to axioms (2) and (3) together with de Morgan's laws[4], a σ-algebra is also closed under countable intersections.

Corollary 2.3.4. Every algebra of sets is also a σ-algebra.

Property 2.3.5. The intersection of a family of σ-algebras is again a σ-algebra.

---

[4]See equations 2.4 and 2.5.

Definition 2.3.6. A σ-algebra $\mathcal{G}$ is said to be generated by a family of sets $\mathcal{A}$ if

$$\mathcal{G} = \bigcap\{\mathcal{F} : \mathcal{F} \text{ is a } \sigma\text{-algebra that contains } \mathcal{A}\} \tag{2.7}$$

It is thus the smallest σ-algebra containing $\mathcal{A}$.

Notation 2.3.7. The σ-algebra generated by a family of sets $\mathcal{A}$ is often denoted by $\mathcal{F}_{\mathcal{A}}$ or $\sigma(\mathcal{A})$.

Definition 2.3.8 (Borel set). Let $\mathcal{B}$ be the σ-algebra generated by all open[5] intervals $O \subseteq \mathbb{R}$. The elements $B \in \mathcal{B}$ are called Borel sets.

Definition 2.3.9 (Product σ-algebra). The smallest σ-algebra containing the products $A_1 \times A_2$ for all $A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2$ is called the product σ-algebra of $\mathcal{F}_1$ and $\mathcal{F}_2$.

Notation 2.3.10. The product σ-algebra of $\mathcal{F}_1$ and $\mathcal{F}_2$ is denoted by $\mathcal{F}_1 \times \mathcal{F}_2$.

Alternative Definition 2.3.11. The product σ-algebra $\mathcal{F}$ can also be equivalently defined in the following two ways:

1. $\mathcal{F}$ is generated by the family of sets

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

2. $\mathcal{F}$ is the smallest σ-algebra such that the following projections are measurable (see 8.1.32):

$$\mathrm{Pr}_1 : \Omega \to \Omega_1 : (\omega_1, \omega_2) \mapsto \omega_1$$
$$\mathrm{Pr}_2 : \Omega \to \Omega_2 : (\omega_1, \omega_2) \mapsto \omega_2$$

Remark. Previous definitions can easily be generalized to higher dimensions.

2.3.2   Monotone class

Definition 2.3.12 (Monotone class). Let $\mathcal{A}$ be a family of sets. $\mathcal{A}$ is called a monotone class if it has the following two properties:

- For every increasing sequence $A_1 \subset A_2 \subset \ldots$

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

- For every decreasing sequence $A_1 \supset A_2 \supset \ldots$

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

Theorem 2.3.13 (Monotone class theorem). *Let $\mathcal{A}$ be an algebra of sets 2.3.1. If $\mathcal{G}_{\mathcal{A}}$ is the smallest monotone class containing $\mathcal{A}$ then it coincides with the σ-algebra generated by $\mathcal{A}$.*

---

[5]Closed and half-open (both types) intervals generate the same σ-algebra.

2.4   Groups

2.4.1   Group

Definition 2.4.1 (Semigroup). Let G be a set equipped with a binary operation $\star$. $(G, \star)$ is a semigroup if it satisfies following axioms:

1. G is closed under $\star$

2. $\star$ is asssociative

Definition 2.4.2 (Monoid). Let M be a set equipped with a binary operation $\star$. $(M, \star)$ is a monoid if it satisfies following axioms:

1. M is closed under $\star$

2. $\star$ is associative

3. M contains an identity element with respect to $\star$

Definition 2.4.3 (Group). Let G be a set equipped with a binary operation $\star$. $(G, \star)$ is a group if it satisfies following axioms:

1. G is closed under $\star$

2. $\star$ is asssociative

3. G has an identity element with respect to $\star$

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

Definition 2.4.4 (Commutative group[6]). Let $(G, \star)$ be a group. If $\star$ is commutative, then G is called a commutative group.

Definition 2.4.5 (Coset). Let G be a group and H a subgroup of G. The left coset of H with respect to $g \in G$ is defined as the set

$$gH = \{gh : h \in H\} \tag{2.8}$$

The right coset is analogously defined as Hg. If for all $g \in G$ the left and right cosets coincide then the subgroup H is said to be a **normal subgroup**. The sets of left and right cosets are denoted by G/H and $G\backslash H$ respectively.

Definition 2.4.6 (Center). The center of a group is defined as follows:

$$Z(G) = \{z \in G : \forall g \in G, zg = gz\} \tag{2.9}$$

This set is a subgroup of G.

---

[6]Also called an Abelian group.

Definition 2.4.7 (Order of a group). The number of elements in the group. It is denoted by $|G|$ or $\text{ord}(G)$.

Definition 2.4.8 (Order of an element). The order of an element $a \in G$ is the smallest integer n such that

$$a^n = e \tag{2.10}$$

where e is the identity element of G.

Definition 2.4.9 (Torsion group). A torsion group is a group for which all element have finite order. The torsion $\text{Tor}(G)$ of a group G is the set of all element $a \in G$ that have finite order. For Abelian groups, $\text{Tor}(G)$ is a subgroup.

Definition 2.4.10 (Quotient group). Let G be a group and N a normal subgroup. The quotient group G/N is defined as the set of left cosets of N in G. This set can be turned into a group itself by equipping it with a product such that the product of aN and bN is $(aN)(bN)$. The fact that N is a normal subgroup can be used to rewrite this as $(aN)(bN) = (ab)N$.

2.4.2 Symmetric and alternating groups

Definition 2.4.11 (Symmetric group). The symmetric group S_n or Sym_n of the set $V = \{1, 2, ..., n\}$ is defined as the set of all permutations of V.

Definition 2.4.12 (Alternating group). The alternating group A_n is the subgroup of S_n containing all even permutations.

Definition 2.4.13 (Cycle). A k-cycle is a permutation of the form $(a_1, a_2, ..., a_k)$ 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.

Formula 2.4.14. Let τ be a k-cycle. The following equality holds:

$$\tau^k = \mathbb{1}_G \tag{2.11}$$

Example 2.4.15. Consider the set $\{1, 2, 3, 4, 5, 6\}$. The permutation $\sigma : x \mapsto x - 2$ can be written as the cycle decomposition $\sigma = (1, 3, 5)(2, 4, 6)$.

Definition 2.4.16 (Transposition). A permutation which exchanges two elements but lets the other ones unchanged.

2.4.3 Direct product

Definition 2.4.17 (Direct product). Let G, H be two groups. The direct product $G \otimes H$ is defined as the set-theoretic Cartesian product $G \times H$ equipped with a binary operation $\cdot$ such that:

$$(g_1, h_1) \cdot (g_2, h_2) = (g_1 g_2, h_1 h_2) \tag{2.12}$$

where the operations on the right hand side are the group operations in G and H. The structure $G \otimes H = (G \times H, \cdot)$ is a group itself.

Notation 2.4.18. When the groups are Abelian, the direct product is often denoted by $\oplus$.

Definition 2.4.19 (Inner semidirect product). Let G be a group, H a subgroup of G and N a normal subgroup of G. G is said to be the direct product of H and N, denoted by $H \rtimes N$, if it satifies the following equivalent statements:

- $G = NH$ where $N \cap H = \{e\}$.

- For every $g \in G$ there exist unique $n \in N, h \in H$ such that $g = nh$.

- For every $g \in G$ there exist unique $h \in H, n \in N$ such that $g = hn$.

- There exists a group homomorphism $\rho : G \to H$ which satisfies $\rho|_H = e$ and $\ker(\rho) = N$.

- The composition of the natural embedding $i : H \to G$ and the projection $\pi : G \to G/N$ is a isomorphism between H and G/N.

G is also said to **split** over N.

Definition 2.4.20 (Outer semidirect product). Let G, H be two groups and let $\varphi : H \to \mathrm{Aut}(G)$ be a group homomorphism. 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:

$$(g_1, h_1) \cdot (g_2, h_2) = (g_1 \varphi(h_1)(g_1), h_1 h_2) \tag{2.13}$$

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

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

Remark 2.4.21. The direct product of groups is a special case of the outer semidirect product where the group homomorphism is given by the trivial map $\varphi : h \mapsto e_G$.

2.4.4   Free groups

Definition 2.4.22 (Free Abelian group). An abelian group G with generators $\{g_i\}_{i \in I}$ is said to be freely generated if every element $g \in G$ can be uniquely written as a formal linear combination of the generators:

$$G = \left\{ \sum_i a_i g_i \,\middle|\, a_i \in \mathbb{Z} \right\} \tag{2.14}$$

The set of generators $\{g_i\}_{i \in I}$ is then called a **basis**[7] of G. The number of elements in the basis is called the **rank** of G.

Property 2.4.23. Consider a free group G. Let $H \subset G$ be a subgroup. Then H is also free.

---

[7]In analogy with the basis of a vector space.

Theorem 2.4.24. *Let G be a finitely generated Abelian group of rank n, i.e. its basis has n elements. This group can be constructed in two different ways:*

$$G = F/H \qquad (2.15)$$

where both F, H are freely and finitely generated Abelian groups. The second decomposition is:

$$G = A \oplus T \qquad where \qquad T = Z_{h_1} \oplus \cdots \oplus Z_{h_m} \qquad (2.16)$$

where A is a freely and finitely generated group of rank $n - m$ and all Z_{h_i} are cyclic groups of order h_i. The group T is called the torsion subgroup[8]. The rank $n - m$ and the numbers h_i are unique.

2.4.5 Group presentations

Definition 2.4.25 (Relations). Let G be a group. If the product of a number of elements $g \in G$ is equal to the identity e then this product is called a relation on G.

Definition 2.4.26 (Complete set of relations). Let H be a group generated by a subgroup G. Let R be a set of relations on G. If H is uniquely (up to an isomorphism) determined by G and R then the set of relations is said to be complete.

Definition 2.4.27 (Presentation). Let H be a group generated by a subgroup G and a complete set of relations R on G. The pair (G, R) is called a presentation of H.

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.

2.4.6 Group actions

Definition 2.4.28 (Group homomorphism). A group homomorphism $\Phi : G \to H$ is a map satisfying $\forall g, h \in G$

$$\Phi(gh) = \Phi(g)\Phi(h) \qquad (2.17)$$

Definition 2.4.29 (Kernel). The kernel of a group homomorphism $\Phi : G \to H$ is defined as the set

$$K = \{g \in G : \Phi(g) = \mathbb{1}_H\} \qquad (2.18)$$

Theorem 2.4.30 (First isomorphism theorem). *Let G, H be a groups and let $\varphi : G \to H$ be a group homomorphism. If φ is surjective than $G/\ker\varphi \cong H$.*

Definition 2.4.31 (Group action). Let G be a group. Let V be a set. A map $\rho : G \times V \to V$ is called an action of G on V if it satisfies the following conditions:

- Identity: $\rho(\mathbb{1}_G, v) = v$

---

[8]See also definition 2.4.9.

- Compatibility: $\rho(gh, v) = \rho(g, \rho(h, v))$

For all $g, h \in G$ and $v \in V$. The set V is called a (left) **G-space**.

Remark 2.4.32. A group action can alternatively be defined as a group homomorphism from G to $\mathrm{Sym}(V)$. It assigns a permutation of V to every element $g \in G$.

Notation 2.4.33. The action $\rho(g, v)$ is often denoted by $g \cdot v$ or even gv.

Definition 2.4.34 (Orbit). The orbit of an element $x \in X$ with respect to a group G is defined as the set:
$$x \cdot G = \{x \cdot g | g \in G\} \tag{2.19}$$

Definition 2.4.35 (Stabilizer). The stabilizer group or **isotropy group** of an element $x \in X$ with respect to a group G is defined as the set:
$$G_x = \{g \in G | g \cdot x = x\} \tag{2.20}$$

This is a subgroup of G.

Definition 2.4.36 (Free action). A group action is free if $g \cdot x = x$ implies $g = e$ for every $x \in X$.

Definition 2.4.37 (Faithful action[9]). A group action is faithful if it is injective.

Definition 2.4.38 (Transitive action). A group action is 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 we can say that there is only one orbit.

Definition 2.4.39 (Homogeneous space). If the group action of a group G on a G-space X is transitive, then X is said to be a homogeneous space.

Property 2.4.40. Let X be a set and let G be a group such that the action of G on X is transitive. Then their exists a bijection $X \cong G/G_x$ where G_x is the stabilizer of any element $x \in X$.[10]

Definition 2.4.41 (G-module). Let G be a group. Let M be a commutative group. M equipped with a left group action $\varphi : G \times M \to M$ is a (left) G-module if φ satisfies the following equation (distributivity):
$$g \cdot (a + b) = g \cdot a + g \cdot b \tag{2.21}$$

where $a, b \in M$ and $g \in G$.

Definition 2.4.42 (G-module homomorphism). A G-module homomorphism is a map $f : V \to W$ satisfying
$$g \cdot f(v) = f(g \cdot v) \tag{2.22}$$

where the $\cdot$ symbol represents the group action in W and V respectively. It is sometimes called a **G-map**, a **G-equivariant map** or an **intertwining map**.

---

[9]A faithful action is also called an **effective** action.
[10]This can be proven by generalizing A.1.1

2.5 Rings

Definition 2.5.1 (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 a commutative group.

2. $(R, \cdot)$ is a monoid.

3. Multiplication is distributive with respect to addition.

Definition 2.5.2 (Unit). An invertible element of ring $(R, +, \cdot)$. The set of units forms a group under multiplication.

2.5.1 Ideal

Definition 2.5.3 (Ideal). Let $(R, +, \cdot)$ be a ring with $(R, +)$ its additive group. A subset $I \subseteq R$ is called an ideal[11] of R if it satisfies the following conditions:

1. $(I, +)$ is a subgroup of $(R, +)$

2. $\forall n \in I, \forall r \in R : (n \cdot r), (r \cdot n) \in I$

Definition 2.5.4 (Unit ideal). Let $(R, +, \cdot)$ be a ring. R itself is called the unit ideal.

Definition 2.5.5 (Proper ideal). Let $(R, +, \cdot)$ be a ring. A subset $I \subset R$ is said to be a proper ideal if it is an ideal of R and if it is not equal to R.

Definition 2.5.6 (Prime ideal). Let $(R, +, \cdot)$ be a ring. A proper ideal I is a prime ideal if for any $a, b \in R$ the following relation holds:

$$ab \in I \implies a \in I \lor b \in I$$

Definition 2.5.7 (Maximal ideal). Let $(R, +, \cdot)$ be a ring. A proper ideal I is said to be maximal if there exists no other proper ideal T in R such that $I \subset T$.

Definition 2.5.8 (Minimal ideal). A proper ideal is said to be minimal if it contains no other nonzero ideal.

Definition 2.5.9 (Generating set). 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:

$$I = \left\{ \sum_{i=1}^{n} l_i x_i r_i \ \middle| \ \forall i \leq n : l_i, r_i \in R \text{ and } x_i \in X \right\} \tag{2.23}$$

Left and right ideals are generated in a similar fashion.

---

[11]More generally: two-sided ideal

2.5.2 Graded ring

Definition 2.5.10 (Graded ring). Let R be a ring that can be written as the direct sum of Abelian groups A_k:

$$R = \bigoplus_{k \in \mathbb{N}} A_k \tag{2.24}$$

If R has the property that for every $i, j \in \mathbb{N} : A_i \star A_j \subseteq A_{i+j}$, where $\star$ is the ring multiplication, then R is said to be a graded ring. The elements of the space A_k are said to be **homogeneous of degree** k.

Formula 2.5.11 (Graded commutativity). Let $m = \deg v$ and let $n = \deg w$. If

$$vw = (-1)^{mn} wv \tag{2.25}$$

for all elements v, w of the graded ring then it is said to be a graded-commutative ring.

2.6 Vector spaces & algebras

See definitions 14.2.1 and 14.2.11 for the definitions of a vector space and algebra respectively.

2.7 Algebraic structures

2.7.1 Direct systems

Definition 2.7.1 (Direct system). Let $(I, \leq)$ be a directed set[12]. Let $\{A_i\}_{i \in I}$ be a family of algebraic objects (groups, rings, ...) and let $\{f_{ij} : A_i \to A_j\}_{i,j \in I}$ be a set of homomorphisms with the following properties:

- For every $i \in I$: $f_{ii} = e_i$, where e_i is the identity in A_i.

- 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 2.7.2 (Direct limit). Consider a direct system (A_i, f_{ij}) over a (directed) set I. The direct limit A of these direct systems is defined as follows:

$$\varinjlim A_i = \left. \bigsqcup_{i \in I} A_i \middle/ \sim \right. \tag{2.26}$$

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 $\phi_i : A_i \to A$ are morphisms.

---

[12]See definition 2.2.11.

2.7.2 Exact sequences

Definition 2.7.3 (Exact sequence). Consider a sequence (finite or infinite) of algebraic structures and their corresponding homomorphisms:

$$A_0 \xrightarrow{\Phi_1} A_1 \xrightarrow{\Phi_2} \cdots \xrightarrow{\Phi_n} A_n \tag{2.27}$$

The sequence is exact if for every $k \in \mathbb{N} : \mathrm{im}(\Phi_k) = \ker(\Phi_{k+1})$. This implies that $\Phi_{k+1} \circ \Phi_k = 0$ for all $h \in \mathbb{N}$.

Definition 2.7.4 (Short exact sequence). A short exact sequence is an exact sequence of the form:

$$0 \to A_0 \xrightarrow{\Phi_1} A_1 \xrightarrow{\Phi_2} A_3 \to 0 \tag{2.28}$$

A long exact sequence is an infinite exact sequence.

Property 2.7.5. Looking at some small examples we can derive some important constraints for certain exact sequences and especially for short exact sequences. Consider the sequence

$$0 \to A \xrightarrow{\Phi} B$$

This sequence can only be exact if Φ is an injective homomorphism (**monomorphism**). This follows from the fact that the only element in the image of the map $0 \to A$ is 0 because the map is a homomorphism. The kernel of Φ is thus trivial which implies that Φ is injective.

Analogously, the sequence

$$A \xrightarrow{\Psi} B \to 0$$

is exact if Ψ is a surjective homomorphism (**epimorphism**). This follows from the fact that the kernel of the map $B \to 0$ and thus the image of Ψ is all of B which implies that Ψ is surjective.

It follows that the sequence

$$0 \to A \xrightarrow{\Sigma} B \to 0$$

is exact if Σ is a **bimorphism** (which is often an isomorphism).

2.8 Functions

2.8.1 Domain

Definition 2.8.1 (Domain). Let $f : X \to Y$ be a function. The set X, containing the arguments of f, is called the domain of f.

Notation 2.8.2. The domain of f is denoted by $\mathrm{dom}(f)$.

Definition 2.8.3 (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 non-zero.

Notation 2.8.4. The support of f is denoted by $\operatorname{supp}(f)$

Remark. The support of a function is a subset of its domain.

Notation 2.8.5. Let X, Y be two sets. The set of functions $\{f : X \to Y\}$ is often denoted by X^Y.

2.8.2 Codomain

Definition 2.8.6 (Codomain). Let $f : X \to Y$ be a function. The set Y, containing (at least) all the output values of f, is called the codomain of f.

Definition 2.8.7 (Image). Let $f : X \to Y$ be a function. The following subset of Y is called the image of f:
$$\{y \in Y \mid \exists x \in X : f(x) = y\}$$
It is denoted by $\operatorname{im}(f)$.

Definition 2.8.8 (Level set). Let $f : X \to \mathbb{R}$ be a real-valued function and let $c \in \mathbb{R}$. The following set is called the level set of f:

$$L_c(f) = \{x \in X : f(x) = c\} \tag{2.29}$$

For $X = \mathbb{R}^2$ the level set is called a **level curve** and for $X = \mathbb{R}^3$ it is called the **level surface**.

2.9   Integers

2.9.1   Partition

Definition 2.9.1 (Composition). Let $n \in \mathbb{N}$. A k-composition of n is a k-tuple $(t_1, ..., t_k)$ such that $\sum_{i=1}^{k} t_k = n$.

Definition 2.9.2 (Partition). Let $n \in \mathbb{N}$. A partition of n is an ordered composition of n.

Definition 2.9.3 (Young diagram[13]). 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.

Definition 2.9.4 (Conjugate partition). Let λ be a partition of n with Young diagram $\mathcal{D}$. The conjugate partition λ' is obtained by reflecting $\mathcal{D}$ across its main diagonal.

Example 2.9.5. Using the diagram 2.1 we obtain the conjugate partition $(4, 3, 3, 3, 1)$ represented by

---

[13]Sometimes called a *Ferrers* diagram.
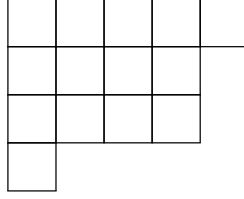
Figure 2.1: A Young diagram representing the partition $(5, 4, 4, 1)$ of 14.
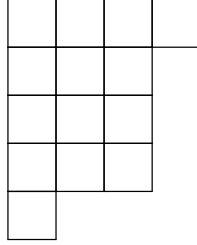
Figure 2.2: A Young diagram representing the partition $(4, 3, 3, 3, 1)$ of 14.

2.9.2 Superpartition

Definition 2.9.6 (Superpartition). Let $n \in \mathbb{N}$. A superpartition in the m-fermion sector is a sequence of integers of the following form:

$$\Lambda = (\Lambda_1, ..., \Lambda_m; \Lambda_{m+1}, ..., \Lambda_N) \tag{2.30}$$

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:

$$\Lambda = (\lambda^a; \lambda^s)$$

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

Notation 2.9.7. A superpartition of degree n in the m-fermion sector is said to be a superpartition of $(n|m)$. To every superpartition Λ we 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.

Part II

Topology

Chapter 3

Metric spaces

3.1 General definitions

Definition 3.1.1 (Metric). A metric (or distance) on a set M is a map $d : M \times M \to \mathbb{R}^+$ that has the following properties:

- Non-degeneracy: $d(x, y) = 0 \iff x = y$

- Symmetry: $d(x, y) = d(y, x)$

- Triangle inequality: $d(x, z) \leq d(x, y) + d(y, z) \quad , \forall x, y, z \in M$

Definition 3.1.2 (Metric space). A set M equipped with a distance function d is called a metric space and is denoted by (M, d).

Definition 3.1.3 (Diameter). The diamater of a subset $U \subset M$ is defined as

$$\text{diam}(U) = \sup_{x,y \in U} d(x, y) \tag{3.1}$$

Definition 3.1.4 (Bounded). A subset $U \subseteq M$ is bounded if $\text{diam}(U) < +\infty$.

Definition 3.1.5 (Open ball). An open ball centered on a point $x_0 \in M$ with radius $R > 0 \in \mathbb{R}$ is defined as the set:

$$\boxed{B(x_0, R) = \{x \in M : d(x, x_0) < R\}} \tag{3.2}$$

Definition 3.1.6 (Closed ball). The closed ball $\overline{B}(x_0, R)$ is defined as the union of the open ball $B(x_0, R)$ and its boundary, i.e. $\overline{B}(x_0, R) = \{x \in M : d(x, x_0) \leq R\}$.

Definition 3.1.7 (Interior point/neighbourhood). Let N be a subset of M. A point $x \in N$ is said to be an interior point of N if there exists an $R > 0$ such that $B(x, R) \subset M$. Furthermore, N is said to be a neighbourhood of x.

Definition 3.1.8 (Open). A subset $N \subset M$ is said to be open if every point $x \in N$ is an interior point of N.

Definition 3.1.9 (Closed). A subset $V \subset M$ is said to be closed if its complement is open.

Definition 3.1.10 (Limit point). Let S be a subset of X. A point $x \in X$ is called a limit point of S if every neighbourhood of x contains at least one point of S different from x.

Definition 3.1.11 (Accumulation point). Let $x \in X$ be a limit point of S. x is an accumulation point of S if every open neighbourhood of x contains infinitely many points of S.

Property 3.1.12. A metric space (M, d) has the following properties:

- M and $\emptyset$ are open.

- The union of open sets is open.

- The intersection of a **finite** number of open sets is open.

Definition 3.1.13 (Convergence). A sequence $(x_n)_{n \in \mathbb{N}} : \mathbb{N} \to M$ in a metric space (M, d) is said to be convergent to a point $a \in M$ if:

$$\forall \varepsilon > 0 : \exists N_0 \in \mathbb{N} : \forall n \geq N_0 : d(x_n, a) < \varepsilon \tag{3.3}$$

Definition 3.1.14 (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 $a \in \text{dom}(f)$ if:

$$\forall \varepsilon > 0 : \exists \delta_\varepsilon : \forall x \in \text{dom}(f) : d(a, x) < \delta_\varepsilon \implies d'(f(a), f(x)) < \varepsilon \tag{3.4}$$

Property 3.1.15. Let (M, d) be a metric space. The distance function $d : M \times M \to \mathbb{R}$ is a continuous function.

3.2 Constructions

Definition 3.2.1 (Product space). The cartesian product

$$M = M_1 \times M_2 \times ... \times M_n \text{ with } \forall n : (M_n, d_n)$$

is a metric space. If equipped with the distance function $d(x, y) = \max_{1 \leq i \leq n} d_i(x_i, y_i)$ it is also a metric space. This space is called the product space.

Definition 3.2.2 (Projection). The projection associated with the set M_j is defined as:

$$\text{pr}_j : M \to M_j : (a_1, ..., a_n) \mapsto a_j \tag{3.5}$$

Remark 3.2.3. A sequence in a product space M converges if and only if every component $(\text{pr}_j(x_m))_{m \in \mathbb{N}}$ converges in (M_j, d_j).

Example 3.2.4 (Supremum distance). Let $K \subset \mathbb{R}^n$ be a compact set. Denote the set of continuous functions $f : K \to \mathbb{C}$ by $\mathcal{C}(K, \mathbb{C})$. The following map defines a metric on $\mathcal{C}(K, \mathbb{C})$:

$$d_\infty(f, g) = \sup_{x \in K} |f(x) - g(x)| \tag{3.6}$$

Example 3.2.5 (p-metric). We can define following set of metrics on $\mathbb{R}^n$:

$$d_p(x, y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{1/p} \tag{3.7}$$

Example 3.2.6 (Chebyshev distance).

$$d_\infty(x, y) = \max_{1 \le i \le n} |x_i - y_i| \tag{3.8}$$

It is also called the **maximum metric**.

Remark 3.2.7. This metric is also an example of a product metric defined on the Euclidean product space $\mathbb{R}^n$. The notation d_∞, as used for the supremum distance, can be justified if the space $\mathbb{R}^n$ is identified with the set of maps $\{1, ..., n\} \to \mathbb{R}$ equipped with the supremum distance. Another justification is the following relation:

$$d_\infty(x, y) = \lim_{p \to \infty} d_p(x, y)$$

Chapter 4

Topology

4.1 Topological spaces

Definition 4.1.1 (Topology). Let Ω be a set. Let $\tau \subseteq 2^\Omega$. The set τ is a topology on Ω if it satisfies following axioms:

1. $\emptyset \in \tau$ and $\Omega \in \tau$

2. $\forall \, \mathcal{F} \subseteq \tau : \bigcup_{V \in \mathcal{F}} V \in \tau$

3. $\forall \, U, V \in \tau : U \cap V \in \tau$

Furthermore we call the elements of τ open sets and the couple (Ω, τ) a topological space.

Remark. On topological spaces the open sets are thus defined by axioms.

Definition 4.1.2 (Relative topology). Let (X, τ_X) be a topological space and Y a subset of X. We can turn Y into a topological space by equipping it with the following topology, called the relative topology:

$$\tau_{\text{rel}} = \{U_i \cap Y : U_i \in \tau_X\} \tag{4.1}$$

Definition 4.1.3 (Disjoint union). Let $\{X_i\}_{i \in I}$ be a family of topological spaces. Now consider the disjoint union

$$X = \bigsqcup_{i \in I} X_i \tag{4.2}$$

together with the canonical inclusion maps $\phi_i : X_i \to X : x_i \mapsto (x_i, i)$. We can turn X into a topological space by equipping it with the following topology:

$$\tau_X = \{U \subseteq X | \forall i \in I : \phi_i^{-1}(U) \text{ is open in } X_i\} \tag{4.3}$$

Definition 4.1.4 (Quotient space). Let X be a topological space and let $\sim$ be an equivalence relation defined on X. The set $X/_\sim$ can be turned into a topological space by equipping it with the following topology:

$$\tau_\sim = \{U \subseteq X/_\sim | \pi^{-1}(U) \text{ is open in } X\} \tag{4.4}$$

where π is the canonical surjective map from X to $X/_\sim$.

Example 4.1.5 (Discrete topology). The discrete topology is the topology such that every subset is open (and thus also closed).

Example 4.1.6 (Product topology). First consider the case where the index set I is finite. The product space $X = \prod_{i \in I} X_i$ can be turned into a topological space by equipping it with the topology generated by the following basis:

$$\mathcal{B} = \left\{ \prod_{i \in I} U_i \ \middle|\ U_i \in \tau_i \right\}$$

For general cases (countably infinite and uncountable index sets) the topology can be defined using the canonical projections $\pi_i : X \to X_i$. The general product topology (also called Tychonoff topology) is the coarsest (finest) topology such that all projections π_i are continuous.

Definition 4.1.7 (Topological group). A topological group is a group G equipped with a topology such that both the multiplication and inversion map are continuous.

Definition 4.1.8 (Pointed topological space). Let $x_0 \in X$. The triple (X, τ, x_0) is called a pointed topological space with base point x_0.

Definition 4.1.9 (Suspension). Let X be a topological space. The suspension of X is defined as the following quotient space:

$$SX = (X \times [0,1])/\{(x,0) \sim (y,0) \text{ and } (x,1) \sim (y,1) | x, y \in X\} \tag{4.5}$$

4.2 Neighbourhoods

Definition 4.2.1 (Neighbourhood). A set $V \subseteq \Omega$ is a neighbourhood of a point $a \in \Omega$ if there exists an open set $U \in \tau$ such that $a \in U \subseteq V$.

Definition 4.2.2 (Basis). Let $\mathcal{B} \subseteq \tau$ be a family of open sets. The family $\mathcal{B}$ is a basis for the topological space (Ω, τ) if every $U \in \tau$ can be written as:

$$U = \bigcup_{V \in \mathcal{F}} V \tag{4.6}$$

where $\mathcal{F} \subseteq \mathcal{B}$.

Definition 4.2.3 (Local basis). Let $\mathcal{B}_x$ be a family of open neighbourhoods of a point $x \in \Omega$. $\mathcal{B}_x$ is a local basis of x if every neighbourhood of x contains at least one element in $\mathcal{B}_x$.

Definition 4.2.4 (First-countability). A topological space (Ω, τ) is first-countable if for every point $x \in \Omega$ there exists a countable local basis.

Property 4.2.5 (Decreasing basis). Let $x \in \Omega$. If there exists a countable local basis for x then there also exists a countable decreasing local basis for x.

Definition 4.2.6 (Second-countability). A topological space (Ω, τ) is second-countable if there exists a countable global basis.

Property 4.2.7. Let X be a topological space. The closure of a subset V is given by:

$$\overline{V} = \{x \in X | \exists \text{ a net } (x_\lambda)_{\lambda \in I} \text{ in } X : x_\lambda \to x\} \tag{4.7}$$

This implies that the topology on X is completely determined by the convergence of nets[1].

Corollary 4.2.8. In first-countable spaces we only have to consider the convergence of sequences.

Definition 4.2.9 (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 U of a point $x \in X$ such that

$$f(u) = g(u) \qquad \qquad \forall u \in U$$

then this property defines an equivalence relation denoted by $f \sim_x g$ and the equivalence classes are called **germs**.

Property 4.2.10. Let the set Y in the previous definition be the set of reals $\mathbb{R}$. Then the germs at a point $p \in X$ satisfy following closure/linearity relations:

- $[f] + [g] = [f + g]$
- $\lambda[f] = [\lambda f]$
- $[f][g] = [fg]$

where $[f], [g]$ are two germs at p and $\lambda \in \mathbb{R}$ is a scalar.

4.2.1   Separation axioms

Definition 4.2.11 (Hausdorff space). A topological space is a Hausdorff space or T_2 space if it satisfies the following axiom:

$$(\forall x, y \in \Omega)(\exists \text{ neighbourhoods } U, V)(x \in U, y \in V, U \cap V = \emptyset) \tag{4.8}$$

This axiom is called the **separation axiom of Hausdorff**.

Property 4.2.12. Every singleton (and thus also every finite set) is closed in a Hausdorff space.

Definition 4.2.13 (Regular space). A topological space is said to be regular if for every closed subset F and every point $x \notin F$ there exist disjoint open subsets U, V such that $x \in U$ and $F \subset V$.

Notation 4.2.14. A space that is both regular and Hausdorff is a T_3 space.

Definition 4.2.15 (Normal space). A topological space is said to be normal if every two closed subsets have disjoint neighbourhoods.

Notation 4.2.16. A space that is both normal and Hausdorff is a T_4 space.

---

[1]See definition 2.2.12.

4.3 Convergence and continuity

Definition 4.3.1 (Convergence). A sequence $(x_n)_{n \in \mathbb{N}}$ in X is said to converge to a point $a \in X$ if:

$$(\forall \text{ neighbourhoods } V \text{ of } a)(\exists \ N > 0)(\forall \ n > N)(x_n \in V) \qquad (4.9)$$

Property 4.3.2. Every subsequence of a converging sequence converges to the same point[2].

Property 4.3.3. Let X be a Hausdorff space. The limit of a converging sequence in X is unique.

Definition 4.3.4 (Continuity). A function $f : X \to Y$ is continuous if the inverse image $f^{-1}(U)$ of every open set U is also open.

Theorem 4.3.5. *Let X be a first-countable space. Consider a function $f : X \to Y$. 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 4.3.6. If the space Y in the previous theorem is Hausdorff then the limit $f(a)$ does not need to be known because the limit is unique (see 4.3.3).

Remark 4.3.7. If the space X is not first-countable, we have to consider the convergence of nets 2.2.12.

Theorem 4.3.8 (Urysohn's lemma). *A topological space X is normal[3] 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.*

$$f(a) = 0, \forall a \in A \qquad \qquad f(b) = 1, \forall b \in B$$

4.3.1 Homeomorphisms

Definition 4.3.9 (Homeomorphism). A function f is called a homeomorphism if both f and f^{-1} are continuous and bijective.

Definition 4.3.10 (Embedding). A function is an embedding if it is homeomorphic onto its image.

---

[2]This limit does not have to be unique. See the next property for more information.
[3]See definition 4.2.15.

4.3.2 Metrizable spaces

Definition 4.3.11 (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 4.3.12 (Urysohn's metrization theorem). *Every second-countable T_3 space is metrizable.*

4.4 Connectedness

Definition 4.4.1 (Connected space). A topological space X is connected if it cannot be written as the disjoint union of non-empty open sets. Equivalently, X is connected if the only clopen sets are X and $\emptyset$.

Property 4.4.2. Let X be a connected space. 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 4.4.3 (Intermediate value theorem). *Let X be a connected space. Let $f : X \to \mathbb{R}$ be a continuous function. If $a, b \in f(X)$ then for every $c \in]a, b[$ we have that $c \in f(X)$.*

Definition 4.4.4 (Path-connected space[4]). Let X be a topological space. If for every two points $x, y \in X$ there exists a continuous function $\varphi : [0,1] \to X$ (i.e. a **path**) such that $\varphi(0) = x$ and $\varphi(1) = y$ then the space is said to be path-connected.

Property 4.4.5. Every path-connected space is connected.

The converse of previous property does not hold. There exists however the following (stronger) relation.

Property 4.4.6. A connected and locally path-connected space is path-connected.

Remark 4.4.7. The notions of connectedness and path-connectedness define equivalence relations on the space X. The equivalence classes are closed in X and form a cover of X.

4.5 Compactness

Definition 4.5.1 (Sequentially compact). A topological space is sequentially compact if every sequence[5] has a convergent subsequence.

---

[4]A similar notion is that of **arcwise-connectedness** where the function φ is required to be a homeomorphism.

Definition 4.5.2 (Finite intersection property). A family $\mathcal{F} \subseteq 2^X$ of subsets has the finite intersection property[6] if every finite subfamily has a non-zero intersection:

$$\bigcap_{i \in I} V_i \neq \emptyset \tag{4.10}$$

for all finite index sets I.

Definition 4.5.3 (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 cover of X.

Property 4.5.4. A first-countable space is sequentially compact if and only if every countable open cover has a finite subcover.

Definition 4.5.5 (Lindelöf space). A space for which every open cover has a countable subcover.

Property 4.5.6. Every second-countable space is also a Lindelöf space.

Definition 4.5.7 (Compact space). A topological space X is compact if every open cover of X has a finite subcover.

Theorem 4.5.8 (Heine-Borel[7]). *If a topological space X is sequentially compact and second-countable then every open cover has a finite subcover. This implies that X is compact.*

Theorem 4.5.9 (Heine-Borel on $\mathbb{R}^n$). *A subset of $\mathbb{R}^n$ is compact if and only if it is closed and bounded.*

Definition 4.5.10 (Locally compact). A topological space is locally compact if every point $x \in X$ has a compact neighbourhood.

Theorem 4.5.11 (Dini's theorem). *Let (X, τ) be a compact space. Let $(f_n)_{n \in \mathbb{N}}$ be an increasing sequence of continuous functions $f_n : X \to \mathbb{R}$. If $(f_n)_n \to f$ pointwise to a continuous function f then the convergence is uniform.*

Definition 4.5.12 (Paracompact space). A topological space is paracompact if every open cover has a locally finite open refinement.

Property 4.5.13 (ω-boundedness). Let X be a topological space. X is said to be ω-bounded if the closure of every countable subset is compact.

Definition 4.5.14 (Partition of unity). Let $\{\varphi_i : X \to [0,1]\}_i$ be a collection of continuous functions such that for every $x \in X$:

- For every neighbourhood V of x, the set $\{f_i : \mathrm{supp} f_i \cap U \neq \emptyset\}$ is finite.

- $\sum_i f_i = 0$

Definition 4.5.15. Consider an open cover $\{V_i\}_{i \in I}$ of X, indexed by a set I. If there exists a partition of unity, also indexed by I, such that $\mathrm{supp}(\varphi_i) \subseteq U_i$, then this partition of unity is said to be **subordinate** to the open cover.

---

[5]The sequence itself does not have to converge.

[6]The family is then called a FIP-family.

[7]Also Borel-Lebesgue.

4.5.1 Compactifications

Definition 4.5.16 (Dense). A subset $V \subseteq X$ is dense in a topological space X if $\overline{V} = X$.

Definition 4.5.17 (Separable space). A topological space is separable if it contains a countable dense subset.

Property 4.5.18. Every second-countable space is separable.

Definition 4.5.19 (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 4.5.20. 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 unique.

Definition 4.5.21 (One-point compactification). Let X be a Hausdorff space. A one-point compactification or **Alexandrov compactification** is a compactification X' such that $X' \setminus X$ is a singleton.

4.5.2 Compactness in metric spaces

Definition 4.5.22 (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 $\forall F \in \mathcal{F} : \text{diam}(F) \leq \varepsilon$.
- For every $\varepsilon > 0$ there exists a finite subset $E \subset M$ such that $M \subseteq \bigcup_{x \in E} B(x, \varepsilon)$.

Property 4.5.23. Every totally bounded set is 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 statement "*a set is compact if and only if it is closed and bounded*" known from Euclidean space $\mathbb{R}^n$.

Theorem 4.5.24. *For a metric space M the following statements are equivalent:*

- *M is compact.*
- *M is sequentially compact.*
- *M is complete and totally bounded.*

Theorem 4.5.25 (Heine-Cantor). *Let M, M' be two metric spaces with M being compact. Every continuous function $f : M \to M'$ is also uniformly continuous.*

Definition 4.5.26 (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 $a \in X$ if for all neighbourhoods U of a:

$$(\forall f \in \mathcal{F})(\forall x \in U)(d(f(x), f(a)) \leq \varepsilon) \tag{4.11}$$

for all $\varepsilon \geq 0$.

Property 4.5.27. Let $I \subseteq \mathbb{R}$ be an open interval. Let $\mathcal{F}$ be a collection of differentiable functions such that $\{f'(t) : f \in \mathcal{F}, t \in I\}$ is bounded. Then $\mathcal{F}$ is equicontinuous.

Theorem 4.5.28 (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[8].*

- *$\mathcal{F}$ is equicontinuous, closed under uniform convergence and $\{f(x) : f \in \mathcal{F}\}$ is totally bounded for every $x \in K$.*

4.6 Completeness

Definition 4.6.1 (Cauchy sequence). A sequence $(x_n)_{n \in \mathbb{N}}$ in a metric space (M, d) is Cauchy (or has the Cauchy property) if

$$(\forall \varepsilon > 0)(\exists N \in \mathbb{N})(\forall m, n \geq N)(d(x_m, x_n) < \varepsilon) \tag{4.12}$$

Property 4.6.2.

- Every closed subset of a complete metric space is complete.

- Every complete subset of a metric space is closed.

Property 4.6.3 (Cauchy criterion). A metric space (M, d) satisfies the Cauchy criterion if a sequence converges to a point $a \in M$ if and only if it is Cauchy.

Definition 4.6.4 (Completeness). A metric space is complete if it satisfies the Cauchy criterion.

4.7 Homotopy theory

Definition 4.7.1 (Homotopy). Let $f, g : X \to Y$ be continuous functions between 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 also induces an equivalence relation on $\mathcal{C}(X, Y)$.

---

[8]See formula 3.6.

Definition 4.7.2 (Homotopy type). Let X, Y be two topological spaces. X and Y are said to be homotopy equivalent, or of the same homotopy type, if there exist 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 4.7.3. Every homeomorphism is a homotopy equivalence.

Definition 4.7.4 (Null-homotopic). A continuous function is null-homotopic if it is homotopic to a constant function.

Definition 4.7.5 (Contractible space). A topological space X is said to be contractible if the identity map $\mathbb{1}_X$ is null-homotopic. Equivalently, the space is homotopy-equivalent to a point.

4.7.1   Fundamental group

In this subsection we will always assume to be working with pointed spaces. The base point will be denoted by x_0.

Definition 4.7.6 (Loop space). The space of all based loops in X. It is denoted by ΩX. We can define a multiplication operation on ΩX corresponding to the concatenation of loops[9].

Definition 4.7.7 (Fundamental group). The fundamental group $\pi_1(X, x_0)$ is defined as the loop space modulo homotopy. 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 we can speak of "the" fundamental group in the case of path-connected spaces.

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.

This definition can be generalized to arbitrary dimensions in the following way[10]:

Definition 4.7.8 (Homotopy group). The homotopy group $\pi_n(X, x_0)$ is defined as the set of homotopy classes of continuous maps $f : S^n \to X$ based at $x_0 \in X$. The set $\pi_0(X, x_0)$ is defined as the set of path-connected components of X.

Property 4.7.9. For $n \geq 1$ the sets $\pi_n(X, x_0)$ are groups.

Property 4.7.10. For $n \geq 2$ the homotopy groups $\pi_n(X, x_0)$ are abelian.

Property 4.7.11. If X is path-connected, then the homotopy groups $\pi_n(X, x_0)$ and $\pi_n(X, x_1)$ are isomorphic for all $x_0, x_1 \in X$ and all $n \in \mathbb{N}$.

---

[9]It should be noted that the speed at which the concatenated loops are traversed is doubled because the parameter t should remain an element of $[0, 1]$.

[10]Note however that we replace the interval $[0, 1]$ by the sphere S^1. This is nonrestrictive as we can construct S^n by mapping (identifying) the boundary of $[0, 1]^n$ to the basepoint x_0.

Property 4.7.12. Homeomorphic spaces have the same homotopy groups π_n.

Formula 4.7.13. 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:

$$\pi_n(X \times Y, (x_0, y_0)) = \pi_n(X, x_0) \otimes \pi_n(Y, y_0) \tag{4.13}$$

where $\otimes$ denotes the direct product of groups 2.4.17.

4.8 Homology and Cohomology

4.8.1 Simplexes

Definition 4.8.1 (Simplex). A k-simplex σ^k is defined as the following set:

$$\sigma^k = \left\{ \sum_{i=0}^{k} \lambda_i t_i \ \middle| \ \sum_{i=0}^{k} \lambda_i = 1 \text{ and } \lambda_i \geq 0 \right\} \tag{4.14}$$

where the points (vertices) t_i are linearly independent, i.e. the vectors $t_i - t_0$ are linearly independent. Equivalently it is the convex hull of the $k + 1$ vertices $\{t_0, ..., t_k\}$.

Definition 4.8.2 (Barycentric coordinates). The coordinates λ_i from previous defintion are called barycentric coordinates. This comes 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.

Definition 4.8.3 (Simplicial complex). A simplicial complex $\mathcal{K}$ is a set of simplexes satisfying following conditions:

- If σ is a simplex in $\mathcal{K}$ then so are its faces.

- If $\sigma_1, \sigma_2 \in \mathcal{K}$ then we have $\sigma_1 \cap \sigma_2 = \emptyset$ 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 4.8.4 (Polyhedron). Let $\mathcal{K}$ be a simplicial complex. The polyhedron associated with $\mathcal{K}$ is the topological spaces constructed by equipping $\mathcal{K}$ with the Euclidean subspace topology.

Definition 4.8.5 (Path-connectedness). Let $\mathcal{K}$ be a simplicial complex. $\mathcal{K}$ is said to be path-connected if every two vertices in $\mathcal{K}$ are connected by edges in $\mathcal{K}$.

Definition 4.8.6 (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 we say that X is **triangulable** and we call $\mathcal{K}$ a **triangulation** of X.

Theorem 4.8.7. *Let $\mathcal{K}$ be a path-connected polyhedron with basepoint a_0. Let $\mathcal{C} \subset \mathcal{K}$ be a contractible 1-dimensional subpolyhedron containing all vertexes of $\mathcal{K}$. Let G be the free group generated by the elements g_{ij} corresponding to the ordered 1-simplexes $[v_i, v_j] \in \mathcal{C}$. If the generators g_{ij} satisfy following two relations:*

- $g_{ij}g_{jk} = g_{ik}$ *for every ordered 2-simplex* $[v_i, v_j, v_k] \in \mathcal{K} \backslash \mathcal{C}$

- $g_{ij} = e$ *if* $[v_i, v_j] \in \mathcal{C}$.

then the group G is isomorphic to fundamental group $\pi_1(\mathcal{K}, a_0)$.

Corollary 4.8.8. From the theorem that homeomorphic spaces have the same homotopy groups it follows that the fundamental group of a triangulable space can be computed by looking at its triangulations.

4.8.2 Simplicial homology

Definition 4.8.9 (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 (additive) group generated by the k-simplexes in $\mathcal{K}$:

$$C_k(\mathcal{K}) = \left\{ \sum_i a_i \sigma_i \,\middle|\, \sigma_i \text{ is a } k\text{-simplex in } \mathcal{K} \text{ and } a_i \in \mathbb{Z} \right\} \tag{4.15}$$

For $k > n$ we define $C_k(\mathcal{K})$ to be $\{0\}$.

Definition 4.8.10 (Boundary operator). The boundary operator $\partial_k : C_k(\mathcal{K}) \to C_{k-1}(\mathcal{K})$ is the homomorphism defined by following properties:

- ∂_k is linear, i.e. $\partial_k(\sum_i a_i \sigma_i) = \sum_i a_i \partial_k(\sigma_i)$

- For every oriented k-simplex $[v_0, ..., v_k]$ we have that

$$\partial_k[v_0, ..., v_k] = \sum_{i=0}^{k} (-1)^i [v_0, ..., \hat{v}_i, ..., v_k] \tag{4.16}$$

- The boundary of every 0-chain is 0.

where $[v_0, ..., \hat{v}_i, ..., v_k]$ denotes the $(k-1)$-simplex obtained by removing the vertex v_i.

Property 4.8.11. The boundary operators satisfy following relation:

$$\partial_k \circ \partial_{k-1} = 0 \tag{4.17}$$

This property turns the system (C_k, ∂_k) into a so-called **chain complex**.

Definition 4.8.12 (Cycle group). The kth cycle group $Z_k(\mathcal{K})$ is defined as the set of k-chains σ_k such that $\partial_k \sigma_k = 0$. These chains are also called **cycles**.

Definition 4.8.13 (Boundary group). The kth boundary group $B_k(\mathcal{K})$ is defined as the set of k-chains σ_k for which there exists a $(k+1)$-chain N such that $\partial_{k+1}N = \sigma_k$. These chains are called **boundaries**.

Definition 4.8.14 (Homology group). From property 4.17 it follows that $B_k(\mathcal{K})$ is a subgroup of $Z_k(\mathcal{K})$. We can thus define the kth homology group $H_k(\mathcal{K})$ as the following quotient:

$$H_k(\mathcal{K}) = Z_k(\mathcal{K})/B_k(\mathcal{K}) \tag{4.18}$$

Theorem 2.4.24 tells us that we can write $H_k(\mathcal{K})$ as $G_k \oplus T_k$. Both of these groups tell us 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 tells us how the space $\mathcal{K}$ is twisted.

Property 4.8.15. If two topological spaces are of the same homotopy type then they have isomorphic homology groups. It follows that homeomorphic spaces have isomorphic homology groups.

Corollary 4.8.16. It follows from the definition of a triangulation that we can (easily) construct the homology groups for a given triangulable space by looking at one of its triangulations.

Definition 4.8.17 (Betti numbers). The numbers $R_k(\mathcal{X})$ from the definition of homology groups are called the Betti numbers of $\mathcal{X}$.

Formula 4.8.18 (Euler characteristic). The Euler characteristic of a triangulable space X is defined as follows[11]:

$$\boxed{\chi(X) = \sum_i (-1)^i R_i(X)} \tag{4.19}$$

Definition 4.8.19. The definition of homology groups can be generalized by letting the (formal) linear combinations used in the definition of the chain group (see 4.8.9) be of the following form:

$$c^k = \sum_i g_i \sigma_i^k \tag{4.20}$$

where $G = \{g_i\}$ is an Abelian group and σ_i^k are k-simplexes. The kth homology group of X with coefficients in G is denoted by $H_k(X; G)$. In case of G being 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 (or field) is given by the *Universal coefficient theorem*.

Formula 4.8.20 (Künneth formula). Let X, Y be two triangulable spaces. The homology groups of the Cartesian product $X \times Y$ with coefficients in a field F is given by:

$$H_k(X \times Y; F) = \bigoplus_{k=i+j} H_i(X; F) \otimes H_j(Y; F) \tag{4.21}$$

When F is replaced by the set of integers the torsion subgroups have to be taken into account. This will not be done here.

---

[11]This formula is sometimes called the *Euler-Poincaré* or *Poincaré* formula.

4.8.3   Relative homology

In this section we use a simplicial complex K and a subcomplex L.

Definition 4.8.21 (Relative chain group). The k-chain group of K modulo L is defined as the following quotient group:

$$C_k(K, L) = C_k(K)/C_k(L) \tag{4.22}$$

Definition 4.8.22 (Relative boundary operator). The relative boundary operator $\overline{\partial}_k$ is defined as follows:

$$\overline{\partial}_k(c_k + C_k(L)) = \partial_k c_k + C_{k-1}(L) \tag{4.23}$$

where $c_k \in C_k(K)$. This operator is, just like the ordinary boundary operator ∂_k, a homomorphism.

Definition 4.8.23 (Relative homology groups). The relative cycle and relative boundary groups are defined analogous to their ordinary counterparts. The relative homology groups are then defined as follows:

$$H_k(K, L) = \frac{\ker \overline{\partial}_k}{\operatorname{im} \overline{\partial}_{k+1}} \tag{4.24}$$

Elements $h_k \in H_k(K, L)$ can thus be written as $h_k = z_k + C_p(L)$ where z_k does not have to be a relative k-cycle. We merely require that $\partial_k z_k$ is a chain in $C_{k-1}(L)$.

Definition 4.8.24 (Homology sequence). Using the relative homology groups we obtain following (long) exact sequence:

$$\cdots \rightarrow H_k(L) \xrightarrow{i_*} H_k(K) \xrightarrow{j_*} H_k(K, L) \xrightarrow{\partial_k} H_{k-1}(L) \rightarrow \cdots \tag{4.25}$$

where i_* and j_* are the homology homomorphisms induced by the inclusions $i : L \rightarrow K$ and $j : K \rightarrow (K, L)$.

Theorem 4.8.25 (Excision theorem). *Let X, U, V be a triangulable spaces such that $U \subset V \subset X$. If the closure $\overline{U}$ is contained in the interior $V°$ then:*

$$H_k(X, V) = H_k(X \backslash U, V \backslash U) \tag{4.26}$$

4.8.4   Singular homology

Definition 4.8.26 (Singular simplex). Consider the **standard** k-simplex Δ^k:

$$\Delta^k = \{(x_0, ..., x_k) \in \mathbb{R}^{k+1} | \sum_i x_i = 1 \text{ and } x_i \geq 0\} \tag{4.27}$$

A singular k-simplex in a topological space X is defined as a continuous map $\sigma^k : \Delta^k \rightarrow X$. The name singular comes from the fact that the maps σ^k need not be invertible.

Definition 4.8.27 (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 $\sum_i g_i \sigma_i^k$. The basis of this freely generated group is in most cases infinite as there are multiple ways to map Δ^k to X.

Definition 4.8.28 (Singular boundary operator). The singular boundary operator ∂ (we use the same notation as for simplicial boundary operators) is defined by its linear action of the singular chain group $S_k(X)$. It follows that we only have to know the action on the singular simplexes σ^k.

We first introduce the notation $[v_0, ..., v_k] := [\sigma^k(e_0), ..., \sigma^k(e_k)]$ where e_i is the ith vertex of the standard simplex Δ^k. The action on the singular simplex σ^k is then given by:

$$\partial \sigma^k = \sum_i^k (-1)^k [v_0, ..., \hat{v}_i, ..., v_k] \tag{4.28}$$

The singular boundary operators satisfy the same relation as in the simplicial case:

$$\partial_k \circ \partial_{k-1} = 0 \tag{4.29}$$

Definition 4.8.29 (Singular homology group). The singular homology groups are defined as follows:

$$H_k(X; G) = \frac{\ker \partial_k}{\operatorname{im} \partial_{k+1}} \tag{4.30}$$

Theorem 4.8.30. *Let X be a triangulable space. The kth singular homology group of X is isomorphic to the kth simplicial homology group of X.*

Remark 4.8.31. When X is not triangulable the previous theorem is not valid. The singular approach to homology is thus a more general construction, but it is often more difficult to compute the homology groups (even in the case of triangulable spaces).

4.8.5 Examples

Example 4.8.32. Let X be a contractible space. We then find that:

$$H_k(X) = \begin{cases} \mathbb{Z} & k = 0 \\ \{0\} & k > 0 \end{cases} \tag{4.31}$$

Example 4.8.33. Let P be a connected polyhedron. We then find that:

$$H_0(P) = \mathbb{Z} \tag{4.32}$$

Example 4.8.34. The homology groups of the n-sphere S^n are given by:

$$H_k(S^n) = \begin{cases} \mathbb{Z} & k = 0 \text{ or } k = n \\ \{0\} & \text{otherwise} \end{cases} \tag{4.33}$$

4.8.6   Axiomatic approach

Definition 4.8.35 (Eilenberg-Steenrod axioms). All homology theories have the following properties in common. By treating these properties as axioms we can construct homology theories as a sequence of functors H_k. The axioms are as follows:

1. **Homotopy**: If f, g are homotopic maps then their induced homology maps are the same.

2. **Excision**[12]: 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. **Dimension**: If X is a singleton then $H_k(X) = \{0\}$ for all $k \geq 1$.

4. **Additivity**: If $X = \bigsqcup_i X_i$ then $H_k(X) \cong \bigoplus_i H_k(X_i)$

5. **Exactness**: Each pair (X, A), where $A \subset X$, induces a long exact sequence

$$\cdots \to H_k(A) \xrightarrow{i_*} H_k(X) \xrightarrow{j_*} H_k(X, A) \xrightarrow{\partial_k} H_{k-1}(A) \to \cdots \tag{4.34}$$

   where i_* and j_* are the homology homomorphisms induced by the inclusions $i : A \to X$ and $j : X \to (X, A)$.

Let X be a singleton. The group $H_0(X)$ is called the **coefficient group** and gives the coefficients used in the construction of the free Abelian chain groups C_k.

Remark 4.8.36. If the dimension axiom is removed from the set of axioms, then we obtain a so-called *extraordinary homology theory*.

4.9   Sheaf theory

4.9.1   Sheafs

Definition 4.9.1 (Sheaf). Let X be a topological space. A sheaf over X is a tuple (S, X, π), where S is a topological space and $\pi : S \to X$ a continuous surjection, such that the following two conditions are satisfied:

- For every point $s \in S$ there exists a neighbourhood U such that $\pi|_U$ is homeomorphism onto some open neighbourhood of $\pi(s) \in X$. This map induces the discrete topology on S

- For every $x \in X$, the set $\pi^{-1}(x)$ is an algebraic structure such that the corresponding algebraic operation is continuous.

Definition 4.9.2 (Stalk). The preimage $\pi^{-1}(x)$ is called the stalk over x and is often denoted by S_x.

---

[12]See also theorem 4.8.25.

Definition 4.9.3 (Homomorphism of sheaves). Let S, S' be two sheaves over the same space with projections π and π'. A homomorphism of sheaves is a map Φ satisfying the following conditions:

- $\Phi : S \to S'$ is continuous.

- $\pi = \pi' \circ \Phi$, i.e. Φ maps stalks in S to corresponding stalks in S'.

- For each $x \in X$, the restriction $\Phi|_x : S_x \to S'_x$ is a homomorphism of the algebraic structures corresponding to the stalks.

4.9.2   Presheafes

Definition 4.9.4 (Presheaf). Let X be a topological space. A presheaf over X consists of an algebraic structure S_U for every open set $U \subseteq X$ and a homomorphism $\Phi_V^U : S_U \to S_V$ for every two open sets $U, V \subseteq X$ with $V \subseteq U$ such that the following conditions are satisfied:

- If $U = \emptyset$ then $S_U = 0$, where 0 is the zero object in the category corresponding to the algebraic structure of S_U.

- $\Phi_U^U = \mathbb{1}_X$

- If $W \subseteq V \subseteq U$ then $\Phi_W^U = \Phi_W^V \circ \Phi_V^U$.

Definition 4.9.5 (Homomorphism of presheaves). Let S, S' be two presheaves. A homomorphism $S \to S'$ is a set of homomorphisms $\Psi_U : S_U \to S'_U$ that commute with the maps Φ_V^U.

Construction 4.9.6. For every presheaf over X we can construct a sheaf (S, X, π). For every $x \in X$ we set the stalk S_x to be the direct limit 2.26 of the direct system (S_U, Φ_V^U). The set S is then defined as the union of all sets S_x and π maps every element of $S_x \subset S$ to x.

The topology on S is defined by means of the following basis. For every $U \in X$ and every element $f \in S_U$ we construct a subset $f_U \subset S$ given by $\{f_x \in S_x : x \in U\}$ where f_x is called the **germ**[13] of f at x. The basis for our topology is then given by the set $\{f_U : U \subset X, f \in S_U\}$.

4.9.3   Sections

Definition 4.9.7 (Section). A section of a sheaf (S, X, π) over an open set U is a continuous map $s : U \to S$ such that $\pi \circ s = \mathbb{1}_U$. The set of all sections carry the same algebraic structure as S.

Remark. A **global** section is a section $s : X \to S$.

---

[13]This is a generalization of definition 4.2.9.

Part III

Calculus

Chapter 5

Calculus

5.1 Sequences

Definition 5.1.1 (Limit superior). Let $(x_i)_{i \in \mathbb{N}}$ be a sequence of real numbers. The limit superior is defined as follows:

$$\limsup_{i \to +\infty} x_i = \inf_{i \geq 1} \left\{ \sup_{k \geq i} x_k \right\} \tag{5.1}$$

Definition 5.1.2 (Limit inferior). Let $(x_i)_{i \in \mathbb{N}}$ be a sequence of real numbers. The limit superior is defined as follows:

$$\liminf_{i \to +\infty} x_i = \sup_{i \geq 1} \left\{ \inf_{k \geq i} x_k \right\} \tag{5.2}$$

Theorem 5.1.3. *A sequence $(x_i)_{i \in \mathbb{N}}$ converges pointwise if and only if $\limsup_{i \to +\infty} x_i = \liminf_{i \to +\infty} x_i$.*

5.2 Continuity

Definition 5.2.1 (Lipschitz continuity). A function $f : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous if there exists a constant $C > 0$ such that

$$|f(x) - f(x')| \leq C|x - x'| \tag{5.3}$$

for all $x, x' \in \mathbb{R}$.

Theorem 5.2.2 (Intermediate value theorem). *Let $I = [a, b] \subset \mathbb{R}$ be a compact interval. Let $f : I \to \mathbb{R}$ be a continuous function. For every $c \in [f(a), f(b)]$ there exists an $x_0 \in I$ such that $f(x_0) = c$.*

Theorem 5.2.3 (Darboux's theorem). *Let f be a differentiable function on a closed interval I. Then f' has the intermediate value property[1].*

Remark 5.2.4 (Darboux function). Functions that have the intermediate value property are called Darboux functions.

Corollary 5.2.5 (Bolzano's theorem). If $f(a) < 0$ and $f(b) > 0$ (or vice versa) then there exists at least one point x_0 where $f(x_0) = 0$.

Corollary 5.2.6. The image of a compact set is also a compact set.

Theorem 5.2.7 (Weierstrass' extreme value theorem). *Let $I = [a, b] \subset \mathbb{R}$ be a compact interval. Let f be a continuous function defined on I. Then f attains a minimum and maximum at least once on I.*

5.3 Convergence

Definition 5.3.1 (Pointwise convergence). Let $(f_n)_{n \in \mathbb{N}}$ be a sequence of functions. The sequence is said to converge pointwise to a limit function $f(x)$ if

$$\forall x \in \operatorname{dom}(f_n) : \lim_{n \to +\infty} f_n(x) = f(x) \tag{5.4}$$

Definition 5.3.2 (Uniform convergence). Let $(f_n)_{n \in \mathbb{N}}$ be a sequence of functions. The sequence is said to converge uniformly to a limit function $f(x)$ if

$$\sup_{x \in \operatorname{dom}(f_n)} \left\{ |\lim_{n \to +\infty} f_n(x) - f(x)| \right\} = 0 \tag{5.5}$$

5.4 Derivative

5.4.1 Single variable

Formula 5.4.1 (Derivative).

$$\boxed{f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}} \tag{5.6}$$

Theorem 5.4.2 (Mean value theorem). *Let f be continuous on the closed interval $[a, b]$ and differentiable on the open interval $]a, b[$. Then there exists a point $c \in]a, b[$ such that:*

$$f'(c) = \frac{f(b) - f(a)}{b - a} \tag{5.7}$$

---

[1]This means that the function satisfies the conclusion of the intermediate value theorem 5.2.2.

Definition 5.4.3 (Differentiablity class). Let I be a set. 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 $i = 1, \ldots, n$) then f is said to be of class $\mathbf{C}^n(I)$.

Definition 5.4.4 (Smooth function). A function f is said to be smooth if it is of class $\mathbf{C}^\infty$.

Definition 5.4.5 (Analytic function). A function f is said to be analytic if it is smooth and if its Taylor series expansion around any point x_0 converges to f in some neighbourhood of x_0. The class of analytic functions defined on I is denoted by $\mathbf{C}^\omega(I)$.

Theorem 5.4.6 (Schwarz's theorem). *Let* $f \in C^2(\mathbb{R}^n, \mathbb{R})$, *then:*

$$\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right) = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right) \tag{5.8}$$

for all indices $i, j \leq n$.

Method 5.4.7 (Derivative of $f(x)^{g(x)}$). Let us consider a function of the form $u(x) = f(x)^{g(x)}$. To find the derivative of this function we can use the derivative of the natural logarithm:

First we take a look at the natural logarithm of the function:

$$\ln[u(x)] = g(x)\ln[f(x)]$$

Then we look at the derivative of the natural logarithm:

$$\frac{d\ln[u(x)]}{dx} = \frac{1}{u(x)}\frac{du(x)}{dx} \implies \frac{du(x)}{dx} = u(x)\frac{d\ln[u(x)]}{dx}$$

But according to the first equation we also have:

$$\frac{d\ln[u(x)]}{dx} = \frac{d}{dx}g(x)\ln[f(x)] = \frac{dg(x)}{dx}\ln[f(x)] + \frac{g(x)}{f(x)}\frac{df(x)}{dx}$$

Combining these two equations gives:

$$\boxed{\frac{d}{dx}\left[f(x)^{g(x)}\right] = f(x)^{g(x)}\left[\frac{dg}{dx}(x)\ln[f(x)] + \frac{g(x)}{f(x)}\frac{df}{dx}(x)\right]} \tag{5.9}$$

Theorem 5.4.8 (Euler's homogeneous function theorem). *Let f be a homogeneous function, i.e.* $f(ax_1, ..., ax_n) = a^n f(x_1, ..., x_n)$. *Then f satisfies following equality:*

$$\sum_k x_k \frac{\partial f}{\partial x_k} = n f(x_1, ..., x_n) \tag{5.10}$$

5.5   Riemann integral

Definition 5.5.1 (Improper Riemann integral).

$$\boxed{\int_{-\infty}^{+\infty} f(x)dx = \lim_{\substack{a \to -\infty \\ b \to +\infty}} \int_a^b f(x)dx}$$ (5.11)

5.6   Fundamental theorems

Theorem 5.6.1 (First fundamental theorem of calculus). *Let $f(x)$ be a continuous function defined on the open interval I. Let $c \in I$. The following theorem establishes a link between integration and differentiation:*

$$\boxed{\exists F(x) = \int_c^x f(x')dx' : F'(x) = f(x)}$$ (5.12)

Furthermore this function $F(x)$ is uniformly continuous on I.

Remark 5.6.2. The function $F(x)$ in the previous theorem is called a **primitive function** of $f(x)$. Remark that $F(x)$ is just 'a' primitive function as adding a constant to $F(x)$ does not change anything because the derivative of a constant is zero.

Theorem 5.6.3 (Second fundamental theorem of calculus). *Let $f(x)$ be a function defined on the interval $[a, b]$. Furthermore, let $f(x) \in C^1[a, b]$. We then find the following important theorem:*

$$\boxed{\int_a^b f'(x)dx = f(b) - f(a)}$$ (5.13)

Theorem 5.6.4 (Differentiation under the integral sign[2]).

$$\boxed{\frac{d}{dx} \int_{a(x)}^{b(x)} f(x, y)dy = f(x, b(x)) \cdot b'(x) - f(x, a(x)) \cdot a'(x) + \int_{a(x)}^{b(x)} \frac{\partial f(x, y)}{\partial x} dy}$$ (5.14)

5.7   Taylor expansion

5.7.1   Examples

Formula 5.7.1 (Exponential function).

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$ (5.15)

---

[2]This is a more general version of the so called 'Leibnitz integral rule'.

5.8 Euler integrals

5.8.1 Euler integral of the first kind

Formula 5.8.1 (Beta function).

$$\boxed{B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt} \qquad (5.16)$$

5.8.2 Euler integral of the second kind

Formula 5.8.2 (Gamma function).

$$\boxed{\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}dt} \qquad (5.17)$$

Formula 5.8.3. $B(x, y) = \dfrac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$

Formula 5.8.4 (Recursion formula). Let $n!$ denote the factorial for integer numbers.

$$\Gamma(n) = (n-1)! \qquad (5.18)$$

Chapter 6

Series

6.1 Convergence tests

Theorem 6.1.1. *A series $\sum_{i=1}^{+\infty} a_i$ can only converge if $\lim_{i \to +\infty} a_i = 0$.*

Property 6.1.2 (Absolute/conditional convergence). If $S' = \sum_{i=1}^{+\infty} |a_i|$ converges then so does the series $S = \sum_{i=1}^{+\infty} a_i$ and S is said to be absolutely convergent. If S converges but S' does not, then S is said to be conditionally convergent.

6.1.1 Comparison test

Definition 6.1.3 (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 sum $S_{a,k} \geq S_{b,k}$.

Remark. S_b is said to minorize S_a.

Theorem 6.1.4 (Comparison test). *Let S_a, S_b be two series such that S_a majorizes S_b. We have the following cases:*

- *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.*

6.1.2 Integral test

Theorem 6.1.5 (MacLaurin-Cauchy integral test). *Let f be a continuous non-negative monotone decreasing function on the interval $[n, +\infty[$. If $\int_n^{+\infty} f(x)dx$ is convergent then so is $\sum_{k=n}^{+\infty} f(k)$. On the other hand, if the integral is divergent, so is the series.*

Remark 6.1.6. The function does not have to be non-negative and decreasing on the complete interval. As long as it does on the interval $[N, +\infty[$ for some $N \geq n$. 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 (the same argument applies for the integral).

Property 6.1.7. If the integral in the previous theorem converges, then the series has following lower and upper bounds:

$$\int_{n}^{+\infty} f(x)dx \leq \sum_{i=n}^{+\infty} a_i \leq f(n) + \int_{n}^{+\infty} f(x)dx \tag{6.1}$$

6.1.3   Ratio test

Theorem 6.1.8 (d'Alembert's ratio test).

$$R = \lim_{n \to +\infty} \left| \frac{a_{n+1}}{a_n} \right| \tag{6.2}$$

Following cases arise:

- $R < 1$: *the series converges absolutely*
- $R > 1$: *the series does not converge*
- $R = 1$: *the test is inconclusive*

6.1.4   Root test

Theorem 6.1.9 (Cauchy's root test).

$$R = \limsup_{n \to +\infty} \sqrt[n]{|a_n|} \tag{6.3}$$

We have the following cases:

- $R < 1$: *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 6.1.10 (Radius of convergences). The number $\frac{1}{R}$, where R is defined as in the Cauchy test, is called the radius of convergence.

6.1.5 Gauss's test

$$\left| \frac{u_n}{u_{n+1}} \right| = 1 + \frac{h}{n} + \frac{B(n)}{n^k} \tag{6.4}$$

where $B(n)$ is a bounded function when $n \to \infty$ en $k > 1$.

Theorem 6.1.11. *If $h \leq 1$, the series converges.*

6.2 Asymptotic expansions

Definition 6.2.1 (Asysmptotic expansion). Let $f(x)$ be a continuous function. A series expansion of order N is called an asymptotic expansion of $f(x)$ if it satisfies:

$$f(x) - \sum_{n=0}^{N} = O(x^{N+1}) \tag{6.5}$$

Method 6.2.2 (Borel transform). Define the function $F(x) = \sum_{n=0}^{+\infty} \frac{a_n}{n!} x^n$. If the integral

$$\int_0^{+\infty} e^{-t} F(xt) dt < +\infty \quad , \quad \forall x \in \mathbb{R} \tag{6.6}$$

then $F(x)$ is called the Borel transform of $f(x)$. Furthermore the integral will give a convergent expression for $f(x)$.

Proof.

$$
\begin{aligned}
\int_0^{+\infty} F(xt) e^{-t} dt &= \sum_{n=0}^{N} \int_0^{+\infty} \frac{a_n}{n!} x^n t^n e^{-t} dt \\
&= \sum_{n=0}^{N} \frac{a_n}{n!} x^n \int_0^{+\infty} t^n e^{-t} dt \\
&= \sum_{n=0}^{N} \frac{a_n}{n!} x^n \Gamma(n+1) \\
&= \sum_{n=0}^{N} a_n x^n
\end{aligned} \tag{6.7}
$$

where we used the definition of the Gamma function on line 3 and the relation between the factorial function and the Gamma function on line 4. $\qquad \square$

Theorem 6.2.3 (Watson's theorem). *The uniqueness of the function $F(x)$ is guaranteed if the function $f(x)$ is holomorphic on the domain $\{z \in \mathbb{C} : |\arg(z)| < \frac{\pi}{2} + \varepsilon\}$.*

Chapter 7

Complex calculus

7.1 Complex algebra

The set of complex numbers $\mathbb{C}$ forms a 2-dimensional vector space over the field of real numbers. Furthermore the operations of complex addition and complex multiplication also turn the complex numbers into a field.

Definition 7.1.1 (Complex conjugate). The complex conjugate $\overline{z} : a + bi \mapsto a - bi$ is an involution, i.e. $\overline{\overline{z}} = z$. It is sometimes denoted by z^* instead of $\overline{z}$.

Formula 7.1.2 (Real/imaginary part). A complex number z can also be written as $\mathrm{Re}(z) + i\,\mathrm{Im}(z)$ where

$$\mathrm{Re}(z) = \frac{z + \overline{z}}{2} \tag{7.1}$$

$$\mathrm{Im}(z) = \frac{z - \overline{z}}{2i} \tag{7.2}$$

Formula 7.1.3 (Roots). Let $z \in \mathbb{C}$. The nth roots[1] of z are given by:

$$z^{1/n} = \sqrt[n]{|z|}\, e^{i\frac{\arg(w) + 2\pi k}{n}} \tag{7.3}$$

where $k \in \{0, 1, ..., n\}$.

Formula 7.1.4 (Complex logarithm).

$$\mathrm{LN}(z) = \ln(r) + i(\theta + 2\pi k) \tag{7.4}$$

Definition 7.1.5 (Branch cuts). From the above two formulas it is clear that the complex roots and logarithms are multivalued functions. To get an unambiguous image it is necessary to fix a value of the parameter k. By doing this there will arise lines (or curves) in the complex plane where the function is discontinuous. These are called branch cuts.

---

[1]Also see theorem 14.1.3.

7.2 Holomorphic functions

Definition 7.2.1 (Holomorphic). A function f is holomorphic on an open set U if it is complex differentiable at every point $z_0 \in U$.

Theorem 7.2.2 (Cauchy-Riemann conditions). *A function f is holomorphic if it satisfies the following conditions:*

$$\boxed{\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad and \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}} \tag{7.5}$$

or equivalently:

$$\boxed{\frac{\partial f}{\partial \overline{z}} = 0} \tag{7.6}$$

Remark 7.2.3. The condition $f(z) \in C^1(\Omega)$ was made redundant by the Cauchy-Goursat theorem.

Property 7.2.4. Functions u, v satisfying the CR-conditions are harmonic functions, i.e. they satisfy Laplace's equation.

Property 7.2.5. Functions u, v satisfying the CR-conditions have orthogonal level curves 2.29.

7.3 Complex integrals

In this and further sections, all contours have been chosen to be evaluated counterclockwise (by convention). To obtain results concerning clockwise evaluation, most of the time adding a minus sign is sufficient.

Definition 7.3.1 (Contour). A contour is a curve $z(t)$ that can be parametrized by

$$\left.\begin{array}{l} x = x(t) \\ y = y(t) \end{array}\right\} \to z(t) = z = x + iy \tag{7.7}$$

Formula 7.3.2 (Complex contour integral). The complex contour integral of a function $f(z) = u(z) + iv(z)$ is defined as the following line integral:

$$\int_{z_1}^{z_2} f(z)dz = \int_{(x_1,y_1)}^{(x_2,y_2)} [u(x,y) + iv(x,y)](dx + idy) \tag{7.8}$$

Theorem 7.3.3 (Cauchy's Integral Theorem). *Let Ω be a simply-connected subset of $\mathbb{C}$ and let f be a holomorphic function on Ω. Then for every closed rectifiable contour C in Ω:*

$$\boxed{\oint_C f(z)dz = 0} \tag{7.9}$$

Corollary 7.3.4. The contour integral of a holomorphic function depends only on the limits of integration and not on the contour connecting them.

Formula 7.3.5 (Cauchy's Integral Formula). Let Ω be a connected subset of $\mathbb{C}$ and let f be a holomorphic function on Ω. Let C be a contour in Ω. For every point z_0 inside C we find:

$$\boxed{f(z_0) = \frac{1}{2\pi i} \oint_C \frac{f(z)}{z - z_0} dz} \tag{7.10}$$

Corollary 7.3.6 (Analytic function). Let Ω be a connected subset of $\mathbb{C}$ and C a closed contour in Ω. If f is holomorphic on Ω then f is analytic[2] on Ω and:

$$\boxed{f^{(n)}(z_0) = \frac{1}{2\pi i} \oint_C f(z) \frac{n!}{(z - z_0)^{n+1}} dz} \tag{7.11}$$

Furthermore, the derivatives are also holomorphic on Ω.

Theorem 7.3.7 (Morera's Theorem). *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.*

Definition 7.3.8 (Meromorphic). A function f is called meromorphic when it is analytic on the whole complex plane with exception of isolated singularities.

7.4   Laurent series

Definition 7.4.1 (Laurent series). If f is function, analytic on an annulus A, then f can be expanded as the following series:

$$f(z) = \sum_{n=-\infty}^{\infty} a_n(z - z_0)^n \qquad \text{with} \qquad a_n = \frac{1}{2\pi i} \oint \frac{f(z')}{(z' - z_0)^{n+1}} dz' \tag{7.12}$$

Remark 7.4.2. The Laurent series of an analytic function f converges uniformly to f in the ring shaped region ('*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 7.4.3 (Principal part). The principal part of a Laurent series is defined as the sum:

$$\sum_{n=-\infty}^{-1} a_n(z - z_0)^n$$

---

[2]See definition 5.4.5.

7.5 Analytic continuation

Theorem 7.5.1 (Schwarz' reflection principle). *Let $f(z)$ be analytic on the upper half plane. If $f(z)$ is real when z is real then*

$$f(\bar{z}) = \overline{f(z)} \tag{7.13}$$

7.6 Singularities

7.6.1 Poles

Definition 7.6.1 (Pole). A function $f(z)$ 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 7.6.2 (Essential singularity). A function $f(z)$ 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. its Laurent series has infinitely many negative degree terms.

Theorem 7.6.3 (Picard's great theorem). *Let $f(z)$ be an analytic function with an essential singularity at z_0. On every punctured neighbourhood of z_0, $f(z)$ takes on all possible complex values, with at most a single exception, infinitely many times.*

Method 7.6.4 (Frobenius transformation). To study the behaviour of a function $f(z)$ at $z \to \infty$, one should apply the Frobenius transformation $h = 1/z$ and study the limit $\lim_{h \to 0} f(h)$.

7.6.2 Branch cuts

Definition 7.6.5 (Branch point). Let $f(z)$ be a complex valued function. A point z_0 such that there exists no neighbourhood $|z - z_0| < \varepsilon$ where $f(z)$ is single valued is called a branch point.

Definition 7.6.6 (Branch cut). A line connecting excatly two branch points is called a branch cut. One of the branch points can be at infinity. In case of multiple branch cuts, they do not cross.

Example 7.6.7. Let $f(z) = \dfrac{1}{\sqrt{(z - z_1)...(z - z_n)}}$. This function has singularities at $z_1, ..., z_n$. Furthermore if n is even, this function will have n branch points and no branch points at infinity and the points will be grouped in pairs connected by a non-crossing branch cut. If n is odd, this function will have n finite branch points and one branch point at infinity. $n - 1$ branch points will be grouped in pairs connected by a non-crossing branch cut and the remaining branch point will be joined to infinity by a branch cut which does not cross the others.(See [5] for the proof.)

7.6.3   Residue theorem

Definition 7.6.8 (Residue). By applying formula 7.8 to a polynomial term we find, where C is a circular contour around a pole at the point z_0: $\int_C (z - z_0)^n dz = 2\pi i \delta_{n,-1}$. This means that integrating a Laurent series around a pole filters the coefficient a_{-1}. This coefficient is therefore called the residue of the function around the given pole. This is often denoted by $\text{Res}[f(z)]_{z=z_0}$.

Formula 7.6.9. For a pole of order m, the residue is calculated as follows:

$$\text{Res}\left[f(z)\right]_{z=z_j} = a_{-1} = \lim_{z \to z_0} \frac{1}{(m-1)!} \left(\frac{\partial}{\partial z}\right)^{m-1} (f(z)(z - z_0)) \tag{7.14}$$

For essential singularities the residue can be found by writing out the Laurent series explicitly.

Theorem 7.6.10 (Residue theorem). *If $f(z)$ is a meromorphic function in Ω and if C is a closed contour in Ω which contains the poles z_j of $f(z)$, then:*

$$\boxed{\oint_C f(z)dz = 2\pi i \sum_j \text{Res}\left[f(z)\right]_{z=z_j}} \tag{7.15}$$

Remark 7.6.11. For poles on the contour C, only half of the residue contributrs to the integral.

7.7   Limit theorems

Theorem 7.7.1 (Small limit theorem). *Let f be a function that is holomorphic almost every where on $\mathbb{C}$. 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 7.7.2 (Great limit theorem). *Let f be a function that is holomorphic almost every where on $\mathbb{C}$. 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 7.7.3 (Jordan's lemma). *Let g be a continuous function with $g(z) = f(z)e^{bz}$. Let the contour C be a semicircle lying in the half-plane bounded by the real axis and oriented away of the point $\bar{b}i$. 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 8

Measure theory and Lebesgue integration

8.1 Measure

8.1.1 General definitions

Definition 8.1.1 (Measure). Let X be a set. 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. Non-negativity: $\forall E \in \Sigma : \mu(E) \geq 0$

2. Null empty set: $\mu(\emptyset) = 0$

3. Countable-additivity[1] : $\forall i \neq j : E_i \cap E_j = \emptyset \implies \mu\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i)$

Definition 8.1.2 (Measure space). The pair (X, Σ) is called a measurable space. The elements $E \in \Sigma$ are called measurable sets. The triplet (X, Σ, μ) is called a measure space.

Definition 8.1.3 (Almost everywhere[2]). Let (X, Σ, μ) be a measure space. A property P is said to hold on X almost everywhere (a.e.) if it satisfies the following equation:

$$\mu\left(\{x \in X : \neg P(x)\}\right) = 0 \tag{8.1}$$

Definition 8.1.4 (Complete measure space). The measure space (X, Σ, μ) is said to be complete if for every $E \in \Sigma$ with $\mu(E) = 0$ the following property holds for all $A \subset E$:

$$A \in \Sigma \quad \text{and} \quad \mu(A) = 0$$

---

[1]also called σ-additivity

[2]In propability often called 'almost surely'.

Definition 8.1.5 (Completion). Let $\mathcal{F}, \mathcal{G}$ be σ-algebras over a set X. $\mathcal{G}$ is said to be the completion of $\mathcal{F}$ if it is the smallest σ-algebra such that the measure space $(X, \mathcal{G}, \mu)$ is complete.

Definition 8.1.6 (Regular Borel measure). Let μ be a non-negative countably additive set function defined on $\mathcal{B}$. μ is called a regular Borel measure if it satisifes following equations for every Borel set B:

$$
\begin{aligned}
\mu(B) &= \inf\{\mu(O) : O \text{ open}, O \supset B\} \\
\mu(B) &= \sup\{\mu(F) : F \text{ closed}, F \subset B\}
\end{aligned}
\tag{8.2}
$$

Definition 8.1.7 (σ-finite measure). Let $(\Omega, \mathcal{F}, P)$ be a measure space. The measure P is said to be σ-finite if there exists a sequence $(A_i)_{i \in \mathbb{N}}$ of measurable sets such that $\bigcup_{i=1}^{+\infty} A_i = \Omega$ with $\forall A_i : P(A_i) < +\infty$.

Method 8.1.8. 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 2.3.13.

8.1.2   Lebesgue measure

Formula 8.1.9 (Length of an interval). The length of an open interval $I = (a, b)$ is defined as:
$$
l(I) = b - a
\tag{8.3}
$$

Definition 8.1.10 (Null set). A set $A \subset \mathbb{R}$ is called a null set if it can be covered by a sequence of intervals of arbitrarily small length: $\forall \varepsilon > 0$ there exists a sequence $(I_n)_{n \in \mathbb{N}}$ such that
$$
A \subseteq \bigcup_{n=1}^{+\infty} I_n
\tag{8.4}
$$
with
$$
\sum_{i=1}^{+\infty} l(I_n) < \varepsilon
\tag{8.5}
$$

Theorem 8.1.11. *Let $(E_i)_{i \in \mathbb{N}}$ be a sequence of null sets. The union $\bigcup_{i=1}^{+\infty} E_i$ is also null.*

Corollary 8.1.12. Any countable set is null.

Definition 8.1.13 (Outer measure). Let $X \subseteq \mathbb{R}$ be an open set. The (Lebesgue) outer measure is defined as:

$$
\boxed{m^*(X) = \inf \left\{ \sum_{i=1}^{+\infty} l(I_i) \text{ with } (I_i)_{i \in \mathbb{N}} \text{ a sequence of open intervals that covers } X \right\}}
\tag{8.6}
$$

Property 8.1.14. Let I be an interval. The outer measure equals the length: $m^*(I) = l(I)$.

Property 8.1.15. The outer measure is translation invariant: $m^*(A + t) = m^*(A) \quad, \forall A, t$

Property 8.1.16. $m^*(A) = 0$ if and only if A is null.

Property 8.1.17. If $A \subset B$ then $m^*(A) \leq m^*(B)$.

Property 8.1.18 (Countable subadditivity). For every sequence of sets $(E_i)_{i \in \mathbb{N}}$ the following inequality holds:

$$m^* \left(\bigcup_{i=1}^{+\infty} E_i \right) \leq \sum_{i=1}^{+\infty} m^*(E_i) \tag{8.7}$$

Theorem 8.1.19 (Carathéodory's criterion / Lebesgue measure). *Let X be a set. If X satisfies the following equation, it is said to be Lebesgue measurable:*

$$\forall E \subseteq \mathbb{R} : m^*(E) = m^*(E \cap X) + m^*(E \cap X^c) \tag{8.8}$$

This is denoted by $X \in \mathcal{M}$ and the outer measure $m^*(X)$ is called the Lebesgue measure of X denoted by $m(X)$.

Property 8.1.20. All null sets and intervals are measurable.

Property 8.1.21 (Countable additivity). For every sequence $(E_i)_{i \in \mathbb{N}}$ with $E_i \in \mathcal{M}$ satisfying $i \neq j : E_i \cap E_j = \emptyset$ the following equation holds:

$$\boxed{m \left(\bigcup_{i=1}^{+\infty} E_i \right) = \sum_{i=1}^{+\infty} m(E_i)} \tag{8.9}$$

Remark. Previous property, together with the properties of the outer measure, implies that the Lebesgue measure is indeed a proper measure as defined in 8.1.1.

Property 8.1.22. $\mathcal{M}$ is a σ-algebra[3] over $\mathbb{R}$.

Theorem 8.1.23. *For every $A \subset \mathbb{R}$ there exists a sequence $(O_i)_{i \in \mathbb{N}}$ of open sets such that:*

$$A \subset \bigcap_i O_i \quad and \quad m \left(\bigcap_i O_i \right) = m^*(A) \tag{8.10}$$

Theorem 8.1.24. *For every $E \in \mathcal{M}$ there exists a sequence $(F_i)_{i \in \mathbb{N}}$ of closed sets such that:*

$$\bigcup_i F_i \subset E \quad and \quad m \left(\bigcup_i F_i \right) = m(E) \tag{8.11}$$

Remark. The previous 2 theorems imply that the Lebesgue measure is a regular Borel measure 8.2.

Theorem 8.1.25. *Let $E \subset \mathbb{R}$. $E \in \mathcal{M}$ if and only if for every $\varepsilon > 0$ there exist an open set $O \supset E$ and a closed set $F \subset E$ such that $m^*(O \backslash E) < \varepsilon$ and $m^*(E \backslash F) < \varepsilon$.*

---

[3]See definition 2.3.2.

Property 8.1.26. Let $(A_i)_{i \in \mathbb{N}}$ be a sequence of sets with $\forall i : A_i \in \mathcal{M}$. The following two properties apply:

$$\forall i : A_i \subseteq A_{i+1} \implies m \left(\bigcup_{i=1}^{+\infty} A_i \right) = \lim_{i \to +\infty} m(A_i) \tag{8.12}$$

$$\forall i : A_i \supseteq A_{i+1} \wedge m(A_1) < +\infty \implies m \left(\bigcap_{i=1}^{+\infty} A_i \right) = \lim_{i \to +\infty} m(A_i) \tag{8.13}$$

Remark 8.1.27. This property is not only valid for the Lebesgue measure but for every countably additive set function.

Property 8.1.28. The Lebesgue measure $m(X)$ is continuous at $\emptyset$, i.e. if $(A_i)_{i \in \mathbb{N}} \to \emptyset$ then $\lim\limits_{i \to +\infty} m(A_i) = 0$.

Theorem 8.1.29. *$\mathcal{M}$ is the completion of $\mathcal{B}$.*

Corollary 8.1.30. *$\mathcal{B} \subset \mathcal{M} \subset \mathcal{F}_{\mathbb{R}}$*

Definition 8.1.31 (Restricted Lebesgue measure). Let $B \subset \mathbb{R}$ be a measurable set with measure $m(B) > 0$. The restriction of the Lebesgue measure to the set B is defined as follows:

$$\mathcal{M}_B = \{A \cap B : A \in \mathcal{M}\} \qquad \text{and} \qquad \forall E \in \mathcal{M}_B : m_B(E) = m(E) \tag{8.14}$$

Furthermore, the measure space $(B, \mathcal{M}_B, m_B)$ is complete.

8.1.3   Measurable functions

Definition 8.1.32 (Measurable function). A function f is (Lebesgue) measurable if for every interval $I \subset \mathbb{R} : f^{-1}(I) \in \mathcal{M}$.

Definition 8.1.33 (Borel measurable function). A function f is called Borel measurable[4] if for every interval $I \subset \mathbb{R} : f^{-1}(I) \in \mathcal{B}$.

Remark 8.1.34. Inclusion 8.1.30 implies that every Borel function is also Lebesgue measurable.

Theorem 8.1.35. *The class of Lebesgue measurable[5] functions defined on $E \in \mathcal{M}$ is closed under multiplication and it forms a vector space.*

Property 8.1.36. Following types of functions are measurable:

- monotone functions

---

[4]These functions are often simply called 'Borel functions'.
[5]This property is also valid for Borel functions.

- continuous functions

- indicator functions

Corollary 8.1.37. Let f, g be measurable functions. 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 8.1.38. Let f be a measurable function. The set[6] $\{x : f(x) = a\}$ is also measurable for all $a \in \mathbb{R}$.

Theorem 8.1.39. *Define following functions, which are measurable if f is measurable as a result of previous properties:*

$$f^+(x) = \begin{cases} f(x) & \text{if } f(x) > 0 \\ 0 & \text{if } f(x) \leq 0 \end{cases} = \max(f, 0) \tag{8.15}$$

$$f^-(x) = \begin{cases} 0 & \text{if } f(x) > 0 \\ -f(x) & \text{if } f(x) \leq 0 \end{cases} = \max(-f, 0) \tag{8.16}$$

The function $f : E \to \mathbb{R}$ is measurable if and only if both f^+ and f^- are measurable. Furthermore f is measurable if $|f|$ is measurable, the converse is false.

8.1.4 Limit operations

Property 8.1.40. Let $(f_i)_{i \in \mathbb{N}}$ be a sequence of measurable[7] functions. The following operations are measurable:

- $\min_{i \leq k} f_i$ and $\max_{i \leq k} f_i$

- $\inf_{i \in \mathbb{N}} f_i$ and $\sup_{i \in \mathbb{N}} f_i$

- $\liminf_{i \to +\infty} f_i$ and $\limsup_{i \to +\infty} f_i$

Remark. The measurability of the limit inferior and limit superior follows from their definitions and from the measurability of the inf / sup and min / max.

Property 8.1.41. Let f be a measurable function. Let g be a function such that $f = g$ almost everywhere. The function g is measurable.

Corollary 8.1.42. A result of the previous two properties is the following: if a sequence of measurable functions converges pointwise a.e. then the limit is also a measurable function.

Definition 8.1.43 (Essential supremum).

$$\text{ess sup} f = \sup\{z : f \geq z \text{ a.e.}\} \tag{8.17}$$

---

[6]This set is called the 'level set' of f.

[7]This property is also valid for Borel functions.

Definition 8.1.44 (Essential infimum).

$$\text{ess inf} f = \inf\{z : f \leq z \text{ a.e.}\} \tag{8.18}$$

Property 8.1.45. Let f be a measurable function. $f \leq \text{ess sup} f$ a.e. and $f \geq \text{ess inf} f$ a.e. We also have that: $\text{ess sup} f \leq \sup f$ and $\text{ess inf} f \geq \inf f$, furthermore this last pair of inequalities becomes a pair of equalities if f is continuous.

Property 8.1.46. Let f, g be measurable functions. $\text{ess sup}(f + g) \leq \text{ess sup} f + \text{ess sup} g$. An analogous inequality holds for the essential infimum.

8.2 Lebesgue integral

8.2.1 Simple functions

Definition 8.2.1 (Indicator function). An important function when working with sets is the following one:

$$\mathbb{1}_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} \tag{8.19}$$

Definition 8.2.2 (Simple function). Let f be a function that takes on a finite number of non-negative values $\{a_i\}$ with for every $i \neq j : f^{-1}(a_i) \cap f^{-1}(a_j) = \emptyset$. f is called a simple function if it can be expanded in the following way:

$$f(x) = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}(x) \tag{8.20}$$

with $A_i = f^{-1}(a_i) \in \mathcal{M}$

Remark 8.2.3 (Step function). If the sets A_i are intervals, the simple function is often called a 'step function'.

Formula 8.2.4 (Lebesgue integral of simple functions). Let φ be a simple function as defined in equation 8.20. Let $\mu : \mathcal{M} \rightarrow \mathbb{R}$ be a Lebesgue measure and let E be a measurable set. The Lebesgue integral of φ over a E with respect to μ is given by:

$$\int_E \varphi d\mu = \sum_{i=1}^{n} a_i \mu(E \cap A_i) \tag{8.21}$$

Example 8.2.5. Let $\mathbb{1}_{\mathbb{Q}}$ be the indicator function of the set of rational numbers. This function is clearly a simple function. Previous formula makes it possible to integrate the rational indicator function over the real line, which is not possible in the sense of Riemann:

$$\int_{\mathbb{R}} \mathbb{1}_{\mathbb{Q}} dm = 1 \times m(\mathbb{Q}) + 0 \times m(\mathbb{R}\backslash\mathbb{Q}) = 0 \tag{8.22}$$

where the measure of the rational numbers is 0 because it is a countable set (see corollary 8.1.12.

8.2.2   Measurable functions

Formula 8.2.6 (Lebesgue integral). Let f be a non-negative measurable function. Let A be measurable set. The Lebesgue integral of f over E is defined as:

$$\int_E f \, dm = \sup \left\{ \int_E \varphi \, dm : \varphi \text{ a simple function such that } \varphi \leq f \right\} \tag{8.23}$$

Property 8.2.7. The Lebesgue integral $\int_E f \, dm$ of a measurable function f is always non-negative.

Notation 8.2.8. The following notation is frequently used (both in the sense of Riemann and Lebesgue):

$$\int f \, dm = \int_{\mathbb{R}} f \, dm \tag{8.24}$$

Formula 8.2.9. The following equality is easily proved as for every set $A \subseteq \mathbb{R} : A \cup A^c = \mathbb{R}$.

$$\int_A f \, dm = \int f \mathbb{1}_A \, dm \tag{8.25}$$

Theorem 8.2.10. *Let f be a non-negative measurable function. Then $f = 0$ a.e. if and only if $\int_{\mathbb{R}} f \, dm = 0$.*

Property 8.2.11. The Lebesgue integral over a null set is 0.

Property 8.2.12. Let f, g me measurable functions. The Lebesgue integral has the following properties:

- $f \leq g$ a.e. implies $\int f \, dm \leq \int g \, dm$.

- Let A be a measurable set. Let $B \subset A$. Then $\int_B f \, dm \leq \int_A f \, dm$.

- The Lebesgue integral is linear.

- For every two disjoint measurable sets A and B we have that $\int_{A \cup B} f \, dm = \int_A f \, dm + \int_B f \, dm$.

- **Mean value theorem**: If $a \leq f(x) \leq b$, then $am(A) \leq \int_A f \, dm \leq bm(A)$.

Theorem 8.2.13. *Let f be a non-negative measurable function. There exists an increasing sequence $(\varphi_i)_{i \in \mathbb{N}}$ of simple functions such that $\varphi_i \nearrow f$.*

Theorem 8.2.14. *Let f be a bounded measurable function defined on the interval $[a, b]$. For every $\varepsilon > 0$ there exists a step function[8] h such that $\int_a^b |f - h| \, dm < \varepsilon$.*

---

[8]See remark 8.2.3.

8.2.3   Integrable functions

Definition 8.2.15 (Integrable function). Let $E \in \mathcal{M}$. A measurable function f is said to be integrable over E if both $\int_E f^+ dm$ and $\int_E f^- dm$ are finite. The Lebesgue integral of f over E is defined as:

$$\int_E f dm = \int_E f^+ dm - \int_E f^- dm \tag{8.26}$$

Remark. The difference between the integral 8.23 and the integral of an integrable function is that with the latter f does not have to be non-negative.

Theorem 8.2.16. *f is integrable if and only if $|f|$ is integrable. Furthermore, $\int_E |f| dm = \int_E f^+ dm + \int_E f^- dm$.*

Property 8.2.17. Let f, g be integrable functions. The following important properties apply:

- $f + g$ is also integrable.

- $\forall E \in \mathcal{M}, \int_E f dm \leq \int_E g dm \implies f \leq g$ a.e.

- Let $c \in \mathbb{R}$. $\int_E (cf) dm = c \int_E f dm$.

- f is finite a.e.

- $|\int f dm| \leq \int |f| dm$

- $f \geq 0 \wedge \int f dm = 0 \implies f = 0$ a.e.

Theorem 8.2.18. *The set of functions integrable over a set $E \in \mathcal{M}$ forms a vector space. It is denoted by $\mathcal{L}^1(E)$.*

Property 8.2.19. Let $f \in \mathcal{L}^1$ and $\varepsilon > 0$. There exists a continuous function g, vanishing outside some finite interval, such that $\int |f - g| dm < \varepsilon$.

Property 8.2.20. Let $f \geq 0$. The mapping $E \mapsto \int_E f dm$ is a measure on E (if it exists, hence if f is integrable). Furthermore, this measure is said to be **absolutely continuous**.

Remark. See section 8.6 for further information.

8.2.4   Convergence theorems

Theorem 8.2.21 (Fatou's lemma). *Let $(f_n)_{n \in \mathbb{N}}$ be a sequence of non-negative measurable functions.*

$$\int_E \left(\liminf_{n \to \infty} f_n \right) dm \leq \liminf_{n \to \infty} \int_E f_n dm \tag{8.27}$$

Theorem 8.2.22 (Monotone convergence theorem). *Let $E \in \mathcal{M}$. Let $(f_n)_{n \in \mathbb{N}}$ be an increasing sequence of non-negative measurable functions such that $f_n \nearrow f$ pointwise a.e. We have the following powerful equality:*

$$\boxed{\int_E f dm = \lim_{n \to \infty} \int_E f_n(x) dm} \tag{8.28}$$

Method 8.2.23. To prove 'linear' results concerning integrable functions in spaces such as $\mathcal{L}^1(E)$ we proceed according to the following steps:

1. Verify that the property holds for indicator functions. (This often follows by definition.)

2. Use the linearity to extend the property to simple functions.

3. Apply the monotone convergence theorem to show that the property holds for all non-negative measurable functions.

4. Extend the property to all integrable functions by writing $f = f^+ - f^-$ and applying the linearity again.

Theorem 8.2.24 (Dominated convergence theorem). *Let $E \in \mathcal{M}$. Let $(f_n)_{n \in \mathbb{N}}$ be a sequence of measurable functions with $\forall n : |f_n| \leq g$ a.e. for a function $g \in \mathcal{L}^1(E)$. If $f_n \to f$ pointwise a.e. then f is integrable over E and*

$$\int_E f \, dm = \lim_{n \to \infty} \int_E f_n(x) \, dm \tag{8.29}$$

Property 8.2.25. Let $(f_n)_{n \in \mathbb{N}}$ be a sequence of non-negative measurable functions. The following equality applies:

$$\int \sum_{n=1}^{+\infty} f_n(x) \, dm = \sum_{n=1}^{+\infty} \int f_n(x) \, dm \tag{8.30}$$

We cannot conclude that the right-hand side is finite a.e., so the series on the left-hand side need not be integrable.

Theorem 8.2.26 (Beppo-Levi). *Suppose that*

$$\sum_{i=1}^{\infty} \int |f_n|(x) \, dm \text{ is finite.}$$

The series $\sum_{i=1}^{\infty} f_n(x)$ converges a.e. Furthermore, the series is integrable and

$$\int \sum_{i=1}^{\infty} f_n(x) \, dm = \sum_{i=1}^{\infty} \int f_n(x) \, dm \tag{8.31}$$

Theorem 8.2.27 (Riemann-Lebesgue lemma). *Let $f \in \mathcal{L}^1$. The sequences*

$$s_k = \int_{-\infty}^{+\infty} f(x) \sin(kx) \, dx$$

and

$$c_k = \int_{-\infty}^{+\infty} f(x) \cos(kx) \, dx$$

both converge to 0.

Remark. This theorem is useful in Fourier analysis.

8.2.5   Relation to the Riemann integral

Theorem 8.2.28 (Fundamental theorem of calculus). *If $f : [a, b] \to \mathbb{R}$ is continuous then f is integrable and the function $F : x \mapsto \int_a^x f \, dm$ is differentiable for $x \in]a, b[$ such that $F' = f$.*

Theorem 8.2.29. *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 $[a, b]$.*

- *Riemann-integrable functions on $[a, b]$ are integrable with respect to the Lebesgue measure on $[a, b]$ and the integrals coincide.*

Theorem 8.2.30. *If $f \geq 0$ and the improper Riemann integral 5.11 exists, then the Lebesgue integral $\int f \, dm$ exists and the two integrals coincide.*

8.3   Examples

Definition 8.3.1 (Dirac measure[9]). We define the Dirac measure as follows:

$$\delta_a(X) = \begin{cases} 1 & \text{if } a \in X \\ 0 & \text{if } a \notin X \end{cases} \tag{8.32}$$

The integration with respect to the Dirac measure has the following nice property[10]:

$$\int g(x) d\delta_a = g(a) \tag{8.33}$$

Example 8.3.2. Let $\mu = \delta_2$, $X = (-4; 1)$ and $Y = (-2; 17)$. The following two integrals are easily computed:

$$\int_X d\mu = 0$$
$$\int_Y d\mu = 1$$

8.4   Space of integrable functions

8.4.1   Distance

To define a distance between functions, we first have to define some notion of length of a function. Normally this would not be a problem, because we now do know how to integrate

---

[9]Compare to 10.6.

[10]This equality can be proved by applying formula 28.14 with $X \equiv a$.

integrable functions, however the fact that two functions differing on a null set have the same integral carries problems with it, i.e. a non-zero function could have a zero length. Therefore we will define the 'length' on a different vector space:

Define the following set of equivalence classes $L^1(E) = \mathcal{L}^1(E)_{/\equiv}$ by introducing the equivalence relation: $f \equiv g$ if and only if $f = g$ a.e.

Property 8.4.1. $L^1(E)$ is a Banach space[11].

Formula 8.4.2. A norm on $L^1(E)$ is given by:

$$||f||_1 = \int_E |f| dm \tag{8.34}$$

8.4.2 Hilbert space L^2

Property 8.4.3. L^2 is a Hilbert space[12].

Formula 8.4.4. A norm on $L^2(E)$ is given by:

$$||f||_2 = \left(\int_E |f|^2 dm \right)^{\frac{1}{2}} \tag{8.35}$$

This norm is induced by the following inner product:

$$\boxed{\langle f|g \rangle = \int_E f\bar{g} dm} \tag{8.36}$$

Now instead of deriving L^2 from $\mathcal{L}^2$ we do the opposite. We define $\mathcal{L}^2$ as the set of measurable functions for which equation 8.35 is finite.

Definition 8.4.5 (Orthogonality). As L^2 is a Hilbert space and thus has an inner product $\langle \cdot | \cdot \rangle$, it is possible to introduce the concept of orthogonality of functions in the following way:

$$\langle f|g \rangle = 0 \implies \text{f and g are orthogonal} \tag{8.37}$$

Furthermore it is also possible to introduce the angle between functions in the same way as equation 14.38.

Formula 8.4.6 (Cauchy-Schwarz inequality). Let $f, g \in L^2(E, \mathbb{C})$. We have that $fg \in L^1(E\mathbb{C})$ and:

$$\boxed{\left| \int_E f\bar{g} dm \right| \leq ||fg||_1 \leq ||f||_2 ||g||_2} \tag{8.38}$$

Remark. This follows immediately from formula 8.40.

Property 8.4.7. If E has finite Lebesgue measure then $L^2(E) \subset L^1(E)$.

---

[11]See definition 16.1.4.
[12]See definition 16.2.1.

8.4.3 L^p spaces

Generalizing the previous two Lebesgue function classes leads us to the notion of L^p spaces with the following norm:

Property 8.4.8. For all $1 \leq p \leq +\infty$ $L^p(E)$ is a Banach space with a norm given by:

$$||f||_p = \left(\int_E |f|^p \, dm \right)^{\frac{1}{p}} \tag{8.39}$$

Remark 8.4.9. Note that L^2 is the only L^p space that is also a Hilbert space. The other L^p spaces do not have a norm induced by an inner product.

Formula 8.4.10 (Hölder's inequality). Let $\frac{1}{p} + \frac{1}{q} = 1$ with $p \geq 1$. For every $f \in L^p(E)$ and $g \in L^q(E)$ we have that $fg \in L^1(E)$ and:

$$||fg||_1 \leq ||f||_p ||g||_q \tag{8.40}$$

Formula 8.4.11 (Minkowski's inequality). For every $p \geq 1$ and $f, g \in L^p(E)$ we have

$$||f + g||_p \leq ||f||_p + ||g||_p \tag{8.41}$$

Property 8.4.12. If E has finite Lebesgue measure then $L^q(E) \subset L^p(E)$ when $1 \leq p \leq q < +\infty$.

8.4.4 L^∞ space of essentially bounded measurable functions

Definition 8.4.13 (Essentially bounded function). Let f be a measurable function satisfying ess sup$|f| < +\infty$. The function f is said to be essentially bounded and the set of all such functions is denoted by $L^\infty(E)$.

Formula 8.4.14. A norm on L^∞ is given by:

$$||f||_\infty = \text{ess sup}|f| \tag{8.42}$$

This norm is called the **supremum norm** and it induces the supremum metric 3.6.

Property 8.4.15. L^∞ is a Banach space.

8.5 Product measures

8.5.1 Real hyperspace $\mathbb{R}^n$

The notions of intervals and lengths from the one dimensional case can be generalized to more dimensions in the following way:

Definition 8.5.1 (Hypercube). Let $I_1, ..., I_n$ be a sequence of intervals.

$$\mathbf{I} = I_1 \times ... \times I_n \tag{8.43}$$

Definition 8.5.2 (Generalized length). Let $\mathbf{I}$ be a hypercube induced by the sequence of intervals $I_1, ..., I_n$. The length of $\mathbf{I}$ is given by:

$$l(\mathbf{I}) = \prod_{i=1}^{n} l(I_i) \tag{8.44}$$

8.5.2   Construction of the product measure

Property 8.5.3 (General condition). The general condition for multi-dimensional Lebesgue measures is given by following equation which should hold for all $A_1 \in \mathcal{F}_1$ and $A_2 \in \mathcal{F}_2$:

$$\boxed{P(A_1 \times A_2) = P_1(A_1)P_2(A_2)} \tag{8.45}$$

Definition 8.5.4 (Section). Let $A = A_1 \times A_2$. The following two sets are called sections:

$$A_{\omega_1} = \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in A\} \subset \Omega_2$$

$$A_{\omega_2} = \{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in A\} \subset \Omega_1$$

Property 8.5.5. Let $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$. If $A \in \mathcal{F}$ then for each ω_1, $A_{\omega_1} \in \mathcal{F}_2$ and for each ω_2, $A_{\omega_2} \in \mathcal{F}_1$. Equivalently the sets $\mathcal{G}_1 = \{A \in \mathcal{F} : \forall \omega_1, A_{\omega_1} \in \mathcal{F}_2\}$ and $\mathcal{G}_2 = \{A \in \mathcal{F} : \forall \omega_2, A_{\omega_2} \in \mathcal{F}_1\}$ coincide with the product σ-algebra $\mathcal{F}$.

Property 8.5.6. The function $A_{\omega_2} \mapsto P(A_{\omega_2})$ is a step function:

$$P(A_{\omega_2}) = \begin{cases} P_1(A_1) & \text{if } \omega_2 \in A_2 \\ 0 & \text{if } \omega_2 \notin A_2 \end{cases}$$

Formula 8.5.7 (Product measure). From previous property it follows that we can write the product measure $P(A)$ in the following way:

$$\boxed{P(A) = \int_{\Omega_2} P_1(A_{\omega_2})dP_2(\omega_2)} \tag{8.46}$$

Property 8.5.8. Let P_1, P_2 be finite. If $A \in \mathcal{F}$ then the functions

$$\omega_1 \mapsto P_2(A_{\omega_1}) \qquad \omega_2 \mapsto P_1(A_{\omega_2})$$

are measurable with respect to $\mathcal{F}_1$ and $\mathcal{F}_2$ respectively and

$$\boxed{\int_{\Omega_2} P_1(A_{\omega_2})dP_2(\omega_2) = \int_{\Omega_1} P_2(A_{\omega_1})dP_1(\omega_1)} \tag{8.47}$$

Furthermore the set function P is countably additive and if any other product measure coincides with P on all rectangles, it is equal to P on the whole product σ-algebra.

8.5.3 Fubini's theorem

Property 8.5.9. Let $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ be a non-negtaive function. If f is measurable with respect to $\mathcal{F}_1 \times \mathcal{F}_2$ then for each $\omega_1 \in \Omega_1$ the function $\omega_2 \mapsto f(\omega_1, \omega_2)$ is measurable with respect to $\mathcal{F}_2$ (and vice versa). There integrals with respect to P_1 and P_2 respectively are also measurable.

Definition 8.5.10 (Section of a function). The functions $\omega_1 \mapsto f(\omega_1, \omega_2)$ and $\omega_2 \mapsto f(\omega_1, \omega_2)$ are called sections of f.

Theorem 8.5.11 (Tonelli's theorem). *Let $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ be a non-negative function. The following equalities apply:*

$$
\begin{aligned}
\int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2) d(P_1 \times P_2)(\omega_1, \omega_2) &= \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) dP_2(\omega_2) \right) dP_1(\omega_1) \\
&= \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2) dP_1(\omega_1) \right) dP_2(\omega_2)
\end{aligned}
\tag{8.48}
$$

Corollary 8.5.12 (Fubini's theorem). Let $f \in L^1(\Omega_1 \times \Omega_2)$. The sections are integrable in the appropriate spaces. Furthermore the functions $\omega_1 \mapsto \int_{\Omega_2} f dP_2$ and $\omega_2 \mapsto \int_{\Omega_1} f dP_1$ are in $L^1(\Omega_1)$ and $L^1(\Omega_2)$ respectively and equality 8.48 holds.

Remark 8.5.13. The previous construction and theorems also apply for higher dimensional product spaces. These thereoms provide a way to construct higher-dimensional Lebesgue measures m_n by defining them as the completion of the product of n one-dimensional Lebesgue measures.

8.6 Radon-Nikodym theorem

Definition 8.6.1. Let $(\Omega, \mathcal{F})$ be a measurable space. Let μ, ν be two measures defined on this space. ν is said to be **absolutely continuous with respect to** μ if

$$
\forall A \in \mathcal{F} : \mu(A) = 0 \implies \nu(A) = 0
\tag{8.49}
$$

Notation 8.6.2. This relation is denoted by $\nu \ll \mu$.

Theorem 8.6.3 (Absolute continuity). *Let μ, ν be finite measures on a measurable space $(\Omega, \mathcal{F})$. Then $\nu \ll \mu$ if and only if*

$$
\forall \varepsilon > 0 : \exists \delta > 0 : \forall A \in \mathcal{F} : \mu(A) < \delta \implies \nu(A) < \varepsilon
\tag{8.50}
$$

Property 8.2.20 can be generalized to arbitrary measure spaces as follows:

Property 8.6.4. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let $f : \Omega \to \mathbb{R}$ be a measurable function such that $\int f d\mu$ exists. Then $\nu(f) = \int_F f d\mu$ defines a measure $\nu \ll \mu$.

Definition 8.6.5 (Dominated measure). Let μ, ν be two measures. μ is said to **dominate** ν if $0 \leq \nu(F) \leq \mu(F)$ for every $F \in \mathcal{F}$.

Theorem 8.6.6 (Radon-Nikodym theorem for dominated measures).
Let μ be a measure such that $\mu(\Omega) = 1$. Let ν be a measure dominated by μ. There exists a non-negative $\mathcal{F}$-measurable function h such that $\nu(F) = \int_F h d\mu$ for all $F \in \mathcal{F}$.

Remark. The assumption $\mu(\Omega) = 1$ is non-restrictive as every other finite measure ϕ can be normalized by putting $\mu = \frac{\phi}{\phi(\Omega)}$.

Definition 8.6.7 (Radon-Nikodym derivative). The function h as defined in previous theorem is called the Radon-Nikodym derivative of ν with respect to μ and we denote it by $\frac{d\nu}{d\mu}$.

Theorem 8.6.8 (Radon-Nikodym theorem). *Let $(\Omega, \mathcal{F})$ be a measurable space. Let μ, ν be two σ-finite measures defined on this space such that $\nu \ll \mu$. There exists a non-negative measurable function $g : \Omega \to \mathbb{R}$ such that $\nu(F) = \int_F g d\mu$ for all $F \in \mathcal{F}$.*

Remark 8.6.9. The function g in the previous theorem is unique up to a μ-null (or ν-null) set.

Property 8.6.10. Let μ, ν be finite measures such that μ dominates ν. Let $h_\nu = \frac{d\nu}{d\mu}$ be the associated Radon-Nikodym derivative. For every non-negative $\mathcal{F}$-measurable function f we have

$$\int_\Omega f d\nu = \int_\Omega f h_\nu d\mu \tag{8.51}$$

Remark 8.6.11. This property also holds for all functions $f \in L^1(\mu)$.

Property 8.6.12. Let λ, ν, μ be σ-finite measures. If $\lambda \ll \mu$ and $\nu \ll \mu$ then we have:

- $\dfrac{d(\lambda + \nu)}{d\mu} = \dfrac{d\lambda}{d\mu} + \dfrac{d\lambda}{d\mu}$ a.e.

- Chain rule: if $\lambda \ll \nu$ then $\dfrac{d\lambda}{d\mu} = \dfrac{d\lambda}{d\nu}\dfrac{d\nu}{d\mu}$ a.e.

8.7 Lebesgue-Stieltjes measure

Chapter 9

Integral transforms

9.1 Fourier series

Definition 9.1.1 (Dirichlet kernel). The Dirichlet kernel is the collection of functions of the form

$$D_n(x) = \frac{1}{2\pi} \sum_{k=-n}^{n} e^{ikx} \tag{9.1}$$

Formula 9.1.2 (Sieve property). If $f \in C^1[-\pi, \pi]$ then

$$\lim_{n \to +\infty} \int_{-\pi}^{\pi} f(x) D_n(x) dx = 0 \tag{9.2}$$

Formula 9.1.3. For 2π-periodic functions, the n-th degree Fourier approximation is given by following convolution:

$$s_n(x) = \sum_{k=-n}^{n} \widetilde{f}(k) e^{ikx} = (D_n * f)(x) \tag{9.3}$$

Theorem 9.1.4 (Convergence of the Fourier series). *Let $f : \mathbb{R} \to \mathbb{R}$ be a 2π-periodic function. If $f(x)$ is piecewise C^1 on $[-\pi, \pi]$ the the limit $\lim_{n \to +\infty} (D_n * f)(x)$ converges to $\frac{f(x+) + f(x-)}{2}$ for all $x \in \mathbb{R}$.*

Formula 9.1.5 (Generalized Fourier series). Let $f(x) \in \mathcal{L}^2[-l, l]$ be a $2l$-periodic function. This function can be approximated by the following series:

$$\boxed{f(x) = \sum_{n=-\infty}^{+\infty} \left(\frac{1}{2l} \int_{-l}^{l} e^{-i\frac{n\pi x'}{l}} f(x') dx' \right) e^{i\frac{n\pi x}{l}}} \tag{9.4}$$

Formula 9.1.6 (Fourier coefficients). As seen in the general formula, the Fourier coefficient $\widetilde{f}(n)$ can be calculated by taking the inner product 16.5 of $f(x)$ and the n-th eigenfunction e_n:

$$\widetilde{f}(n) = \langle e_n | f \rangle = \int_{-l}^{l} e_n^*(x) f(x) dx \qquad \text{with} \qquad e_n = \sqrt{\frac{1}{2l}} e^{i\frac{n\pi x}{l}} \tag{9.5}$$

Definition 9.1.7 (Periodic extension). Let $f(x)$ be piecewise C^1 on $[-L, L]$. The periodic extension $f^L(x)$ is defined by repeating the restriction of $f(x)$ to $[-L, L]$ every $2L$. The **normalized periodic extension** is defined as

$$f^{L,\nu}(x) = \frac{f^L(x+) + f^L(x-)}{2} \tag{9.6}$$

Theorem 9.1.8. *If $f : \mathbb{R} \to \mathbb{R}$ is piecewise C^1 on $[-L, L]$ then the Fourier series approximation of $f(x)$ converges to $f^{L,\nu}(x)$ for all $x \in \mathbb{R}$.*

9.2 Fourier transform

The Fourier series can be used to expand a $2l$-periodic function as an infinite series of exponentials. For expanding a non-periodic function we need the Fourier integral:

$$\boxed{\mathcal{F}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt} \tag{9.7}$$

$$f(t) = \mathcal{F}^{-1}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathcal{F}(\omega) e^{i\omega t} d\omega \tag{9.8}$$

Equation 9.7 is called the (forward) Fourier transform of $f(t)$ and equation 9.8 is called the inverse Fourier transform.

Notation 9.2.1. The Fourier transform of a function $f(t)$, as seen in equation 9.7, is often denoted by $\widetilde{f}(\omega)$.

Theorem 9.2.2 (Convergence of the Fourier integral). *If $f : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous (see 5.3) and if $\int_{-\infty}^{+\infty} |f(x)| dx$ is convergent then the Fourier integral converges to $f(x)$ for all $x \in \mathbb{R}$.*

Theorem 9.2.3 (Fourier inversion theorem). *If both $f(t), \mathcal{F}(\omega) \in \mathcal{L}^1(\mathbb{R})$ are continuous then the Cauchy principal value in 9.8 can be replaced by a normal integral.*

Remark 9.2.4. Schwartz functions (see 10.1) are continuous elements of $\mathcal{L}^1(\mathbb{R})$ and as such the Fourier inversion theorem also holds for these functions. This is interesting because checking the conditions for Schwartz functions is often easier then checking the more general conditions of the theorem.

Property 9.2.5. From the Riemann-Lebesgue lemma 8.2.27 it follows that

$$\mathcal{F}(\omega) \to 0 \quad \text{if} \quad |\omega| \to 0 \tag{9.9}$$

Property 9.2.6 (Parceval's theorem). Let $(f, \widetilde{f})$ and $(g, \widetilde{g})$ be two Fourier transform pairs.

$$\int_{-\infty}^{+\infty} f(x) g(x) dx = \int_{-\infty}^{+\infty} \widetilde{f}(k) \widetilde{g}(k) dk \tag{9.10}$$

Corollary 9.2.7 (Plancherel theorem). 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:

$$\int_{-\infty}^{+\infty} |f(x)|^2 dx = \int_{-\infty}^{+\infty} |\widetilde{f}(k)|^2 dk \tag{9.11}$$

9.2.1 Convolution

Formula 9.2.8 (Convolution).

$$\boxed{(f * g)(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau} \tag{9.12}$$

Property 9.2.9 (Commutativity).

$$f * g = g * f \tag{9.13}$$

Theorem 9.2.10 (Convolution Theorem).

$$\widetilde{f * g} = \widetilde{g}\widetilde{f} \tag{9.14}$$

9.3 Laplace transform

Formula 9.3.1 (Laplace transform).

$$\mathcal{L}\{F(t)\}_{(s)} = \int_{0}^{\infty} f(t)e^{-st}dt \tag{9.15}$$

Formula 9.3.2 (Bromwich integral).

$$f(t) = \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} \mathcal{L}\{F(t)\}_{(s)}e^{st}ds \tag{9.16}$$

Notation 9.3.3. The Laplace transform as defined in equation 9.15 is sometimes denoted by $f(s)$.

9.4 Mellin transform

Formula 9.4.1 (Mellin transform).

$$\mathcal{M}\{f(x)\}(s) = \int_{0}^{+\infty} x^{s-1}f(x)dx \tag{9.17}$$

Formula 9.4.2 (Inverse Mellin transform).

$$f(x) = \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} \mathcal{M}\{f(x)\}_{(s)}x^{-s}ds \tag{9.18}$$

9.5 Integral representations

Formula 9.5.1 (Heaviside step function).

$$\theta(x) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{e^{ikx}}{x - i\varepsilon} dk \tag{9.19}$$

Formula 9.5.2 (Dirac delta function).

$$\delta^{(n)}(\vec{x}) = \frac{1}{(2\pi)^n} \int_{-\infty}^{\infty} e^{i\vec{k}\cdot\vec{x}} d^n k \tag{9.20}$$

Chapter 10

Distributions

10.1 Generalized function

Definition 10.1.1 (Schwartz space). The Schwartz space or **space of rapidly decreasing functions**[1] $S(\mathbb{R})$ is defined as:

$$S(\mathbb{R}) = \left\{ f(x) \in C^\infty(\mathbb{R}) : \forall i, j \in \mathbb{N} : \forall x \in \mathbb{R} : |x^i f^{(j)}(x)| < +\infty \right\} \qquad (10.1)$$

Remark 10.1.2. This definition can be generalized to functions of the class $C^\infty(\mathbb{R}^n)$ or functions $f : \mathbb{R} \to \mathbb{C}$. The Schwartz space is then denoted by $S(\mathbb{R}, \mathbb{C})$.

Definition 10.1.3 (Functions of slow growth). The set of functions of slow growth $N(\mathbb{R})$ is defined as:

$$N(\mathbb{R}) = \{ f(x) \in C^\infty(\mathbb{R}) : \forall i \in \mathbb{N}, \exists M_i > 0 : |f^{(i)}(x)| = O(|x|^i) \text{ for } |x| \to +\infty \} \qquad (10.2)$$

Remark. It is clear that all polynomials belong to $N(\mathbb{R})$ but not to $S(\mathbb{R})$.

Property 10.1.4. If $f(x) \in S(\mathbb{R})$ and $a(x) \in N(\mathbb{R})$ then $a(x)f(x) \in S(\mathbb{R})$.

Definition 10.1.5 (Generalized function). Let $g(x) \in S(\mathbb{R})$ be a test function. Let $\{f_n(x) \in S(\mathbb{R})\}, \{h_n(x) \in S(\mathbb{R})\}$ be sequences such that

$$\lim_{n \to +\infty} \langle f_n(x)|g(x) \rangle = \lim_{n \to +\infty} \int_{-\infty}^{+\infty} f_n(x)g(x)dx$$

and similarly for h_n. Define the equivalence relation $\{f_n(x) \in S(\mathbb{R})\} \sim \{h_n(x) \in S(\mathbb{R})\}$ by saying that the two sequences, satisfying the previous condition, are equivalent if and only if

$$\lim_{n \to +\infty} \langle f_n(x)|g(x) \rangle = \lim_{n \to +\infty} \langle h_n(x)|g(x) \rangle$$

A generalized function is defined as a complete equivalence class under previous relation.

---

[1]These functions are said to be rapidly decreasing because every derivative $f^{(j)}(x)$ decays faster than any polynomial x^i for $x \to +\infty$.

Notation 10.1.6. Let ψ be a generalized function. Let $f \in S(\mathbb{R})$. The inner product 8.36 is generalized by following functional:

$$\langle \psi | f \rangle = \lim_{n \to +\infty} \int_{-\infty}^{+\infty} \psi_n(x) f(x) dx \tag{10.3}$$

Property 10.1.7. Let ψ be a generalized function. Let $f(x) \in S(\mathbb{R})$. The previous functional has following properties:

- $\forall i \in \mathbb{N} : \langle \psi^{(i)} | f \rangle = (-1)^i \langle \psi | f^{(i)} \rangle$

- $\forall a, b \in \mathbb{R}, a \neq 0 : \langle \psi(ax+b) | f(x) \rangle = |a|^{-1} \langle \psi(x) | f(x - b/a) \rangle$

- $\forall a(x) \in N(\mathbb{R}) : \langle a\psi | f \rangle = \langle \psi | af \rangle$

Property 10.1.8 (Ordinary function as generalized function). Let $f : \mathbb{R} \to \mathbb{C}$ be a function such that $\exists M \geq 0 : (1+x^2)^{-M} |f(x)| \in L(\mathbb{R}, \mathbb{C})^2$. There exists a generalized function $\psi \sim \{f_n(x) \in S(\mathbb{R}, \mathbb{C})\}$ such that for every $g(x) \in S(\mathbb{R}, \mathbb{C})$:

$$\langle \psi | g \rangle = \langle f | g \rangle$$

Furthermore if $f(x)$ is continuous on an interval, then $\lim_{n \to +\infty} f_n(x) = f(x)$ converges pointwise on that interval.

10.2 Dirac Delta distribution

Definition 10.2.1 (Heaviside function). Define the generalized function $H \sim \{H_n(x) \in S(\mathbb{R})\}$ as:

$$H(x) = \begin{cases} 0 & if \quad x < 0 \\ 1 & if \quad x \geq 0 \end{cases} \tag{10.4}$$

From this definition it follows that for every $f \in S(\mathbb{R})$:

$$\langle H | f \rangle = \int_0^{+\infty} f(x) dx \tag{10.5}$$

Remark 10.2.2. For the above integral to exist, $f(x)$ does not need to be an element of $S(\mathbb{R})$. It is a sufficient condition, but not a necessary one.

Definition 10.2.3 (Generalized delta function). The Dirac delta function is defined as a representant of the equivalence class of generalized functions $\{H_n'(x) \in S(\mathbb{R})\}$. By equations 10.1.7 and 10.5 we have for every $f \in S(\mathbb{R})$:

$$\begin{aligned} \langle \delta | f \rangle &= \langle H' | f \rangle \\ &= -\langle H | h \rangle \\ &= -\int_0^{+\infty} f'(x) dx \\ &= f(0) \end{aligned} \tag{10.6}$$

---

[2]The space of Lebesgue integrable functions 8.2.18.

Property 10.2.4 (Sampling property). The result from previous definition can be generalized in the following way:

$$\boxed{f(x_0) = \int_{\mathbb{R}} f(x)\delta(x - x_0)dx} \tag{10.7}$$

Example 10.2.5 (Dirac comb).

$$III_b(x) = \sum_n \delta(x - nb) \tag{10.8}$$

Property 10.2.6. Let $f(x) \in C^1(\mathbb{R})$ be a function with roots at $x_1, x_2, ..., x_n$ such that $f'(x_i) \neq 0$. The Dirac delta distribution has the following property:

$$\delta[f(x)] = \sum_{i=1}^n \frac{1}{|f'(x_i)|}\delta(x - x_i) \tag{10.9}$$

Property 10.2.7 (Convolution with delta function). Let $f(x) \in S(\mathbb{R})$. Let $\otimes$ denote the convolution.

$$\delta(x) \otimes f(x) = \int_{-\infty}^{+\infty} \delta(x - \alpha)f(\alpha)d\alpha = f(x) \tag{10.10}$$

Formula 10.2.8 (Differentiation across discontinuities). Let $f(x)$ be a piecewise continuous function with discontinuities at $x_1, ..., x_n$. Let f satisfy the conditions to be a generalized function. Define $\sigma_i = f^+(x_i) - f^-(x_i)$ which are the jumps of f at its discontinuities. Next, define the function

$$f_c(x) = f(x) - \sum_{i=1}^n \sigma_i H(x - x_i)$$

which is a continuous function. Differentiation gives

$$f'(x) = f_c'(x) + \sum_{i=1}^n \sigma_i \delta(x - x_i)$$

It follows that the derivative in a generalized sense of a piecewise continuous function equals the derivative in the classical sense plus a summation of delta functions at every jump discontinuity.

10.3 Fourier transform

Theorem 10.3.1. *Let $f(x), F(k)$ be a Fourier transform pair. If $f(x) \in S(\mathbb{R}, \mathbb{C})$, then $F(k) \in S(\mathbb{R}, \mathbb{C})$. It follows that for a sequence $\{f_n(x) \in S(\mathbb{R}, \mathbb{C})\}$ the sequence of Fourier*

transformed functions $\{F_n(x) \in S(\mathbb{R}, \mathbb{C})\}$ is also a subset of the Schwartz space. Furthermore Parceval's theorem 9.10 gives

$$\int_{-\infty}^{+\infty} f_n(x)g(x)dx = \int_{-\infty}^{+\infty} F_n(x)G(x)dx \in \mathbb{R}$$

where $g(x) \in S(\mathbb{R}, \mathbb{C})$. From these two properties it follows that the Fourier transform of a generalized functions is also a generalized functions.

Property 10.3.2. Let ψ be a generalized function with Fourier transform Ψ. Let $f(x) \in S(\mathbb{R}, \mathbb{C})$ with Fourier transform $F(k)$. We have the following equality:

$$\langle \psi | F \rangle = \langle \Psi | f \rangle \tag{10.11}$$

Formula 10.3.3 (Fourier representation of delta function).

$$\delta(x - a) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{ik(x-a)} dk \tag{10.12}$$

Chapter 11

Ordinary differential equations

11.1   Boundary conditions

Unigue solutions of a differential equation are obtained by supplying additional conditions. These are called boundary conditions.

11.1.1   Periodic boundary conditions

Periodic boundary conditions are conditions of the following form:

$$y(x) = y(x + \varphi) \tag{11.1}$$

By induction it follows that for every n:

$$y(x) = y(x + n\varphi) \tag{11.2}$$

11.1.2   Dirichlet boundary conditions

Dirichlet boundary conditions are conditions of the following form:

$$y(x) = f(x) \qquad , \qquad x \in \partial\Omega \tag{11.3}$$

where Ω is the domain of the problem.

Remark 11.1.1. When y is a function of multiple variables, α can be a function as well. For example (in spherical coordinates: ρ, ϕ, θ):

$$y(x, \phi, \theta) = \alpha(\phi, \theta) \tag{11.4}$$

11.1.3 Neumann boundary conditions

Neumann boundary conditions are conditions of the following form:

$$y'(a) = \alpha \tag{11.5}$$

Remark 11.1.2. When y is a function of multiple variables, we obtain the following form (where S is the boundary of the domain and $\hat{n}$ a normal vector to this boundary):

$$\frac{\partial y}{\partial \hat{n}}(\vec{x}) = f(\vec{x}) \quad , \quad \vec{x} \in S \tag{11.6}$$

11.2 First order ODE's

Formula 11.2.1 (First order ODE).

$$\boxed{y'(t) + a(t)y(t) = R(t)} \tag{11.7}$$

If the function $R(t)$ is identically zero, then the ODE is said to be **homogenous**.

Theorem 11.2.2. *Let $U \subseteq \mathbb{R}$ be an open set. Let the functions $a(t), R(t) : U \to \mathbb{R}$ be continuous. The solutions $\varphi(t) : U \to \mathbb{R}$ of equation 11.7 are given by:*

$$\boxed{\varphi(t) = e^{-\int a(t)dt}\left(c + \int R(t)e^{\int a(t)dt}dt\right)} \tag{11.8}$$

where c is a constant.

11.3 Second order ODE's

Formula 11.3.1 (Second order ODE).

$$\boxed{y''(t) + a(t)y'(t) + b(t)y(t) = R(t)} \tag{11.9}$$

Formula 11.3.2 (Homogeneous second order ODE).

$$y''(t) + a(t)y'(t) + b(t)y(t) = 0 \tag{11.10}$$

11.3.1 General solution

Formula 11.3.3. Let $\varphi : U \to \mathbb{R}$ be a nowhere zero solution of the homogeneous equation 11.10. The general solution of equation 11.9 is then given by:

$$\boxed{y(t) = c_1 \, \varphi + c_2 \, \varphi \int \frac{e^{-\int a}}{\varphi^2} + \psi_0} \tag{11.11}$$

where ψ_0 is a particular solution of equation 11.9.

Theorem 11.3.4. *Let ψ_0 be a solution of equation 11.9. The set of all solutions is given by the affine space:*

$$\{\psi_0 + \chi : \chi \text{ is a solution of the homogeneous equation 11.10}\} \tag{11.12}$$

Theorem 11.3.5. *Two solutions of the homogeneous equation 11.10 are independent if the wronskian is nonzero:*

$$W(\varphi_1(x), \varphi_2(x)) = \left| \begin{array}{cc} \varphi_1(x) & \varphi_2(x) \\ \varphi_1'(x) & \varphi_2'(x) \end{array} \right| \neq 0 \tag{11.13}$$

Formula 11.3.6 (Abel's identity). An explicit formula for the wronskian is given by:

$$W(x) = W(x_0) \exp\left(- \int_{x_0}^{x} a(x')dx' \right) \tag{11.14}$$

11.3.2 Constant coefficients

Theorem 11.3.7. *A map $\varphi : U \to \mathbb{C}$ is a complex solution of equation 11.10 if and only if $Re\{\varphi\}$ and $Im\{\varphi\}$ are real solutions of equation 11.10.*

Formula 11.3.8 (Characteristic equation). When having an ODE of the form[1] :

$$y''(t) + py'(t) + qy(t) = 0 \tag{11.15}$$

where p and q are constants, we define the characteristic equation as follows:

$$\lambda^2 + p\lambda + q = 0 \tag{11.16}$$

This polynomial equation generally[2] has two distinct (complex) roots λ_1 and λ_2. From these roots we can derive the solutions of equation 11.15 using the following rules (c_1 and c_2 are constants):

- $\lambda_1 \neq \lambda_2$, λ_1 and $\lambda_2 \in \mathbb{R}$: $y(t) = c_1 \ e^{\lambda_1 t} + c_2 \ e^{\lambda_2 t}$
- $\lambda_1 = \lambda_2$: $y(t) = c_1 \ e^{\lambda t} + c_2 \ te^{\lambda t}$
- $\lambda_1 = \lambda_2^*$, where $\lambda_1 = a + ib$: $y(t) = c_1 \ e^{at} \cos(bt) + c_2 \ e^{at} \sin(bt)$

---

[1] Any other form of homogeneous second order ODE's with constant coefficients can be rewritten in this form.

[2] See theorem 14.1.3 ("Fundamental theorem of algebra").

11.3.3 Method of Frobenius

Formula 11.3.9 (Method of Frobenius). To find a solution of the homogeneous equation 11.10 we assert a solution of the form:

$$y(x) = \sum_{i=0}^{\infty} a_i (x - x_0)^{i+k} \tag{11.17}$$

where k is a constant.

Definition 11.3.10 (Indicial equation). After inserting the solution 11.17 into the homogeneous equation 11.10 we obtain[3] an equation of the form $\sum_{i=n}^{\infty} H_i(k) x^i = 0$ where $n \in \mathbb{R}$ and $H_i(k)$ is a polynomial in k. This means that for every i we obtain an equation of the form $H_i(k) = 0$, due to the independence of polynomial terms. The equation for the lowest power will be quadratic in k and it is called the indicial equation.

Theorem 11.3.11. *The indicial equation generally has two roots k_1, k_2. The following possibilities arise:*

- $k_1 = k_2$: *Only one solution will be found with the method of Frobenius (another one can be found as in the second term of equation 11.11)*

- $k_1 - k_2 \in \mathbb{Z}$: *A second independent solution might be obtained using this method. If not, then 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 11.3.12 (Fuch's theorem). *If $a(z)$ and $b(z)$ are analytic at $z = z_0$ then the general solution $y(z)$ can be expressed as a Frobenius' series.*

11.4 Sturm-Liouville theory

Definition 11.4.1 (Sturm-Liouville boundary value problem). The following ODE, subject to mixed boundary conditions, is called a Sturm-Liouville boundary value problem:

$$\frac{d}{dx}\left[p(x)\frac{dy}{dx}\right] + \left[g(x) + \lambda r(x)\right] y(x) = 0 \tag{11.18}$$

where $p(x), q(x)$ and $r(x)$ are continuous on $a \leq x \leq b$. $p(x) \in C^1(a, b)$ with $p(x) < 0$ or $p(x) > 0$ for $a \leq x \leq b$. $r(x) \geq 0$ or $r(x) \leq 0$ for $a \leq x \leq b$ and $r(x)$ is not identically zero on any subinterval.

The boundary conditions are given by

$$\begin{cases} \alpha_1 y(a) + \beta_1 y'(a) = 0 \\ \alpha_2 y(b) + \beta_2 y'(b) = 0 \end{cases} \tag{11.19}$$

where at least one of the constants $\alpha_1, \alpha_2, \beta_1$ or β_2 is non-zero.

---

[3]It is important to 'sync' the power of all terms in order to obtain one 'large' coefficient.

Formula 11.4.2. The solutions are of the form

$$y(x) = c_1 u_1(\lambda; x) + c_2 u_2(\lambda; x)$$

Only for certain values of λ will these solutions (u_1, u_2) be non-trivial. The values of λ for which the solutions are non-trivial are called **eigenvalues** and the associated solutions are called **eigenfunctions**. Substituting this form in the boundary conditions gives the following determinant condition for non-trivial solutions, which is also the defining equation of the eigenvalues λ:

$$\begin{vmatrix} \alpha_1 u_1(a; \lambda) + \beta_1 u_1'(a; \lambda) & \alpha_1 u_2(a; \lambda) + \beta_1 u_2'(a; \lambda) \\ \alpha_1 u_1(b; \lambda) + \beta_1 u_1'(b; \lambda) & \alpha_1 u_2(b; \lambda) + \beta_1 u_2'(b; \lambda) \end{vmatrix} = 0 \qquad (11.20)$$

The independent eigenfunctions can be found by substituting the found eigenvalues in the ODE 11.18.

Definition 11.4.3 (Self-adjoint form). The SL-problem can be rewritten as

$$\left[\hat{\mathcal{L}} + \lambda r(x) \right] y(x) = 0$$

The operator $\hat{\mathcal{L}} = \frac{d}{dx} \left[p(x) \frac{d}{dx} + g(x) \right]$ is called the self-adjoint form. Now consider the general linear ODE

$$\left[a_2(x) \frac{d^2}{dx^2} + a_1(x) \frac{d}{dx} + a_0(x) \right] y(x) = 0 \qquad (11.21)$$

This equation can be rewritten in a self-adjoint form by setting:

$$p(x) = e^{\int \frac{a_1}{a_2} dx} \qquad \text{and} \qquad g(x) = \frac{a_0}{a_2} e^{\int \frac{a_1}{a_2} dx} \qquad (11.22)$$

Property 11.4.4. The eigenfunctions corresponding to distinct eigenvalues are orthogonal with respect to the weight function $r(x)$.

Theorem 11.4.5 (Oscillation theorem). *Let f_n be the nth eigenfunction of a Sturm-Liouville boundary condition problem. Then f_n has precisely $n - 1$ roots.*

Chapter 12

Partial differential equations

12.1   General linear equations

Formula 12.1.1 (Cramer's rule). Let $Ax = b$ be a system of linear equations where the matrix A has a nonzero determinant. Then Cramer's rule gives a unique solution where the unknowns are given by;

$$x_i = \frac{\det(A_i)}{\det(A)} \tag{12.1}$$

where A_i is the matrix obtained by replacing the ith column of A by the column matrix b.

Definition 12.1.2 (Characteristic curve). Curve along which the highest order partial derivatives are not uniquely defined.

12.2   First order PDE

Formula 12.2.1 (First order quasilinear PDE).

$$\boxed{P(x,y,z)\frac{\partial z}{\partial x} + Q(x,y,z)\frac{\partial z}{\partial y} = R(x,y,z)} \tag{12.2}$$

Formula 12.2.2 (Characteristic curve). The PDE will have no unique solution if

$$\begin{vmatrix} P & Q \\ dx & dy \end{vmatrix} = 0 \tag{12.3}$$

and will have a non-unique solution if

$$\begin{vmatrix} P & R \\ dx & dz \end{vmatrix} = 0 \tag{12.4}$$

The characteristic curves are thus defined by $\frac{dx}{P} = \frac{dy}{Q}$ and along thse curves the condition $\frac{dx}{P} = \frac{dz}{R}$ should hold to ensure a solution.

Theorem 12.2.3. *The general solution of 12.2 is implicitly given by $F(\xi, \eta) = 0$ with $F(\xi, \eta)$ an arbitrary differentiable function where $\xi(x, y, z) = c_1$ and $\eta(x, y, z) = c_2$ are solutions of the equation*

$$\frac{dx}{P} = \frac{dy}{Q} = \frac{dz}{R} \tag{12.5}$$

where c_1, c_2 are constants which are fixed by boundary conditions.

Remark 12.2.4. Looking at the defining equations of the characteristic curve, it is clear that these fix the general solution of the PDE.

12.3   Characteristics

Formula 12.3.1 (Second order quasilinear PDE). Consider the following pseudolinear differential equation for the function $u(x, y)$:

$$R(x, y)u_{xx} + S(x, y)u_{xy} + T(x, y)u_{yy} = W(x, y, u, p, q) \tag{12.6}$$

where $p = u_x$ and $q = u_y$.

Formula 12.3.2 (Equation of characteristics). Consider the following two differential equations

$$\begin{cases} u_{xx}dx + u_{xy}dy = dp \\ u_{xy}dx + u_{yy}dy = dq \end{cases} \tag{12.7}$$

According to Cramer's rule 12.1 these equations, together with the PDE 12.6, give the following condition for the characteristic curves:

$$\begin{vmatrix} R(x, y) & S(x, y) & T(x, y) \\ dx & dy & 0 \\ 0 & dx & dy \end{vmatrix} = 0 \tag{12.8}$$

which is equivalent to following equation:

$$\boxed{R\left(\frac{dy}{dx}\right)^2 - S\left(\frac{dy}{dx}\right) + T = 0} \tag{12.9}$$

Definition 12.3.3 (Types of characteristics). Equation 12.9 is quadratic in $\frac{dy}{dx}$. If this equation has two distinct real roots then 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 12.3.4 (Canonical form). Consider the general change of variables $\xi = \xi(x, y)$, $\eta = \eta(x, y)$ and $z = \zeta$. With this change, the PDE 12.6 becomes:

$$A(\xi_x, \xi_y)\frac{\partial^2\zeta}{\partial\xi^2} + 2B(\xi_x, \xi_y, \eta_x, \eta_y)\frac{\partial^2\zeta}{\partial\xi\partial\eta} + A(\eta_x, \eta_y)\frac{\partial^2\zeta}{\partial\eta^2} = F(\xi, \eta, \zeta, \zeta_\xi, \zeta_\eta) \tag{12.10}$$

where $A(a, b) = Ra^2 + Sab + Tb^2$ and $B = R\xi_x\eta_x + \frac{1}{2}S(\xi_x\xi_y + \eta_x\eta_y) + Tbd$. Solving the quadratic equation 12.9 will lead to the following three canonical forms:

- **hyperbolic PDE**: With the solutions $\lambda_1(x, y)$ and $\lambda_2(x, y)$ the defining equation can be separated into two ODE's

$$\left(\frac{dy}{dx} + \lambda_1(x, y) \right) \left(\frac{dy}{dx} + \lambda_2(x, y) \right) = 0$$

It is clear that the solutions of these ODE's are also roots of the $A(a, b)$ coefficients such that the change of variables $\xi = f_1(x, y)$ and $\eta = f_2(x, y)$ gives the canonical hyperbolic form

$$\frac{\partial^2 \zeta}{\partial \xi \partial \eta} = H(\xi, \eta, \zeta, \zeta_\xi, \zeta_\eta) \qquad (12.11)$$

where $H = \frac{F}{2B}$.

- **parabolic PDE**: As in the hyperbolic case we perform the change of variable $\xi = f(x, y)$, however there is only one root of the defining equation so the second variable can be chosen randomly, yet indepedent of $f_1(x, y)$. From the condition $S^2 + 4RT = 0$ it is alos possible to derive the condition 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

$$\frac{\partial^2 \zeta}{\partial \eta^2} = G(\xi, \eta, \zeta, \zeta_\xi, \zeta_\eta) \qquad (12.12)$$

where $G = \frac{F}{A(\eta_x, \eta_y)}$.

- **elliptic PDE**: Again there are two (complex) roots, so the A coefficients will dissappear. 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 hyperbolic case results in the following elliptic canonical form

$$\frac{\partial^2 \zeta}{\partial \alpha^2} + \frac{\partial^2 \zeta}{\partial \beta^2} = K(\alpha, \beta, \zeta, \zeta_\alpha, \zeta_\beta) \qquad (12.13)$$

Theorem 12.3.5 (Maximum principle). *Consider a PDE of the parabolic or elliptic type. The maximum of the solution on a domain is to be found on the boundary of that domain.*

12.3.1 D'Alemberts method

Consider the wave equation

$$\frac{\partial^2 u}{\partial x^2}(x, t) = \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2}(x, t) \qquad (12.14)$$

By applying the method from previous subsection, it is clear that the characteristics are given by

$$\xi = x + ct \qquad \text{and} \qquad \eta = x - ct \qquad (12.15)$$

Furthermore, it follows that the wave equation is a hyperbolic equation which can be rewritten in the canonical form:

$$\frac{\partial^2 u}{\partial \xi \partial \eta}(\xi, \eta) = 0 \tag{12.16}$$

Integration with respect to ξ and η and rewriting the solution in terms of x and t gives

$$u(x,t) = f(x + ct) + g(x - ct) \tag{12.17}$$

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 conditions

$$u(x,0) = v(x) \qquad \text{and} \qquad \frac{\partial u}{\partial t}(x,0) = q(x) \tag{12.18}$$

By applying these conditions to the general solution 12.17 it can be shown that the general solution subject to the given boundary conditions is given by:

$$\boxed{u(x,t) = \frac{1}{2}\left[v(x+ct) + v(x-ct)\right] + \frac{1}{2c}\int_{x-ct}^{x+ct} q(z)dz} \tag{12.19}$$

Remark 12.3.6. Because x is not bounded, this solution is only valid for infinite strings.

12.4 Separation of variables

Remark. We begin this section with the remark that solutions obtained by this method 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 obtained solutions.

12.4.1 Cartesian coordinates

Method 12.4.1 (Separation of variables). Let $\hat{\mathcal{L}}$ be the operator associated with a partial diferential equation such that $\hat{\mathcal{L}}u(\vec{x}) = 0$ where $\vec{x} = (x_1, ..., x_n)$ is the set of variables. A useful method is to propose a solution of the form

$$u(\vec{x}) = \prod_{i=1}^{n} u_i(x_i)$$

By substituting this form in the PDE and using (basic) algebra it is sometimes (!!) possible to reduce the partial differential equation to a system of n ordinary differential equations.

Example 12.4.2. Consider following PDE:

$$\frac{\partial u}{\partial t} - a\frac{\partial^2 u}{\partial x^2} = 0 \tag{12.20}$$

Substituting a solution of the form $u(x,t) = X(x)T(t)$ gives

$$X(x)\frac{dT(t)}{dt} - aT(t)\frac{d^2 X(x)}{dx^2} = 0$$

which can be rewritten as (the arguments are dropped for convenience)

$$\frac{1}{aT}\frac{dT}{dt} = \frac{1}{X}\frac{d^2 X}{dx^2}$$

As both sides are independent, it is clear that they are equal to a constant, say λ. This results in the following system of ordinary differential equations:

$$\begin{cases} X''(x) &= \lambda X(x) \\ T'(t) &= a\lambda T(t) \end{cases}$$

12.4.2 Dirichlet problem

The (interior) Dirichlet problem[1] 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. The uniqueness of this solution can be proven with the maximum principle 12.3.5 if the PDE is of the elliptic kind (!!) such as the Laplace equation[2].

Proof. Let ϕ, ψ be two solutions of the interior Dirichlet problem. Due to the 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 $\phi \leq \psi$ in the complete domain. Which means that $\phi = \psi$ in the complete domain. $\square$

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.

12.5 Non-homogeneous boundary conditions

Formula 12.5.1 (Non-homogeneous boundary condition).

$$\alpha u(a,t) + \beta\frac{\partial u}{\partial x}(a,t) = h(t) \tag{12.21}$$

When $h(t)$ is identically zero, the boundary condition becomes homogeneous.

---

[2]Think of the Dirichlet boundary condition 11.3.
[2]The Dirichlet boundary problem originated with the Laplace equation.

Method 12.5.2 (Steady-state solution). Assume that the function $h(t)$ is constant. In this case it is useful to rewite the solution as

$$u(x,t) = v(x) + w(x,t)$$

The 'time'-independent function is called the steady-state solution and the function $w(x,t)$ represents the deviation of this steady-state scenario.

As the PDE is linear, we require the partial solutions $v(x)$ and $w(x,t)$ to individually satisfy the equation. Furthermore we require the function $v(x)$ to also satisfy the given non-homogeneous boundary conditions. This results in $w(x,t)$ 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, we also have $\alpha v(a) + \beta \frac{\partial v}{\partial x}(a) = u_0$. Combining these two conditions gives

$$\alpha \left[v(a) + w(a,t) \right] + \beta \left[\frac{\partial v}{\partial x}(a) + \frac{\partial w}{\partial x}(a,t) \right] = \alpha v(a) + \beta \frac{\partial v}{\partial x}(a)$$

which can be reduced to

$$\alpha w(a,t) + \beta \frac{\partial w}{\partial x}(a,t) = 0$$

The steady-state deviation $w(x,t)$ thus satisfies homogeneous boundary conditions. □

Method 12.5.3. If the function $h(t)$ is not a constant, we use a different method. Rewrite the solution as $u(x,t) = v(x,t) + w(x,t)$ where we only require $v(x,t)$ to be some function that satisfies the boundary conditions (and not the PDE)[3]. This will lead to $w(x,t)$ satisfying the homogeneous boundary conditions as in the previous method. After substituting the function $v(x,t)$ in the PDE, we obtain a differential equation for $w(x,t)$ but it can be non-homogeneous.

Method 12.5.4. A third, sometimes useful, method is the following. If the problem consists of 3 homogeneous and 1 non-homogeneous boundary condition then the problem can be solved by first applying the homogenous conditions to restrict the values of the separation constant and obtain a series expansion. Afterwards the obtained series can be fitted to the non-homogeneous condition to obtain the final remaining coefficients.

If there is more than 1 non-homogeneous boundary condition, the method can be extended. Let there be j boundary conditions. Rewrite the general solution as $u(x,t) = \sum_{i=1}^{j} v_j(x,t)$ where $v_j(x,t)$ satisfies the jth non-homogeneous 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 $v_j(x,t)$ to obtain a series expansion.

Method 12.5.5 (Non-homogeneous PDE). A possible way to solve non-homogeneous second order partial differential equations of the form

$$\hat{\mathcal{L}} u(x,t) = f(x,t)$$

---

[3]As there are infinitely many possible functions that satisfy the boundary conditions, the best choice for $v(x,t)$ is the one that makes the equation for $w(x,t)$ as simple as possible.

given a set of homogeneous boundary conditions and inital value conditions $w(x, 0) = \psi(x)$, is the following, where we assume all involved functions to be expandable as a generalized Fourier series:

1. Solve the homogeneous version of the PDE, which will result in a series expansion $\sum_n w_n(t)e_n(x)$, where $e_n(x)$ are a complete set of eigenfunctions in the variable x. This solution should satisfy the (homogeneous[4]) boundary conditions.

2. Expand the function $f(x, t)$ in the same way as $u(x, t)$. The coefficients f_n can be found by using the orthogonality realtions of the functions $e_n(x)$.

3. Inserting these expansions in the original PDE and rewriting the equation will lead to a summation of the form:

$$\sum_n \left[\left(\hat{D} w_n(t) \right) e_n(x) \right] = 0$$

   where $\hat{D}$ is a linear first order differential operator. As all terms are independent, this gives n first order ODE's to obtain the functions $w_n(t)$. These can be generally solved by using formula 11.8.

4. Initial value conditions for the functions $w_n(t)$ are applied by setting $t = 0$ in the series expansion of $u(x, t)$ and equating it with the series expansion of $\psi(x)$. This results in $w_n() = \Psi_n$.

5. The obtained ODE's together with the found boundary conditions $w_n(0) = \Psi_n$ will give the solutions of $w_n(t)$.

6. Entering these solutions in the series expansion of $u(x, t)$ will give the general solution of the non-homogeneuous PDE.

Remark 12.5.6. It is clear that the requirement that all involved functions are expandable as a generalized Fourier series is restricting. Not all non-homogeneous PDE's are solvable with this method.

---

[4]Non-homogeneous boundary conditions can be turned homogeneous by the previous two methods.

Chapter 13

Bessel functions

13.1 Bessel's differential equation (BDE)

A Bessel's differential equation is an ordinary differential equation of the following form:

$$\boxed{z^2 y'' + z y' + (z^2 - n^2) y = 0} \tag{13.1}$$

The solutions of this ODE are the Bessel functions of the first and second kind (also called respectivelly Bessel and Neumann functions).

$$J_n(z) = \sum_{m=0}^{\infty} \frac{(-1)^m}{m!(m+n)!} \left(\frac{z}{2}\right)^{2m+n} \tag{13.2}$$

$$N_n(z) = \lim_{\nu \to n} \frac{cos(\nu\pi) J_N(z) - J_{-n}(z)}{sin(\nu\pi)} \tag{13.3}$$

Remark. Solution 13.2 can be found by using Frobenius' method.

Property 13.1.1. For $n \notin \mathbb{N}$ the solutions $J_n(z)$ and $J_{-n}(z)$ are independent.

Remark 13.1.2. For $n \notin \mathbb{N}$ the limit operation in function 13.3 is not necessary as $\sin(n\pi)$ will never become 0 in this case.

13.2 Generating function

Define the following function:

$$g(x,t) = exp\left[\frac{x}{2}\left(t - \frac{1}{t}\right)\right] \tag{13.4}$$

If we expand this function as a Laurent series, we obtain the following formula:

$$g(x,t) = \sum_{n=-\infty}^{+\infty} J_n(x)t^n \tag{13.5}$$

By applying the residue theorem 7.6.10, we can express the functions $J_n(x)$ as follows:

$$J_n(x) = \frac{1}{2\pi i} \oint_C \frac{g(x,t)}{t^{n+1}} dt \tag{13.6}$$

The function $g(x,t)$ is called the generating function of the Bessel functions.

13.3 Applications

13.3.1 Laplace equation

When solving the Laplace equation in cylindrical coordinates we obtain a BDE with integer n, which has the **cylindrical Bessel functions** 13.2 and 13.3 as solutions.

13.3.2 Helmholtz equation

When solving the Helmholtz equation in spherical coordinates we obtain a variant of the BDE for the radial part:

$$z^2 y'' + 2zy' + [z^2 - n(n+1)]y = 0 \tag{13.7}$$

where n is an integer. The solutions, called **spherical Bessel functions**, are related to the cylindrical Bessel functions in the following way:

$$j_n(r) = \sqrt{\frac{\pi}{2x}} J_{n+\frac{1}{2}}(r) \tag{13.8}$$

and similarly for the Neumann functions.

Part IV

Linear Algebra

Chapter 14

Linear Algebra

14.1 General

14.1.1 Polynomes

Definition 14.1.1 (Degree). The exponent of the highest order power in x. It is often denoted by $\deg(f)$.

Definition 14.1.2 (Monic polynome). A polynome of which the highest order term has coefficient 1.

Theorem 14.1.3 (Fundamental theorem of algebra). *Let $f(x) \in K[x]$ with $\deg(f) \geq 1$. Then $f(x)$ has at least 1 root in $\mathbb{C}$.*

Theorem 14.1.4. *If $f(x) \in \mathbb{C}[x]$ is a monic polynome with $\deg(f) \geq 1$, we can write:*

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

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

14.2 Vector spaces

In this and coming sections all vector spaces can be finite- or infinite-dimensional. If necessary, the dimension will be specified.

Definition 14.2.1 (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 8 axioms:

1. V is an Abelian group under vector addition.

2. $a(b\vec{v}) = (ab)\vec{v}$

3. $1_K\vec{v} = \vec{v}$ where 1_K is the identity element of the field K

4. Distributivity of scalar multiplication with respect to vector addition: $a(\vec{v}+\vec{w}) = a\vec{v}+a\vec{w}$

Definition 14.2.2 (Linear combination). The vector w is a linear combination of elements in the set $\{v_n\}$ if it can be written as:

$$w = \sum_n \lambda_n v_n \tag{14.1}$$

Definition 14.2.3 (Linear independence). A set $\{v_n\}$ is said to be linearly independent if the following relation holds:

$$\sum_n \lambda_n v_n = 0 \iff \forall n : \lambda_n = 0 \tag{14.2}$$

Definition 14.2.4 (Span). A set of vectors $\{v_n\}$ is said to span V if every vector $v \in V$ can be written as a linear combination of $\{v_n\}$.

Definition 14.2.5 (Basis). A set $\{v_n\}$ is said to be a basis of V if all the elements v_n are linearly independent and if $\{v_n\}$ spans V.

Remark 14.2.6. Every set T that spans V contains a basis of V.

Here it is time for a little side note. In the previous definition we implicitly used the concept of a *Hamel* basis, which is based on the condition that all linear combinations contain only finitely many terms[1]. It follows that for finite-dimensional spaces we do not have to worry. In infinitely-dimensional spaces however we have to keep this in mind.

We continue by defining this peculiar type of basis.

Definition 14.2.7 (Hamel basis). Consider the set of all linearly independent subsets of V. Under the relation of inclusion this set becomes a partially ordered set[2]. From Zorn's lemma[3] it follows that there exists at least one maximal linearly independent set. Such a set is called a Hamel basis of V.

Remark. This construction clearly assumes the ZFC axioms of 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 (and thus to Zorn's lemma).

14.2.1 Subspaces

Definition 14.2.8 (Subspace). Let V be a K-vector space. A subset W of V is a subspace if W itself is a K-vector space under the operations of V. Alternatively we can write this as:

$$W \leq V \iff \forall w_1, w_2 \in W : \forall \lambda, \mu \in K : \lambda w_1 + \mu w_2 \in W \tag{14.3}$$

---

[1]An alternative, which allows infinitely many terms, is given by the concept of a *Schauders* basis.

[2]See definition 2.2.2.

[3]See theorem 2.2.7.

Definition 14.2.9 (Grassmannian). Let V be a K-vector space. The set consisting of all k-dimensional subspaces of V is denoted by $\mathrm{Gr}(k, V)$.

Property 14.2.10 (†). The action of $GL(V)$ acts transitively[4] on all k-dimensional subspaces of V. It follows that the coset space[5] $GL(V)/H_W$ for any stabilizer H_W of some $W \in \mathrm{Gr}(k, V)$ is isomorphic (as a set) to $\mathrm{Gr}(k, V)$.

14.2.2 Algebra

Definition 14.2.11 (Algebra). Let V be a K-vector space. Let V be equipped with the binary operation $\star : V \times V \to V$. $(V, \star)$ is called an algebra over K if it satisfies the following conditions[6]:

1. Right distributivity: $(\vec{x} + \vec{y}) \star \vec{z} = \vec{x} \star \vec{z} + \vec{y} \star \vec{z}$

2. Left distributivity: $\vec{x} \star (\vec{y} + \vec{z}) = \vec{x} \star \vec{y} + \vec{x} \star \vec{z}$

3. Compatibility with scalars: $(a\vec{x}) \star (b\vec{y}) = (ab)(\vec{x} \star \vec{y})$

These conditions turn the binary operation into a bilinear operation.

Definition 14.2.12 (Unital algebra). An algebra V is said to be unital if it contains an identity element with respect to the bilinear map $\star$.

Definition 14.2.13 (Clifford algebra). Let V be a unital associative algebra over the field K. If the bilinear map is a quadratic form then V is called a Clifford algebra.

Notation 14.2.14. Let V be an algebra and Q a quadratic form. The Clifford algebra is denoted by $C\ell(V, Q)$.

14.2.3 Sum and direct sum

Definition 14.2.15 (Sum). Let V be a K-vector space. Let $W_1, W_2, ..., W_k$ be subspaces of V. The 'sum' of the subspaces $W_1, ..., W_k$ is defined as follows:

$$W_1 + ... + W_k := \left\{ \sum_{i=1}^{k} w_i : w_i \in W_i \right\} \tag{14.4}$$

Definition 14.2.16 (Direct sum). If every element v of the sum as defined in definition 14.2.15 can be written as a unique linear combination, then the sum is called a direct sum.

---

[4]See definition 2.4.38

[5]See definition 2.4.5.

[6]These conditions imply that the binary operation is a bilinear map.

Notation 14.2.17 (Direct sum).

$$W_1 \oplus \ldots \oplus W_k = \bigoplus_{i=1}^{k} W_i$$

Theorem 14.2.18. *Let V be a K-vector space. Let W, W_1, W_2 be three subspaces of V such that $W = W_1 \oplus W_2$. We have the following properties:*

- *If $\mathcal{B}_1$ is a basis of W_1 and if $\mathcal{B}_2$ is a basis of W_2, $\mathcal{B}_1 \cup \mathcal{B}_2$ is a basis of W.*
- $\dim(W) = \dim(W_1) + \dim(W_2)$

Theorem 14.2.19. *Let V be a finite-dimensional K-vector space. Let W_1, W_2 be two subspaces of V. Then the following relation holds:*

$$\dim(W_1 + W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 \cap W_2) \tag{14.5}$$

The second item in previous property is a direct consequence of this property.

Definition 14.2.20 (Complement). Let V be a K-vector space. Let W be a subspace of V. A subspace W' of V is called a complement of W if $V = W \oplus W'$.

Theorem 14.2.21. *Let V be a K-vector space. Let U, W be two subspaces of V. If $V = U + W$, then there exists a subspace $Y \leq U$ such that $V = W \oplus Y$. Furthermore every subset W of V has a complement in V.*

14.2.4 Graded vector space

Similar to definition 2.5.10 we can define the following:

Definition 14.2.22 (Graded vector space). Let V_n be a vector space for all $n \in \mathbb{N}$. The vector space

$$V = \bigoplus_{n \in \mathbb{N}} V_n \tag{14.6}$$

is called a graded vector space.

Definition 14.2.23 (Graded algebra). Let V be a graded vector space with the additional structure of an algebra given by the multiplication $\star$. Then V is a graded algebra if $\star$ maps $V^k \times V^l$ to V^{k+l}.

Example 14.2.24 (Superalgebra). Let A be a $\mathbb{Z}_2$-graded algebra, i.e.:

$$A = A_0 \oplus A_1 \tag{14.7}$$

such that for all $i, j \mod 2$:

$$A_i \star A_j \subseteq A_{i+j} \tag{14.8}$$

14.3   Linear maps[7]

Definition 14.3.1 (Zero map). Let $f : A \to B$ be a (linear) map. The map f is called a zero map if:

$$\forall a \in A : f(a) = 0 \tag{14.9}$$

Definition 14.3.2 (Restriction). Let $f : A \to B$ be a (linear) map. Let $C \subset A$. The (linear) map $f|_C : C \to B : c \to f(c)$ is called the restriction of f to C.

Definition 14.3.3 (Injective). A map $f : A \to B$ is called injective if the following condition is satisfied:

$$\forall a, a' \in A : f(a) = f(a') \implies a = a' \tag{14.10}$$

Notation 14.3.4 (Injective map).

$$f : A \hookrightarrow B$$

Definition 14.3.5 (Surjective). A map $f : A \to B$ is called surjective if the following condition is satisfied:

$$\forall b \in B, \exists a \in A : f(a) = b \tag{14.11}$$

Notation 14.3.6 (Surjective map).

$$f : A \twoheadrightarrow B$$

Definition 14.3.7 (Bijective). A map is called bijective if it is both injective and surjective.

Notation 14.3.8 (Bijective map).

$$f : A \xrightarrow{\sim} B$$

Definition 14.3.9 (Isomorphism). A linear bijective map f between two K-vector spaces is called an isomorphism.

Notation 14.3.10 (Isomorphic). If two K-vector spaces V, W are isomorphic we denote it as following:

$$V \cong W$$

Definition 14.3.11 (Automorphism). An isomorphism from V to V is called an automorphism. The set of all automorphisms on V, which is in fact a group, is denoted by $\text{Aut}(V)$.

Definition 14.3.12 (C^r-diffeomorphism). An isomorphism of class $C^r(K)$ with an inverse that is also of class $C^r(K)$ is called a C^r-diffeomorphism.

Theorem 14.3.13. *Let $f : A \to B$ be a map. The following statements are equivalent:*

1. *There exists a map[8] $g : B \to A$ such that $f \circ g = \mathbf{1}_B$ and $g \circ f = \mathbf{1}_A$.*

---

[7]Other names are 'linear mapping' and 'linear transformation'.

  2. f is bijective.

Corollary 14.3.14. From theorem 14.3.13 and the definition of isomorphisms we can conclude that isomorphisms are precisely those maps that are invertible.

Definition 14.3.15 (General linear group[9]). The set of all automorphisms $f : V \to V$ is called the general linear group $GL_K(V)$ of $GL(V)$.

Definition 14.3.16 (Rank). The dimension of the image of a linear map is called the rank.

Definition 14.3.17 (Kernel). The kernel of a linear map $f : V \to W$ is the following subset of V:

$$\ker(f) = \{v \in V \mid f(v) = 0\} \tag{14.12}$$

Definition 14.3.18 (Nullity). The dimension of the kernel is called the nullity.

Theorem 14.3.19. *A linear map $f : V \to W$ is injective if and only if $ker(f) = \{0\}$.*

Property 14.3.20. Let $f : V \to W$ be a linear map. Let $U \leq V$. We have the following two properties of the restriction $f|_U$ of f to U:

- $\ker(f|_U) = \ker(f) \cap U$

- $\operatorname{im}(f|_U) \leq \operatorname{im}(f)$

14.3.1   Linear operator

Definition 14.3.21 (Linear operator). A linear automorphism $f : V \to V$ is called a linear operator. It is also more generally known as an **endomorphism** on V.

Property 14.3.22. Let $\lambda, \mu \in K$. An operator $f : V \to V$ is called linear if it satisfies the following condition:

$$f(\lambda v_1 + \mu v_2) = \lambda f(v_1) + \mu f(v_2) \tag{14.13}$$

Theorem 14.3.23. *Let V be finite-dimensional K-vector space. Let $f : V \to V$ be a linear operator. The following statements are equivalent:*

- *f is injective*

- *f is surjective*

- *f is bijective*

---

[8]The map g is called the **inverse** of f.

[9]This group is isomorphic to the general linear group of invertable matrices, hence the similar name and notation. (See definition 14.5.7)

14.3.2   Dimension

Definition 14.3.24 (Dimension). Let V be a finite-dimensional K-vector space. Let $\{v_n\}$ be a basis for V that contains n elements. We then define the dimension of V as following:

$$\boxed{\dim(V) = n} \tag{14.14}$$

Property 14.3.25. Let V be a finite-dimensional K-vector space. Every basis of V has the same number of elements.

Theorem 14.3.26 (Dimension theorem[10]). *Let $f : V \to W$ be a linear map.*

$$\dim(\mathrm{im}(f)) + \dim(\ker(f)) = \dim(\mathrm{V}) \tag{14.15}$$

Theorem 14.3.27. *Two K-vector spaces are isomorphic if and only if they have the same dimension.*

14.3.3   Homomorphisms

Definition 14.3.28 (Homomorphism space). Let V, W be two K-vector spaces. The set of all linear maps between V and W is called the homomorphism space of V to W, or shorter: the 'hom-space' of V to W.

$$Hom_K(V, W) = \{f : V \to W \mid \text{f is a linear map}\} \tag{14.16}$$

Theorem 14.3.29. *If V, W are two finite-dimensional K-vector spaces we have:*

$$\dim\left(Hom_K(V, W)\right) = \dim(V) \cdot \dim(W) \tag{14.17}$$

Definition 14.3.30 (Endomorphism ring). The space $Hom_K(V, V)$ with the composition as multiplication forms a ring, the endomorphism ring. It is denoted as $\mathrm{End}_K(V)$ or $\mathrm{End}(V)$.

Definition 14.3.31 (Minimal polynomial). Let $f \in \mathrm{End}(V)$ and V a finite-dimensional K-vector space. The monic polynomial $\mu_f(x)$ of the lowest order such that $\mu_f(f) = 0$ is called the minimal polynomial of f.

Property 14.3.32. Let $f \in \mathrm{End}(V)$. Let $\mu_f(x)$ be the minimal polynomial of f. Let $\varphi(x) \in K[x]$. If $\varphi(f) = 0$, then the minimal polynomial $\mu_f(x)$ divides $\varphi(x)$.

---

[10]Also called the 'rank-nullity theorem'.

14.3.4   Dual space

Definition 14.3.33 (Dual space). Let V be a K-vector space. The dual space V^* of V is the following vector space:

$$V^* := \text{Hom}_K(V, K) = \{f : V \to K \; : \; f \text{ is a linear map}\} \tag{14.18}$$

Definition 14.3.34 (Linear form). The elements of V^* are called *linear forms.*

Property 14.3.35. From theorem 14.3.29 it follows that $\dim(V^*) = \dim(V)$.

Remark 14.3.36. If V is infinite-dimensional, theorem 14.3.35 is not valid. In the infinite-dimensional case we **always** have $|V^*| > |V|$ (where we now use the cardinality instead of the dimension).

Definition 14.3.37 (Dual basis). Let $\mathcal{B} = \{e_1, e_2, ..., e_n\}$ be a basis for a finite-dimensional K-vector space V. We can define a basis $\mathcal{B}^* = \{\varepsilon_1, \varepsilon_2, ..., \varepsilon_n\}$ for V^*, called the dual basis of $\mathcal{B}$, as follows:

$$\boxed{\varepsilon_i : V \to K : \sum_{j=1}^{n} a_i e_i \mapsto a_i} \tag{14.19}$$

Definition 14.3.38 (Dual map). Let $f : V \to W$ be a linear map. The linear map $f^* : W^* \to V^* : \varphi \to \varphi \circ f$ is called the dual map or **transpose** of f.

Notation 14.3.39 (Transpose). When $V = W$ the dual map f^* is often denoted by f^T.

Definition 14.3.40 (Natural pairing). The natural pairing of V and its dual V^* is defined as the following bilinear map:

$$\langle v, v^* \rangle = v^*(v) \tag{14.20}$$

14.3.5   Convex functions

Definition 14.3.41 (Convex function). Let X be a convex subset of V. A function $f : X \to \mathbb{R}$ is convex if for all $x, y \in X$ and $t \in [0, 1]$:

$$f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y) \tag{14.21}$$

Remark 14.3.42. For a concave function we have to turn the inequality around.

Corollary 14.3.43. A linear map $f : X \to \mathbb{R}$ is both convex and concave.

Theorem 14.3.44 (Karamata's theorem). *Let $I \subset \mathbb{R}$ be an interval and let $f : I \to \mathbb{R}$ be a convex function. If $(x_1, ..., x_n)$ is a tuple that majorizes $(y_1, ..., y_n)$ then:*

$$\sum_{k=1}^{n} f(x_k) \geq \sum_{k=1}^{n} f(y_k) \tag{14.22}$$

14.4 Inner product

In the following section all vector spaces V will be $\mathbb{R}$- or $\mathbb{C}$-vector spaces.

Notation 14.4.1 (Inner product). Let v, w be two vectors in V. The map $\langle \cdot | \cdot \rangle : V \times V \to K$ is called an inner product on V if it satisfies the following 3 properties:

1. **Conjugate symmetry:** $\langle v | w \rangle = \langle w | v \rangle^*$

2. **Linearity in the first argument:** $\langle \lambda u + v | w \rangle = \lambda \langle u | w \rangle + \langle v | w \rangle$

3. **Non-degeneracy:** $\langle v | v \rangle = 0 \iff v = 0$

4. **Positive-definiteness** $\langle v | v \rangle \geq 0$

Remark 14.4.2. Inner products are special cases of **non-degenerate Hermitian forms** which do not possses the positive-definiteness property.

Corollary 14.4.3. The first two properties have the result of conjugate linearity in the second argument:
$$\langle f | \lambda g + \mu h \rangle = \overline{\lambda} \langle f | g \rangle + \overline{\mu} \langle f | h \rangle \tag{14.23}$$

14.4.1 Inner product space

Definition 14.4.4 (Inner product space[11]). A vector space equipped with an inner product $\langle \cdot | \cdot \rangle$ is called an inner product space.

Definition 14.4.5 (Metric dual[12]). Using the inner product (or any other non-degenerate Hermitian form) one can define the metric dual of a vector v by the following map:
$$L : V \to V^* : v \mapsto \langle v | \cdot \rangle \tag{14.24}$$

Definition 14.4.6 (Adjoint operator). Let A be a linear operator on V. Let v, w be two vectors in V. The *Hermitian* adjoint of A is defined as the linear operator $A^\dagger$ that satisfies:
$$\langle A^\dagger v, w \rangle = \langle v, A w \rangle \tag{14.25}$$

Alternatively one can define the adjoint using the metric dual $L(\cdot)$ as follows:
$$\boxed{A^\dagger = L^{-1} \circ A^T \circ L} \tag{14.26}$$

If $A = A^\dagger$ then A is said to be **Hermitian** or **self-adjoint**.

Corollary 14.4.7. The Hermitian adjoint of a complex matrix $A \in \mathbb{C}^{m \times n}$ is given by:
$$A^\dagger = \overline{A}^T \tag{14.27}$$

where $\overline{A}$ denotes the complex conjugate of A and A^T the transpose of A.

---

[11]Sometimes called a **prehilbert space**.
[12]See also definition 26.1.

The definition of an adjoint operator 14.4.6 can be generalized to the case where $A^\dagger$ is not unique (for example when A is not globally defined) in the following way:

Definition 14.4.8 (Conjugate operators). Two operators B and C are said to be conjugate if:

$$\langle Bx, y \rangle = \langle x, Cy \rangle \tag{14.28}$$

Example 14.4.9. The Lie algebra associated with the group of isometries $\text{Isom}(V)$ of a non-degenerate Hermitian form satisfies following condition:

$$\langle Xv, w \rangle = -\langle v, Xw \rangle \tag{14.29}$$

for all Lie algebra elements X. It follows that the Lie algebra consists of all anti-hermitian operators.

14.4.2   Orthogonality

Definition 14.4.10 (Orthogonal). Let $v, w \in V$. The vectors v and w are said to be orthogonal, denoted by $v \perp w$, if they obey the following relation:

$$\langle v|w \rangle = 0 \tag{14.30}$$

An orthogonal **system** is a set of vectors, none of them the null vector, that are mutually orthogonal.

Property 14.4.11. Orthogonal systems are linearly independent.

Definition 14.4.12 (Orthonormal). A set of vectors $\{v_n\}$ is said to be orthonormal if it is orthogonal and if all the elements v_n obey the following relation:

$$\langle v|v \rangle = 1 \tag{14.31}$$

Definition 14.4.13 (Orthogonal complement[13]). Let W be a subspace of V. The following subspace is called the orthogonal complement of W:

$$W^\perp = \{v \in V \mid \forall w \in W : \langle v|w \rangle = 0\} \tag{14.32}$$

Property 14.4.14. The inner-product is invariant under transformations between orthonormal bases.

Property 14.4.15.
$$W \cap W^\perp = \{0\} \tag{14.33}$$

Property 14.4.16. Let V be a finite-dimensional K-vector space. The orthogonal complement $W^\perp$ is a complementary subspace[14] to W, i.e. $W \leq V$: $W \oplus W^\perp = V$.

---

[13]$W^\perp$ is pronunciated as 'W-perp'.

Corollary 14.4.17. Let $W \leq V$ where V is a finite-dimensional K-vector space. We have the following relation:

$$(W^\perp)^\perp = W \tag{14.34}$$

Definition 14.4.18 (Orthogonal projection). Let V be a finite-dimensional K-vector space. Let $W \leq V$. Let $w \in W$ and let $\{w_1, ..., w_k\}$ be an orthonormal basis of W. We define the projection of $v \in V$ on W and $w \in W$ as follows:

$$\text{proj}_W(v) = \sum_{i=1}^{k} \langle v|w_i \rangle w_i \tag{14.35}$$

$$\text{proj}_w(v) = \frac{\langle v|w \rangle}{\langle w|w \rangle} w \tag{14.36}$$

Property 14.4.19.

1. $\forall w \in W : \text{proj}_W(w) = w$

2. $\forall u \in W^\perp : \text{proj}_W(u) = 0$

Method 14.4.20 (Gram-Schmidt orthonormalisation). Let $\{u_n\}$ be a set of linearly independent vectors. We can construct an orthonormal set $\{e_n\}$ out of $\{u_n\}$ in the following way:

$$
\begin{aligned}
w_1 &= u_1 & e_1 &= \frac{w_1}{||w_1||} \\
w_2 &= u_2 - \frac{\langle u_2|w_1 \rangle}{||u_2||^2} w_1 & e_2 &= \frac{w_2}{||w_2||} \\
&\vdots & &\vdots \\
w_n &= u_n - \sum_{k=1}^{n-1} \frac{\langle u_n|w_k \rangle}{||u_n||^2} w_k & e_n &= \frac{w_n}{||w_n||}
\end{aligned}
\tag{14.37}
$$

14.4.3 Angle

Definition 14.4.21 (Angle). Let v, w be elements of an inner product space. The angle θ between v and w is defined as:

$$\boxed{\cos\theta = \frac{\langle v|w \rangle}{||v|| ||w||}} \tag{14.38}$$

14.5 Matrices

Notation 14.5.1. The set of all $m \times n$-matrices defined over the field K is denoted as $M_{m,n}(K)$ or $\text{Mat}_{m,n}(K)$. If $m = n$, the set is denoted as $M_n(K)$ or $Mat_n(K)$.

---

[14]hence the name

Property 14.5.2 (Dimension). The dimension of $M_{m,n}(K)$ is mn.

Definition 14.5.3 (Trace). Let $A = (a_{ij}) \in M_n(K)$. We define the trace of A as follows:

$$\boxed{\operatorname{tr}(A) = \sum_{i=1}^{n} a_{ii}} \tag{14.39}$$

Property 14.5.4. Let $A, B \in M_n(K)$. We have the following properties of the trace:

1. $\operatorname{tr} : M_n(K) \to K$ is a linear map

2. $\operatorname{tr}(AB) = \operatorname{tr}(BA)$

3. $\operatorname{tr}(AB) \neq \operatorname{tr}(A)\operatorname{tr}(B)$

4. $\operatorname{tr}(A^T) = \operatorname{tr}(A)$

Formula 14.5.5 (Hilbert-Schmidt norm). Also called the Frobenius norm. This matrix norm is given by following formula

$$||A||_{HS}^2 = \sum_{i,j} |A_{ij}|^2 = \operatorname{tr}(A^\dagger A) \tag{14.40}$$

If one identifies $M_n(\mathbb{C})$ with $\mathbb{C}^{2n}$ then this norm equals the standard Hermitian norm.

Formula 14.5.6 (Hadamard product). Let A, B be two matrices in $M_{m \times n}(K)$. The Hadamard product is defined as the entry-wise product:

$$(A \circ B)_{ij} = A_{ij} B_{ij} \tag{14.41}$$

Definition 14.5.7 (General linear group). The set of invertable matrices is called the general linear group and is denoted by $GL_n(K)$.

Property 14.5.8. For all $A \in GL_n(K)$ we have:

- $A^T \in GL_n(K)$

- $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$

Property 14.5.9. Let $A \in M_{m,n}(K)$. Denote the set of columns of A as $\{A_1, A_2, ..., A_n\}$ and the set of rows of A as $\{R_1, R_2, ..., R_m\}$. The set of columns is a subspace of K^m and the set of rows is a subspace of K^n. Furthermore we have:

$$\dim(\operatorname{span}(A_1, ..., A_n)) = \dim(\operatorname{span}(R_1, ..., R_m))$$

Definition 14.5.10 (Rank of a matrix). We can define the rank of matrix $A \in M_{m,n}(K)$ as follows:

$$\operatorname{rk}(A) := \dim(\operatorname{span}(A_1, ..., A_n)) \overset{14.5.9}{=} \dim(\operatorname{span}(R_1, ..., R_m)) \tag{14.42}$$

Property 14.5.11. The rank of a matrix has the folowing properties:

1. Let $A \in M_{m,n}(K)$ and $B \in M_{n,r}(K)$. We have $\text{rk}(AB) \leq \text{rk}(A)$ and $\text{rk}(AB) \leq \text{rk}(A)$.

2. Let $A \in GL_n(K)$ and $B \in M_{n,r}(K)$. We have $\text{rk}(AB) = \text{rk}(B)$.

3. Let $A \in GL_n(K)$ and $B \in M_{r,n}(K)$. We have $\text{rk}(BA) = \text{rk}(B)$.

Property 14.5.12. Let $A \in M_{m,n}(K)$. First define the following linear map:

$$\boxed{L_A : K^n \to K^m : v \mapsto Av}$$

(14.43)

This map has the following properties:

1. $\text{im}(L_A) = \text{span}(A_1, ..., A_n)$

2. $\dim(\text{im}(L_A)) = \text{rk}(A)$

Remark. The second property is a direct consequence of the first one and definition 14.42.

14.5.1 System of equations

Theorem 14.5.13. *Let $AX = w$ with $A \in M_{m,n}(K)$, $w \in K^m$ and $X \in K^n$ be a system of m equations in n variables. Let L_A be the linear map as defined in equation 14.43. We then have the following properties:*

1. *The system is false if and only if $w \notin \text{im}(L_A)$.*

2. *If the system is not false, the solution set is an affine space. If $v_0 \in K^n$ is a solution, then the solution set is given by: $L_A^{-1}(w) = v_0 + \text{ker}(L_A)$.*

3. *If the system is homogeneous ($AX = 0$), then the solution set is equal to $\text{ker}(L_A)$.*

Theorem 14.5.14 (Uniqueness). *Let $AX = w$ with $A \in M_n(K)$ be a system of n equations in n variables. If $rk(A) = n$, then the system has a unique solution.*

14.5.2 Coordinates and matrix representations

Definition 14.5.15 (Coordinate vector). Let $\mathcal{B} = \{b_1, ..., b_n\}$ be a basis of V. Let $v \in V$ such that $v = \sum_{i=1}^{n} \lambda_i b_i$. We define the coordinate vector of v with respect to $\mathcal{B}$ as $(\lambda_1, ..., \lambda_n)^t$. The λ_i's are called the **coordinates** of v with respect to $\mathcal{B}$.

Definition 14.5.16 (Coordinate isomorphism). With the previous definition in mind we can define the coordinate isomorphism of v with respect to $\mathcal{B}$ as follows:

$$\boxed{\beta : V \to K^n : \sum_{i=1}^{n} \lambda_i b_i \mapsto (\lambda_1, ..., \lambda_n)^t}$$

(14.44)

Definition 14.5.17 (Matrix representation). Let V be an n-dimensional K-vector space and W an m-dimensional K-vector space. Let $f : V \to W$ be a linear map. Let $\mathcal{B} = \{b_1, ..., b_n\}, \mathcal{C} = \{c_1, ..., c_m\}$ be a basis for V, respectively W. The matrix representation of f with respect to $\mathcal{B}$ and $\mathcal{C}$ can be derived as follows: For every $j \in \{1, ..., n\}$ we can write $f(b_j) = \sum_{i=1}^{m} a_{ij} c_i$, so with this in mind we can define the matrix $(a_{ij}) \in M_{m,n}(K)$ as the matrix represenation of f.

Notation 14.5.18 (Matrix representation of a linear map). The matrix representation of f with respect to $\mathcal{B}$ and $\mathcal{C}$ is denoted by $A_{f,\mathcal{B},\mathcal{C}}$.

Method 14.5.19 (Construction of a matrix representation). From definition 14.5.17 we can see that j-th column of $A_{f,\mathcal{B},\mathcal{C}}$ coincides with the coordinate vector of $f(b_j)$ with respect to $\mathcal{C}$. We use this relation to construct $A_{f,\mathcal{B},\mathcal{C}}$ by writing for every $j \in \{1, ..., n\}$ the coordinate vector of $f(b_j)$ in the j-th column.

Theorem 14.5.20. Let $(\lambda_1, ..., \lambda_n)^t$ be the coordinate vector of $v \in V$ with respect to $\mathcal{B}$. Let $(\mu_1, ..., \mu_m)^t$ be the coordinate vector of $f(v)$ with respect to $\mathcal{C}$. Then the following relation holds:

$$\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix} = A_{f,\mathcal{B},\mathcal{C}} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \tag{14.45}$$

Theorem 14.5.21. For every matrix $A \in M_{m,n}(K)$ there exists a linear map $f : V \to W$ such that $A_{f,\mathcal{B},\mathcal{C}} = A$.

On the other hand we also have the following theorem:

Theorem 14.5.22. Let $f : K^n \to K^m$ be a linear map. There exists a matrix $A \in M_{m,n}(K)$ such that $f = L_A$.

Theorem 14.5.23. Let β and γ be the coordinate isomorphisms with respect to respectively $\mathcal{B}$ and $\mathcal{C}$. From theorem 14.5.20 it follows that:

$$\gamma(f(v)) = A_f \cdot \beta(v) \tag{14.46}$$

or alternatively

$$\gamma \circ f = L_{A_f} \circ \beta \tag{14.47}$$

Theorem 14.5.24. The map $Hom_K(V, W) \to M_{m,n}(K) : f \mapsto A_f$ is an isomorphism and for every $f \in Hom_K(V, W)$ and $g \in Hom_K(W, U)$ we have:

$$A_{g \circ f} = A_g A_f \tag{14.48}$$

Theorem 14.5.25. The map $End_K(V) \to M_n(K) : f \mapsto A_{f,\mathcal{B},\mathcal{B}}$ is an isomorphism and for every $f, g \in End_K(V)$ we have:

$$A_{g \circ f} = A_g A_f \tag{14.49}$$

Theorem 14.5.26. *Let $f \in \text{End}_K(V)$. 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, we have that*

$$(A_f)^{-1} = A_{f^{-1}}$$

In other words, the following map is an isomorphism[15]:

$$GL_K(V) \to GL_n(K) : f \mapsto A_f$$

Remark 14.5.27. The sets $GL_K(V)$ and $GL_N(K)$ are groups. So the previous theorem states that the map $f \mapsto A_f$ is a group isomorphism.

Theorem 14.5.28. *Let $V = K^n$. Let $f \in V^*$. From construction 14.5.19 it follows that $A_f = (f(e_1), ..., f(e_n)) \in M_{1,n}(K)$ with respect to the standard basis of V. This combined with theorem 14.5.20 gives:*

$$f(\lambda_1, ..., \lambda_n)^t = (f(e_1), ..., f(e_n))(\lambda_1, ..., \lambda_n)^t = \sum_{i=1}^{n} f(e_i)\lambda_i \qquad (14.50)$$

or alternatively with $\{\varepsilon_1, ..., \varepsilon_n\}$ the dual basis to the standard basis of V:

$$\boxed{f = \sum_{i=1}^{n} f(e_i)\varepsilon_i} \qquad (14.51)$$

Theorem 14.5.29. *Let $f : V \to W$ be a linear map. 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 $\mathcal{C}$, then 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 $\mathcal{B}$.*

14.5.3 Coordinate transforms

Definition 14.5.30 (Transition matrix). Let $\mathcal{B} = \{b_1, ..., b_n\}$ and $\mathcal{B}' = \{b'_1, ..., b'_n\}$ be two bases of V. Every element of $\mathcal{B}'$ can be written as a linear combination of elements in $\mathcal{B}$:

$$b'_j = q_{1j}b_1 + ... + q_{nj}b_n$$

The matrix $Q = (q_{ij}) \in M_n(K)$ is the transition matrix from the 'old' basis $\mathcal{B}$ to the 'new' basis $\mathcal{B}'$.

Theorem 14.5.31. *Let $\mathcal{B}, \mathcal{B}'$ be two basis of V. Let Q be the transition matrix from $\mathcal{B}$ to $\mathcal{B}'$. We find the following statements:*

---

[15]Follows from theorem 14.5.25.

1. *Let $\mathcal{C}$ be an arbitrary basis of V with γ the corresponding coordinate isomorphism. Define the following matrices:*

$$B = (\gamma(b_1), ..., \gamma(b_n)) \quad and \quad B' = (\gamma(b'_1), ..., \gamma(b'_n))$$

Then $BQ = B'$.

2. *$Q \in GL_n(K)$ and Q^{-1} is the transition matrix from $\mathcal{B}'$ to $\mathcal{B}$.*

3. *Let $v \in V$ with $(\lambda_1, ..., \lambda_n)^t$ the coordinate vector with respect to $\mathcal{B}$ and $(\lambda'_1, ..., \lambda'_n)^t$ the coordinate vector with respect to $\mathcal{B}'$. Then:*

$$Q\begin{pmatrix} \lambda'_1 \\ \vdots \\ \lambda'_n \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \quad and \quad \begin{pmatrix} \lambda'_1 \\ \vdots \\ \lambda'_n \end{pmatrix} = Q^{-1}\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

Theorem 14.5.32. *Let V, W be two finite-dimensional K-vector spaces. Let $\mathcal{B}, \mathcal{B}'$ be two bases of V and $\mathcal{C}, \mathcal{C}'$ two bases of W. Let Q, P be the transition matrices from $\mathcal{B}$ to $\mathcal{B}'$ and from $\mathcal{C}$ to $\mathcal{C}'$ respectively. Let $A = A_{f,\mathcal{B},\mathcal{C}}$ and $A' = A_{f,\mathcal{B}',\mathcal{C}'}$. Then:*

$$A' = P^{-1}AQ$$

Corollary 14.5.33. Let $f \in \text{End}_K(V)$ and let Q be the transition matrix. From theorem 14.5.32 it follows that:

$$A' = Q^{-1}AQ$$

Definition 14.5.34 (Matrix conjugation). Let $A \in M_n(K)$. The set

$$\{Q^{-1}AQ \mid Q \in GL_n(K)\}$$

is called the conjugacy class[16] of A. Another name for conjugation is **similarity transformation**.

Remark 14.5.35. If A is a matrix representation of a linear operator f, then the conjugacy class of A consists out of every possible matrix representation of f.

Property 14.5.36. From property 14.5.4 it follows that the trace of a matrix is invariant under similarity transformations:

$$\boxed{\text{tr}(Q^{-1}AQ) = \text{tr}(A)} \tag{14.52}$$

Definition 14.5.37 (Matrix congruence). Let $A, B \in M_n(K)$. If there exists a matrix P such that

$$A = P^T BP \tag{14.53}$$

then the matrices are said to be congruent.

Property 14.5.38. Every matrix congruent to a symmetric matrix is also symmetric.

Theorem 14.5.39. *Let $(V, \langle .|. \rangle)$ be an inner-product space defined over $\mathbb{R}$ (or $\mathbb{C}$). Let $\mathcal{B}, \mathcal{B}'$ be two orthonormal bases of V and let Q be the transition matrix. We can find the following result:*

$$Q^T Q = \mathbb{1}_n$$

[16]This is the general definition of conjugacy classes for groups. Furthermore, these classes induce a partitioning of the group.

14.5.4 Determinant

Definition 14.5.40 (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 14.5.41 (Cofactor). The cofactor α_{ij} of the matrix element a_{ij} is equal to:

$$(-1)^{i+j}\det(A_{ij})$$

where $\det(A_{ij})$ is the minor as previously defined.

Definition 14.5.42 (Adjugate matrix). The adjugate matrix of $A \in M_n(K)$ is defined as follows:

$$\mathrm{adj}(A) := \begin{pmatrix} \alpha_{11} & \alpha_{21} & \cdots & \alpha_{n1} \\ \alpha_{12} & \alpha_{22} & \cdots & \alpha_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{1n} & \alpha_{2n} & \cdots & \alpha_{nn} \end{pmatrix} \tag{14.54}$$

or shorter: $\mathrm{adj}(A) = (\alpha_{ij})^t$.

Remark. It is important to notice that we have to transpose the matrix after the elements have been replaced by their cofactor.

Property 14.5.43. Let $A, B \in M_n(K)$. Denote the columns of A as $A_1, \ldots, A_n$. We have the following properties of the determinant:

1. $\det(A^t) = \det(A)$

2. $\det(AB) = \det(BA) = \det(A)\det(B)$

3. $\det(A_1, \ldots, A_i + \lambda A_i', \ldots, A_n) = \det(A_1, \ldots, A_i, \ldots, A_n) + \lambda \det(A_1, \ldots, A_i', \ldots, A_n)$ for all $A_i, A_i' \in M_{n,1}(K)$.

4. If two columns of A are equal then $\det(A) = 0$.

5. $\det(A_{i_1}, \ldots, A_{i_n}) = \mathrm{sgn}(i_1, \ldots, i_n)\det(A_1, \ldots, A_n)$

6. $\det(A_1, \ldots, A_n) = \det(A_1, \ldots, A_i + \lambda A_k, \ldots, A_n)$ where A_i, A_k are columns of A.

7. The determinant can be evaluated as follows:

$$\det(A) = \sum_{i=1}^{n}(-1)^{i+k}a_{ik}\det(A_{ik}) \tag{14.55}$$

Theorem 14.5.44. *Let $A \in M_n(K)$, the following statements are equivalent:*

1. $\det(A) \neq 0$

   2. $rk(A) = n$

   3. $A \in GL_n(K)$

Theorem 14.5.45. *For all $A \in M_n(K)$ we find $A\,adj(A) = adj(A)A = \det(A)I_n$.*

Theorem 14.5.46. *For all $A \in GL_n(K)$ we find $A^{-1} = \det(A)^{-1}\ adj(A)$.*

An alternative definition of a $k \times k$-minor is:

Definition 14.5.47. Let $A \in M_{m,n}(K)$ and $k \leq \min(m, n)$. 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.

Theorem 14.5.48. *Let $A \in M_{m,n}(K)$ and $k \leq \min(m, n)$. We find that $rk(A) \geq k$ if and only if A contains a $k \times k$-minor different from 0.*

Theorem 14.5.49. *Let $f \in End_K(V)$. The determinant of the matrix representation of f is invariant under basis transformations.*

Definition 14.5.50 (Determinant of a linear operator). The previous theorem allows us to unambiguously define the determinant of f as follows:

$$\det(f) := \det(A)$$

where A is some matrix representation of f.

14.5.5   Characteristic polynomial

Definition 14.5.51 (Characteristic polynomial[17]). Let V be a finite-dimensional K-vector space. Let $f \in End_K(V)$ be a linear operator with the matrix representation A (with respect to some arbitrary basis). We then find:

$$\chi_f(x) := \det(x\mathbb{1}_n - A) \in K[x] \tag{14.56}$$

is a monic polynomial of degree n in the variable x and the polynomial does not depend on the choice of basis.

Definition 14.5.52 (Characteristic equation[18]). The following equation is called the characteristic equation of f:

$$\boxed{\chi_f(x) = 0} \tag{14.57}$$

---

[17]This polynomial can also be used directly for a matrix A as theorem 14.5.21 matches every matrix A with some linear operator f.

[18]This equation is sometimes called the **secular equation**.

Formula 14.5.53. Let $A = (a_{ij}) \in M_n(K)$ with characteristic polynomial:

$$\chi_A(x) = x^n + c_{n-1}x^{n-1} + \ldots + c_1 x + c_0$$

We then have the following result:

$$\begin{cases} c_0 = (-1)^n \det(A) \\ c_{n-1} = -\text{tr}(A) \end{cases}$$

Theorem 14.5.54 (Cayley-Hamilton).

1. Let $A \in M_n(K)$ with characteristic polynomial $\chi_A(x)$. We find the following relation:

$$\chi_A(A) = A^n + \sum_{i=1}^{n-1} c_i A^i = 0 \tag{14.58}$$

2. Let $f \in End_K(V)$ with characteristic polynomial $\chi_f(x)$. We find that

$$\chi_f(f) = f^n + \sum_{i=1}^{n-1} c_i f^i = 0 \tag{14.59}$$

Corollary 14.5.55. From theorem 14.3.32 and the Cayley-Hamilton theorem it follows that the minimal polynomial $\mu_f(x)$ is a divisor of the characteristic polynomial $\chi_f(x)$.

14.5.6 Linear groups

Definition 14.5.56 (Elementary matrix). An elementary matrix is a matrix of the following form:

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & c_{ij} & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \vdots & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & c_{ij} & \ddots & \vdots \\ 0 & 0 & \vdots & 1 \end{pmatrix}, \ldots$$

i.e. equal to the sum of an identity matrix and a multiple of a matrix unit U_{ij}, $i \neq j$.

Notation 14.5.57 (Elementary matrix). $E_{ij}(c)$ is the elementary matrix with element c on the i, j-th position.

Property 14.5.58. We have the following property:

$$\det(E_{ij}(c)) = 1 \tag{14.60}$$

which implies that $E_{ij}(c) \in GL_n(K)$.

Property 14.5.59. We find the following results concerning the multiplication by an elementary matrix:

1. 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.

2. 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.

Theorem 14.5.60. *Every matrix $A \in GL_n(K)$ can be written in the following way:*

$$A = SD$$

where S is a product of elementary matrices and $D = diag(1, \dots, 1, \det(A))$.

Definition 14.5.61 (Special linear group). The following subset of $GL_n(K)$ is called the special linear group:
$$SL_n(K) = \{A \in GL_n(K) \mid \det(A) = 1\} \tag{14.61}$$

Theorem 14.5.62. *Every $A \in SL_n(K)$ can be written as a product of elementary matrices.*[19]

Definition 14.5.63 (Orthogonal group). The orthogonal and special orthogonal group are defined as follows:

$$O_n(K) = \{A \in GL_n(K) \mid AA^t = A^t A = I_n\}$$
$$SO_n(K) = O_n(K) \cap SL_n(K)$$

Property 14.5.64. For orthogonal matrices, conjugacy 14.5.34 and congruency 14.53 are equivalent.

Definition 14.5.65 (Unitary group). The unitary and special unitary group are defined as follows:

$$U_n(K, \sigma) = \{A \in GL_n(K) \mid A\overline{A}^t = \overline{A}^t A = I_n\}$$
$$SU_n(K, \sigma) = U_n(K) \cap SL_n(K)$$

where σ denotes the *involution*[20] $a^\sigma \equiv \bar{a}$.

Remark. If $K = \mathbb{C}$ where the involution is taken to be the complex conjugate, the σ is often ommited in the definition: $U_n(K)$ and $SU_n(K)$.

Definition 14.5.66 (Unitary equivalence). Let A, B be two matrices in $M_n(K)$. If there is a unitary matrix U such that
$$A = U^\dagger B U$$

then the matrices A and B are said to be **unitarily equivalent**.

---

[19]Follows readily from theorem 14.5.60.

[20]An involution is an operator that is its own inverse: $f(f(x)) = x$.

14.6 Eigenvectors

Definition 14.6.1 (Eigenvector). A vector $v \in V \setminus \{0\}$ is called an **eigenvector** of the linear operator $f : V \to V$ if it satisfies the following equation:

$$f(v) = \lambda v \tag{14.62}$$

Where $\lambda \in K$ is the **eigenvalue** belonging to v.

Definition 14.6.2 (Eigenspace). The subspace of V consisting of the zero vector and the eigenvectors of an operator is called the **eigenspace** associated with that operator:

$$\ker(\lambda \mathbf{1}_V - f) \tag{14.63}$$

Theorem 14.6.3 (Characteristic equation[21]). *Let $f \in End_K(V)$ be a linear operator. A scalar $\lambda \in K$ is an eigenvalue of f if and only if it satisfies the characteristic equation 14.57.*

Theorem 14.6.4. *A linear operator $f \in End_K(V)$ defined over an n-dimensional K-vector space V has at most n different eigenvalues.[22]*

Method 14.6.5 (Finding the eigenvectors of a matrix A).

1. First we find the eigenvalues λ_i of $\mathbf{A}$ by applying theorem 14.6.3.

2. Then we find the eigenvector v_i belonging to the eigenvalue λ_i by using the following equation:
$$(\mathbf{A} - \lambda_i \mathbf{1}_V) v_i = 0 \tag{14.64}$$

14.6.1 Diagonalization

Definition 14.6.6 (Diagonalizable operator). An operator $f \in End_K(V)$ on a finite-dimensional K-vector space V is diagonalizable if there exists a matrix representation $A \in M_n(K)$ of f such that A is a diagonal matrix.

Theorem 14.6.7. *A linear operator f defined on a finite-dimensional K-vector space V has a diagonal matrix as matrix representation if and only if the set of eigenvectors of f is a basis of V.*

Theorem 14.6.8. *A matrix $A \in M_n(K)$ is diagonalizable if and only if there exists a matrix $P \in GL_n(K)$ such that $P^{-1}AP$ is diagonal.*

[21]This theorem also holds for the eigenvalues of a matrix $A \in M_n(K)$.
[22]This theorem also holds for a matrix $A \in M_n(K)$.

Corollary 14.6.9. Using the fact that the trace of a linear operator is invariant under similarity transformations (see property 14.52) we get following useful formula:

$$\boxed{\operatorname{tr}(f) = \sum_i \lambda_i} \tag{14.65}$$

where $\{\lambda_i\}_{0 \leq i \leq n}$ are the eigenvalues of f.

Property 14.6.10. Let V be an n-dimensional K-vector space. Let $f \in \operatorname{End}_K(V)$ be a linear operator. We find the following properties of the eigenvectors/eigenvalues of f:

1. The eigenvectors of f belonging to different eigenvalues are linearly independent.

2. If f has exactly n eigenvalues, f is diagonalizable.

3. If f is diagonalizable, V is the direct sum of the eigenspaces of f belonging to the different eigenvalues of f.

Definition 14.6.11 (Multiplicity). Let V be a K-vector space. Let $f \in \operatorname{End}_K(V)$ be a linear operator with characteristic polynomial[23]:

$$\chi_f(x) = \prod_{i=1}^n (x - \lambda_i)^{n_i}$$

We can define the following multiplicities:

1. The *algebraic multiplicity* of an eigenvalue λ_i is equal to n_i.

2. The *geometric multiplicity* of an eigenvalue λ_i is equal to the dimension of the eigenspace belonging to that eigenvalue.

Remark 14.6.12. The geometric multiplicity is always at least 1.

Property 14.6.13. The algebraic multiplicity is always greater than or equal to the geometric multiplicity.

Theorem 14.6.14. *Let $f \in End_K(V)$ be a linear operator. f is diagonalizable if and only if for every eigenvalue the algebraic multiplicity is equal to the geometric multiplicity.*

Property 14.6.15. Every Hermitian operator $f \in \operatorname{End}_K(\mathbb{C}^n)$ has the following properties:

1. All the eigenvalues of f are real.

2. Eigenvectors belonging to different eigenvalues are orthogonal.

3. f is diagonalizable and there always exists an orthonormal basis of eigenvectors of f.

---

[23]We assume that the characteristic polynomial can be written in this form. This depends on the possibility to completely factorize the polynomial in K (i.e. it has 'enough' roots in K). If not, f cannot even be diagonalized. However, there always exists a field F containing K, called a **splitting field**, where the polynomial has 'enough' roots.

Property 14.6.16. Let $A, B \in \mathrm{End}_K(V)$ be two linear operators. If the commutator $[A, B] = 0$, then the two operators have a common eigenbasis.

Theorem 14.6.17 (Sylvester's law of inertia). *Let S be a symmetric matrix. The number of positive and negative eigenvalues is invariant with respect to similarity transformations*[24].

14.7 Euclidean space $\mathbb{R}^n$

A finite-dimensional $\mathbb{R}$-vector space is called a **Euclidean space**.

14.7.1 Angle

Definition 14.7.1 (Angle). Let $(V, \langle .|. \rangle)$ be a real inner-product space. For every $u, v \in V \setminus \{0\}$ we can define the angle between them as[25]:

$$\angle(u, v) = \mathrm{acos} \frac{\langle u|v \rangle}{||u|| \cdot ||v||} \tag{14.66}$$

where we set the range of acos as $[0, \pi]$.

Notation 14.7.2. When working in a Euclidean space the inner product $\langle v|w \rangle$ is often written as $v \cdot w$ or even vw.

14.7.2 Vector product

Definition 14.7.3 (Orientation). Let $\mathcal{B}, \mathcal{B}'$ be two ordered bases of $\mathbb{R}^n$. Let Q be the transition matrix from $\mathcal{B}$ to $\mathcal{B}'$. If $\det(Q) > 0$ then the bases are said to have the same orientation (or be *consistently oriented*). If $\det(Q) < 0$ then the bases are said to have an opposite orientation.

Corollary 14.7.4 (Positive orientation). The previous definition imposes an equivalence relation on the set of bases of $\mathbb{R}^n$. The set of bases consists out of two equivalence classes. Take one class and call the bases in it *positively* or *directly* oriented. The bases in the other class are then said to be *negatively* or *indirectly* oriented.

Remark 14.7.5. It is convenient to take the standard basis $(e_1, \ldots, e_n)$ to be positively oriented.

---

[24]Also with respect to conjugation, which are equivalent to similarity transformations according to property 14.5.64.

[25]This formula follows readily from the Cauchy-Schwarz inequality (see theorem 16.2.6).

Formula 14.7.6 (Cross product).

$$\boxed{(v \times w)_i = \varepsilon_{ijk} v_j w_k} \tag{14.67}$$

where ε_{ijk} is the 3-dimensional Levi-Civita symbol.

Remark 14.7.7. It is important to note that the previous construction is only valid in 3 dimenensions.

Chapter 15

Vector calculus

15.1 Nabla-operator

Definition 15.1.1 (Nabla).

$$\nabla \equiv \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) \tag{15.1}$$

Following 'definitions' can be found by using basic properties of (vector) calculus.

Formula 15.1.2 (Gradient).

$$\nabla V = \left(\frac{\partial V_x}{\partial x}, \frac{\partial V_y}{\partial y}, \frac{\partial V_z}{\partial z} \right) \tag{15.2}$$

Formula 15.1.3. Let φ be a scalar field. The total differential $d\varphi$ can be rewritten as

$$d\varphi = \nabla \varphi \cdot d\vec{r} \tag{15.3}$$

Property 15.1.4. The gradient of a scalar function V is perpendicular to the level surfaces 2.29 of V.

Definition 15.1.5 (Directional derivative). Let $\vec{a}$ be a unit vector. The directional derivative $\nabla_a V$ is defined as the change of the function V in the direction of $\vec{a}$:

$$\nabla_a V \equiv (\vec{a} \cdot \nabla) V \tag{15.4}$$

Example 15.1.6. Let φ be a scalar field. Let $\vec{t}$ denote the tangent vector to a curve $\vec{r}(s)$ with s natural parameter. The variation of the scalar field φ along $\vec{r}(s)$ is given by

$$\frac{\partial \varphi}{\partial s} = \frac{d\vec{r}}{ds} \cdot \nabla \varphi \tag{15.5}$$

Definition 15.1.7 (Conservative vector field). Vector field obtained as the gradient of a scalar function.

Property 15.1.8. A vector field is conservative if and only if its line integral is path independent.

Formula 15.1.9 (Gradient of tensor). Let T be a tensor field with coordinates ξ_i. Let $\vec{g}^i(\xi^1, \xi^2, \xi^3)$ be a curvilinear orthogonal frame[1]. The gradient of T is defined as follows:

$$\nabla T = \frac{\partial T}{\partial \xi^i} \otimes \vec{g}^i \tag{15.6}$$

Formula 15.1.10 (Divergence).

$$\nabla \cdot \vec{A} = \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \tag{15.7}$$

Definition 15.1.11 (Solenoidal vector field). Also known as a **divergence free vector field**. It satisifies the equation

$$\nabla \cdot \vec{V} = 0$$

Formula 15.1.12 (Rotor / curl).

$$\nabla \times \vec{A} = \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}, \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x}, \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \right) \tag{15.8}$$

Definition 15.1.13 (Irrotational vector field).

$$\nabla \times \vec{V} = 0$$

Remark 15.1.14. Conservative vector fields are also irrotational but irrotational vector fields are only conservative if the domain is simply connected.

15.1.1   Laplacian

$$\triangle V \equiv \nabla^2 V = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} \tag{15.9}$$

$$\nabla^2 \vec{A} = \nabla \left(\nabla \cdot \vec{A} \right) - \nabla \times \left(\nabla \times \vec{A} \right) \tag{15.10}$$

Remark 15.1.15. Equation 15.10 is called the 'vector laplacian'.

Formula 15.1.16 (Laplacian in different coordinate systems).

- Cylindrical coordinates (ρ, ϕ, z):

$$\frac{1}{\rho} \frac{\partial}{\partial \rho} \left(\rho \frac{\partial}{\partial \rho} \right) + \frac{1}{\rho^2} \frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2} \tag{15.11}$$

- Spherical coordinates (r, ϕ, θ):

$$\frac{1}{r^2} \left[\frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} + \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) \right] \tag{15.12}$$

---

[1]See definition 24.3.19.

15.1.2   Mixed properties

$$\nabla \times (\nabla V) = 0 \tag{15.13}$$

$$\nabla \cdot \left(\nabla \times \vec{\boldsymbol{V}}\right) = 0 \tag{15.14}$$

In Cartesian coordinates equation 15.10 can be rewritten as follows:

$$\nabla^2 \vec{\boldsymbol{A}} = (\triangle A_x, \triangle A_y, \triangle A_z) \tag{15.15}$$

15.1.3   Helmholtz decomposition

Formula 15.1.17 (Helmholtz decomposition). Let $\vec{\boldsymbol{P}}$ be a vector field that decays rapidly (more than $1/r$) when $r \to \infty$. $\vec{\boldsymbol{P}}$ can be written as follows:

$$\vec{\boldsymbol{P}} = \nabla \times \vec{\boldsymbol{A}} + \nabla V \tag{15.16}$$

15.2   Line integrals

Formula 15.2.1 (Line integral of a continuous scalar field). Let f be a continuous scalar field. Let Γ be a piecewise smooth curve with the parametrization $\vec{\boldsymbol{\varphi}}(t), t \in [a, b]$. We define the line integral of f over Γ as follows:

$$\int_\Gamma f(s)ds = \int_a^b f(\vec{\boldsymbol{\varphi}}(t))||\vec{\boldsymbol{\varphi}}'(t)||dt \tag{15.17}$$

Formula 15.2.2 (Line integral of a continuous vector field). Let $\vec{\boldsymbol{F}}$ be a continuous vector field. Let Γ be a piecewise smooth curve with the parametrization $\vec{\boldsymbol{\varphi}}(t), t \in [a, b]$. We define the line integral of F over Γ as follows:

$$\int_\Gamma \vec{\boldsymbol{F}}(\vec{\boldsymbol{s}}) \cdot d\vec{\boldsymbol{s}} = \int_a^b \vec{\boldsymbol{F}}(\vec{\boldsymbol{\varphi}}(t)) \cdot \vec{\boldsymbol{\varphi}}'(t)dt \tag{15.18}$$

15.3   Integral theorems

Theorem 15.3.1 (Fundamental theorem of calculus for line integrals).
Let $\vec{\Gamma} : \mathbb{R} \to \mathbb{R}^3$ be a smooth curve.

$$\int_{\Gamma(a)}^{\Gamma(b)} \nabla f(\vec{\boldsymbol{r}}) \cdot d\vec{\boldsymbol{r}} = \varphi(\Gamma(b)) - \varphi(\Gamma(a)) \tag{15.19}$$

Theorem 15.3.2 (Kelvin-Stokes' theorem).

$$\oint_{\partial S} \vec{A} \cdot d\vec{l} = \iint_S \left(\nabla \times \vec{A} \right) dS \tag{15.20}$$

Theorem 15.3.3 (Divergence theorem). *Also known as Gauss' theorem or the Gauss-Ostrogradsky theorem.*

$$\oiint_{\partial V} \vec{A} \cdot d\vec{S} = \iiint_V \left(\nabla \cdot \vec{A} \right) dV \tag{15.21}$$

Corollary 15.3.4 (Green's identity).

$$\oiint_{\partial V} \left(\psi \nabla \phi - \phi \nabla \psi \right) \cdot d\vec{S} = \iiint_V \left(\psi \nabla^2 \phi - \phi \nabla^2 \psi \right) dV \tag{15.22}$$

15.4   Curvilinear coordinates

In this section the differential operators are generalized to curvilinear coordinates. To do this we need the scale factors as formally defined in equation 20.14. Also there is no Einstein summation used, all summations are written explicitly.

Formula 15.4.1 (Unit vectors).

$$\frac{\partial \vec{r}}{\partial q^i} = h_i \hat{e}_i \tag{15.23}$$

Formula 15.4.2 (Gradient).

$$\nabla V = \sum_{i=1}^3 \frac{1}{h_i} \frac{\partial V}{\partial q^i} \hat{e}_i \tag{15.24}$$

Formula 15.4.3 (Divergence).

$$\nabla \cdot \vec{A} = \frac{1}{h_1 h_2 h_3} \left(\frac{\partial}{\partial q^1} (A_1 h_2 h_3) + \frac{\partial}{\partial q^2} (A_2 h_3 h_1) + \frac{\partial}{\partial q^3} (A_3 h_1 h_2) \right) \tag{15.25}$$

Formula 15.4.4 (Rotor).

$$(\nabla \times \vec{A})_i = \frac{1}{h_j h_k} \left(\frac{\partial}{\partial q^j} (A_k h_k) - \frac{\partial}{\partial q^k} (A_j h_j) \right) \tag{15.26}$$

where $i \neq j \neq k$.

Chapter 16

Banach spaces and Hilbert spaces

In this chapter the term "linear operator", which is normally reserved for maps of the form $f : V \to V$, is used instead of "linear map". This was done to keep the vocabulary in track with that of the standard literature on Banach and operator spaces.

For a revision of inner product spaces see section 14.4.

16.1 Banach spaces

Definition 16.1.1 (Norm). Let V be a K-vector space. A function $||\vec{v}|| : V \to [0, +\infty[$ is called a norm if it satisfies following conditions:

- **Non-degeneracy:** $||\vec{v}|| = 0 \iff \vec{v} = 0$

- **Homogeneity:** $||a\vec{v}|| = |a|||\vec{v}||$ for all scalars $a \in K$

- **Triangle equality (subadditivity):** $||\vec{v} + \vec{w}|| \leq ||\vec{v}|| + ||\vec{w}||$

Remark 16.1.2. A norm $|| \cdot ||$ clearly induces a metric[1] by setting $d(x, y) = ||x - y||$.

Definition 16.1.3 (Normed vector space). A K-vector space equipped with a norm $||\cdot||$.

Definition 16.1.4 (Banach space). A normed vector space that is complete[2] with respect to the norm $|| \cdot ||$.

Definition 16.1.5 (Reflexive space). A Banach space V for which its dual coincides with the dual of its dual, i.e. $V^* = (V^*)^*$.

Property 16.1.6. Every finite-dimensional Banach spaces is reflexive. This follows from property 14.3.35.

---

[1]See definition 3.1.1.
[2]See condition 4.12.

Property 16.1.7. Let (x_n) be a Cauchy sequence in a normed space V. Then $(||x_n||)$ is a convergent sequence in $\mathbb{R}$. This implies that every Cauchy sequence in a normed space is bounded.

Property 16.1.8. The topological (continuous) dual of a Banach space is also a Banach space.

16.1.1 Bounded operators

Definition 16.1.9 (Bounded operator). Let $L : V \to W$ be a linear operator between two Banach spaces. The operator is said to be bounded if there exists a scalar M that satisfies the following condition:

$$\boxed{\forall v \in V : ||Lv||_W \leq M||v||_V} \tag{16.1}$$

Notation 16.1.10. The space of bounded linear operators from V to W is denoted by $\mathcal{B}(V, W)$.

Property 16.1.11. If V is a Banach space then $\mathcal{B}(V)$ is also a Banach space.

Definition 16.1.12 (Operator norm). The operator norm of L is defined as follows:

$$||L||_{op} = \inf\{M | M \text{ satisfies condition 16.1}\} \tag{16.2}$$

As the name suggests it is a norm on $\mathcal{B}(V, W)$. The topology induced by this norm is called the norm topology.

Equivalent definitions of the operator norm are:

$$||L||_{op} = \sup_{||x|| \leq 1} ||L(x)|| = \sup_{||x|| = 1} ||L(x)|| = \sup_{x \neq 0} \frac{||L(x)||}{||x||} \tag{16.3}$$

Following property reduces the problem of continuity to that of boundedness.

Property 16.1.13. Let $f \in \mathcal{L}(V, W)$. 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 16.1.14. Let A be a bounded linear operator with eigenvalue λ. We then have:

$$|\lambda| \leq ||A||_{op} \tag{16.4}$$

Property 16.1.15. Let A be a bounded linear operator. Let $A^\dagger$ denote its adjoint[3]. Then $A^\dagger$ is bounded and $||A||_{op} = ||A^\dagger||_{op}$.

---

[3]See definition 14.4.6.

16.1.2   Theorems

Property 16.1.16. Let X be a general TVR. Every linear map $\varphi : \mathbb{K}^n \to X$ is continuous.

Property 16.1.17. Let X be a finite-dimensional normed vector space. Every linear bijection $\varphi : \mathbb{K}^n \to X$ is a homeomorphism.

Corollary 16.1.18. 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 16.1.19 (Open mapping theorem[4]). *Let $f : V \to W$ be a continuous linear operator between two Banach spaces. If f is surjective then it also open.*

16.1.3   Spectrum

Definition 16.1.20 (Resolvent set). Let A be a bounded linear operator on a normed space V. The resolvent set $\rho(A)$ consists of all scalar $\lambda \in \mathbb{C}$ such that $(A - \lambda \mathbb{1})^{-1}$ is a bounded linear operator, called the resolvent of A, on a dense subset of V. These scalars λ are called **regular values** of A.

Definition 16.1.21 (Spectrum). The set of scalars $\mu \notin \rho(A)$ is called the spectrum of A.

Remark 16.1.22. It is obvious from the definition of an eigenvalue that every eigenvalue of A belongs to the spectrum of A. The converse however is not true.

Definition 16.1.23 (Point spectrum). The set of scalars $\mu \in \mathbb{C}$ for which the resolvent of A fails to be injective is called the point spectrum of A. This set contains exactly the eigenvalues of A.

Definition 16.1.24 (Continuous spectrum). The set of scalars $\mu \in \mathbb{C}$ for which the resolvent of A fails to be surjective but for which the range of the resolvent is dense in V is called the continuous spectrum of A. The scalars for which the range is not dense is called the **residual spectrum** $\sigma_r(A)$.

Definition 16.1.25 (Compression spectrum). The set of scalars $\mu \in \mathbb{C}$ for which the resolvent of A fails to have a dense range in V is called the compression spectrum $\sigma(A)$. It follows that $\sigma_r(A) \subseteq \sigma(A)$.

16.2   Hilbert space

Definition 16.2.1 (Hilbert space). A vector space that is both a Banach space and an inner product space (where the norm is induced by the inner product).

---

[4]Sometimes called the *Banach-Schauder* theorem.

Example 16.2.2. Let $f, g \in \mathcal{L}^2([a, b], \mathbb{C})$, the inner product of f and g is defined as:

$$\boxed{\langle f|g \rangle = \int_a^b f^*(x)\overline{g(x)}dx} \tag{16.5}$$

Remark 16.2.3. See section 8.4.2 for a more formal treatment of this subject.

Formula 16.2.4. It is also possible to define an inner product with respect to a weight function $\phi(x)$:

$$\int_a^b f^*(x)g(x)\phi(x)dx \tag{16.6}$$

Using this formula it is possible to define orthogonality with respect to a weight function.

16.2.1   Inner products and norms

Formula 16.2.5. Let V be an inner product space. A norm on V can be induced by the inner product in the following way:

$$||v||^2 = \langle v|v \rangle \tag{16.7}$$

However not every norm induces an inner product. Only norms that satisfy the parallellogram law 16.9 induce an inner product. This inner product can be recovered through the polarization identity 16.10 (see below).

Property 16.2.6 (Cauchy-Schwarz inequality).

$$\boxed{|\langle v|w \rangle| \leq ||v|| \, ||w||} \tag{16.8}$$

where the equality holds if and only if v and w are linearly dependent.

Corollary 16.2.7. 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 also a normed space.

Formula 16.2.8 (Parallellogram law).

$$||v + w||^2 + ||v - w||^2 = 2(||v||^2 + ||w||^2) \tag{16.9}$$

Formula 16.2.9 (Polarization identity).

$$4\langle v|w \rangle = ||v + w||^2 - ||v - w||^2 + i\left(||v + iw||^2 - ||v - iw||^2\right) \tag{16.10}$$

Formula 16.2.10 (Pythagorean theorem). In an inner product space the triangle equality reduces to the well-known Pythagorean theorem for orthogonal vectors v, w:

$$||v + w||^2 = ||v||^2 + ||w||^2 \tag{16.11}$$

This formula can be extended to any set of orthogonal vectors $x_1, ..., x_n$:

$$\boxed{\left|\left|\sum_{i=1}^n x_i\right|\right|^2 = \sum_{i=1}^n ||x_i||^2} \tag{16.12}$$

16.2.2 Generalized Fourier series

Property 16.2.11 (Bessel's inequality). First of all we have following general equality for orthonormal vectors $x_1, ..., x_n$ and complex scalars $a_1, ..., a_n$:

$$\left\| x - \sum_{i=1}^{n} a_i x_i \right\|^2 = ||x||^2 - \sum_{i=1}^{n} |\langle x, x_i \rangle|^2 + \sum_{i=1}^{n} |\langle x, x_i \rangle - a_i|^2 \qquad (16.13)$$

This expression becomes minimal for $a_i = \langle x, x_i \rangle$ (last term becomes 0). This leads to Bessel's inequality:

$$\boxed{\sum_{i=1}^{n} |\langle x, x_i \rangle|^2 \leq ||x||^2} \qquad (16.14)$$

Corollary 16.2.12. The sum in 16.14 is bounded for all n, so the series $\sum_{i=1}^{+\infty}$ converges for all x. This implies that the sequences $(\langle x, x_n \rangle)$ belongs to the space l^2 of square-summable sequences.

This result does however not imply that the generalized Fourier series $\sum_{i=1}^{+\infty} \langle x, x_i \rangle x_i$ converges to x. The following theorem gives a necessary and sufficient condition for the convergence.

Theorem 16.2.13. *Let $\mathcal{H}$ be a Hilbert space. Let (x_n) be an orthonormal sequence in $\mathcal{H}$ and let (a_n) be a sequence in $\mathbb{C}$. The expansion $\sum_{i=1}^{+\infty} a_i x_i$ converges in $\mathcal{H}$ if and only if $(a_n) \in l^2$. Furthermore the expansion satisfies following equality:*

$$\left\| \sum_{i=1}^{+\infty} a_i x_i \right\|^2 = \sum_{i=1}^{+\infty} |a_i|^2 \qquad (16.15)$$

As we noted the sequence $(\langle x, x_n \rangle)$ belongs to l^2 so the generalized Fourier series converges of $x \in \mathcal{H}$ converges in $\mathcal{H}$.

Remark 16.2.14. Although the convergence of the generalized Fourier series of $x \in \mathcal{H}$ can be established using previous theorem, it does not follow that the expansion converges to x itself. We can merely say that the Fourier expansion is the best approximation of x with respect to the norm on $\mathcal{H}$.

16.2.3 Complete sets

Definition 16.2.15 (Complete set). Let $\{e_i\}_{i \in I}$ be a set (possibly a sequence) of orthonormal vectors in an inner product space V. This set is said to be complete if every vector $x \in V$ can be expressed as follows:

$$x = \sum_{i \in I} \langle x, x_i \rangle x_i \qquad (16.16)$$

This implies that a complete set is a basis for the vector space.

Another characterization is the following.

Alternative Definition 16.2.16. A complete set of orthonormal vectors is a set $S \subset V$ such that we cannot add another vector w to it satisfying:

$$\forall v_i \in S : \langle v_i, w \rangle = 0 \qquad \wedge \qquad w \neq 0 \tag{16.17}$$

Property 16.2.17. For complete sequences (x_n) the inequality of Bessel 16.14 becomes an equality. Furthermore, the generalized Fourier series with respect to the complete sequence is unique.

Using previous property we can prove the following theorem due to Parceval.

Theorem 16.2.18 (Parceval). *Let (x_n) be a complete sequence in a Hilbert space $\mathcal{H}$. Every vector $x \in \mathcal{H}$ has a unique Fourier series representation $\sum_{i=1}^{+\infty} a_i x_i$ where the Fourier coefficients (a_i) belong to l^2 and the inequality of Bessel is an equality.*

Conversely if the inequality of Bessel becomes an equality for every $x \in \mathcal{H}$ then the sequence (x_n) is complete.

Property 16.2.19. A sequence (x_n) in a Hilbert space $\mathcal{H}$ is complete if and only if $\langle x, x_i \rangle = 0$ for all x_i implies that $x = 0$.

16.2.4   Orthogonality and projections

The basic notions on orthogonality in inner product space can be found in section 14.4.2.

Property 16.2.20. 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 16.2.21. The previous property implies that the orthogonal complemement of some arbitrary subset of a Hilbert space is a Hilbert space itself.

Theorem 16.2.22 (Projection theorem). *Let H be a Hilbert space and $K \leq H$ a complete subspace. For every $h \in H$ there exists a unique $h' \in K$ such that $h - h'$ is orthogonal to every $k \in K$, i.e $h - h' \in K^\perp$.*

Remark 16.2.23. An equivalent definition for the unique $h' \in K$ is $||h - h'|| = \inf\{||h - k|| : k \in K\}$.

Corollary 16.2.24. It follows that given a complete (or closed) subspace S the Hilbert space $\mathcal{H}$ can be decomposed as $\mathcal{H} = S \oplus S^\perp$.

16.2.5   Separable Hilbert spaces

The definition of separable spaces in the sense of point-set topology is given in 4.5.17. An equivalent definition for Hilbert spaces is the following.

Alternative Definition 16.2.25 (Separable Hilbert space). A Hilbert space is separable if it contains a complete sequence of orthonormal vectors.

Corollary 16.2.26. Using the Gram-Schmidt method it follows from previous definition that every finite-dimensional Hilbert space is separable.

The following theorem shows that (up to an isomorphism) there are only 2 distinct types of separable Hilbert spaces.

Theorem 16.2.27. *Let $\mathcal{H}$ be separable. If $\mathcal{H}$ is finite-dimensional with dimension n then it is isometrically isomorphic to $\mathbb{C}^n$. If $\mathcal{H}$ is infinite-dimensional then it is isometrically isomorphic to l^2.*

Property 16.2.28. Every orthogonal subset of a separable Hilbert space is countable.

16.2.6   Compact operators

Definition 16.2.29 (Compact operator). Let A be a linear operator on a Hilbert space $\mathcal{H}$. A is said to be compact if for every sequence (x_n) in $\mathcal{H}$ the sequence $(A[x_n])$ has a convergent subsequence.

Property 16.2.30. Every compact operator on a Hilbert space is bounded.

Corollary 16.2.31. Every linear operator on a finite-dimensional Hilbert space is bounded.

16.2.7   Linear functionals

Property 16.2.32. Let f be a continuous linear functional. Then $\dim(\ker f)^\perp$ is 0 or 1 where the former case only arises when $f \equiv 0$.

Theorem 16.2.33 (Riesz' representation theorem). *Let $\mathcal{H}$ be a Hilbert space. For every continuous linear functional $\rho : \mathcal{H} \to \mathbb{R}$ there exists a unique element $x_0 \in \mathcal{H}$ such that*

$$\rho(h) = \langle h, x_0 \rangle \tag{16.18}$$

for all $h \in \mathcal{H}$. This implies that $\mathcal{H}$ and $\mathcal{H}^*$ are isometrically isomorphic. Furthermore the operator norm of ρ is equal to the norm of x_0.

Corollary 16.2.34. This theorem justifies the bra-ket notation used in quantum mechanics where one associates to every ket $|\psi\rangle \in \mathcal{H}$ a bra $\langle\psi| \in \mathcal{H}^*$.

Chapter 17

Operator algebras

17.1   Involutive algebras

Definition 17.1.1 (Involution). Let $*$ be an automorphism of an algebra A. If $*(*\,a) = a$ for all $a \in A$ then $*$ is called an involution of A.

Definition 17.1.2 (Involutive algebra[1]). An involutive algebra is an associative algebra A over a commutative ring R with involution $^-$ together with an operator $* : A \rightarrow A$ such that:

- $(a + b)^* = a^* + b^*$

- $(ab)^* = b^* a^*$

- $(\lambda a)^* = \overline{\lambda} a^*$

where $\lambda \in R$.

17.2   C*-algebras

Definition 17.2.1 (C*-algebra). A C*-algebra is a involutive Banach algebra[2] A such that the **C*-identity**

$$||a^*a|| = ||a||\,||a^*|| \tag{17.1}$$

is satisfied.

Definition 17.2.2 (Positive). An element of a C*-algebra is called positive if it is self-adjoint and if its spectrum is contained in $[0, +\infty[$. A linear functional on a C*-algebra is called positive if every positive element is mapped to a positive number.

---

[1]Also called a *-**algebra**.
[2]See definition 16.1.4.

Definition 17.2.3 (State). Let A be a C*-algebra. A state ψ on A is a positive linear functional of unit norm.

Chapter 18

Tensor calculus

18.1 General definition

Definition 18.1.1 (Einstein summation convention). This notation is a useful tool to reduce long formulas. Summations over indices are assumed implicitly if they occur once as a subscript and once as a superscript.

$$\sum_i p_i x^i = p_i x^i \tag{18.1}$$

In the following sections we will use this notation without further notice.

Definition 18.1.2 (Contravariant). A tensor component that transforms by the following rule is called contravariant:

$$v^i = v'^j \frac{\partial x^i}{\partial x'^j} \tag{18.2}$$

Remark 18.1.3. As can be seen in the equation, a superscript in the denominator can be regarded as a normal subscript.

Definition 18.1.4 (Covariant). A tensor component that transforms by the following rule is called covariant:

$$p_i = p'_j \frac{\partial x'^j}{\partial x^i} \tag{18.3}$$

Example 18.1.5 (Mixed tensor). As an example of a mixed tensor we give the transformation formula for the mixed third-order tensor T^k_{ij}:

$$T^k_{ij} = T'^w_{uv} \frac{\partial x'^u}{\partial x^i} \frac{\partial x'^v}{\partial x^j} \frac{\partial x^k}{\partial x'^w}$$

18.1.1   Second-order tensors

Definition 18.1.6 (Tensor of rank 2). A contravariant tensor T^{ij} transforms as:

$$T^{ij} = \frac{\partial x^i}{\partial x'^k} \frac{\partial x^j}{\partial x'^l} T'^{kl} \tag{18.4}$$

A covariant tensor T_{ij} transforms as:

$$T_{ij} = \frac{\partial x'^k}{\partial x^i} \frac{\partial x'^l}{\partial x^j} T'_{kl} \tag{18.5}$$

Remark 18.1.7. These definitions can easily be generalized to higher dimensions

Formula 18.1.8 (Product of second-order tensor and vector). Let $\vec{v}$ be a vector and let $\mathbf{T}$ be a second-order tensor.

$$\mathbf{T}(\vec{v}) = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} T_{ij} v_j \right) \vec{e}_i \tag{18.6}$$

18.1.2   Transformation

Property 18.1.9 (Inverse transformation). From equation 18.2 we know that a contravariant tensor transforms as follows

$$(A')^j = \frac{\partial (x')^j}{\partial x^i} A^i$$

The inverse transformation is given by[1]:

$$A^i = \frac{\partial x^i}{\partial (x')^j} (A')^j \tag{18.7}$$

Remark. It is important to note that we do not just invert the derivatives as that is only valid in Cartesian coordinates.

Theorem 18.1.10 (Quotient rule). *Assume we have an equation such as* $K_i A^{jk} = B_i{}^{jk}$ *or* $K_i{}^j A_{jl}{}^k = B_{il}{}^k$ *with A and B two known tensors[2]. The quotient rule asserts the following: "If the equation of interest holds under all transformations, then K is a tensor of the indicated rank and covariant/contravariant character".*

Remark. This rule is a useful substitute for the "illegal" division of tensors.

---

[1]This can be checked by substitution of the second formula in the first one.
[2]This rule does not necessarily hold when $B = 0$ as transformations rules are not defined for the null-tensor.

18.2   Tensor product

18.2.1   Tensor product

There are two possible ways to introduce the components of a tensor (on finite dimensional spaces). One way is to interpret tensors as multilinears maps another way is to interpret the components as expansion coefficients with respect to the tensor space basis.

Definition 18.2.1 (Multilinear map). One way to define the tensor components is as follows: Let $\mathbf{T}$ be a tensor that takes r vectors and s covectors as input and returns a scalar. The different components are given by $\mathbf{T}(e_i, ..., e_j, e^k, ..., e^l) = T_{i...j}{}^{k...l}$.

Definition 18.2.2. The tensor product of vector spaces V and W is defined as[3] the set of multilinear maps on the Cartesian product $V^* \times W^*$. Let v, w be vectors in respectively V and W. Let g, h be vectors in the corresponding dual spaces. The tensor product of v and w is then defined as:

$$\boxed{(v \otimes w)(g, h) = v(g)w(h)} \tag{18.8}$$

Property 18.2.3 (Universal property). A set X together with a bilinear map $\mathcal{T} : V \times W \to X$ is said to have the universal property if for every bilinear map $f : V \times W \to Z$, where Z is some other vector space, there exists a unique linear map $f' : X \to Z$ such that $f = f' \circ \mathcal{T}$.

Corollary 18.2.4. The tensor product is unique up to a linear isomorphism. This results in

$$V \otimes W \cong W \otimes V \tag{18.9}$$

Corollary 18.2.5. Let $v \in V$ and $f \in V^*$.

$$v(f) \equiv f(v) \tag{18.10}$$

Notation 18.2.6 (Tensor power).

$$V^{\otimes n} = \underbrace{V \otimes ... \otimes V}_{n \text{ copies}} \tag{18.11}$$

Remark 18.2.7. More generally, the tensor product of r copies of V and s copies of V^* is the vector space $\mathcal{T}_s^r(V) = V^{\otimes r} \otimes V^{*\otimes s}$. These tensors are said to be of **type** (r, s).

Definition 18.2.8 (Tensor product basis). The tensor space $\mathcal{T}_s^r(V)$ is spanned by the basis

$$\underbrace{e_i \otimes ... \otimes e_j}_{r \text{ basis vector}} \otimes \underbrace{e^k \otimes ... \otimes e^l}_{s \text{ dual basis vectors}}$$

where the operation $\otimes$ satisfies following properties:

---

[3]"isomorphic to" would be a better terminology. See the "universal property" 18.2.3. For a complete proof and explanation, see [10].

1. Associativity: $u \otimes (v \otimes w) = u \otimes v \otimes w$

2. Multilinearity: $a(v \otimes w) = (av) \otimes w = v \otimes (aw)$ and $v \otimes (u + w) = v \otimes u + v \otimes w$

The expansion coefficients in this basis are written as $T^{i...j}{}_{k...l}$

Property 18.2.9 (Dimension of tensor product). From the previous construction it follows that the dimension of $\mathcal{T}^r_s(V)$ is equal to rs.

Theorem 18.2.10. *We now have to proof that the values of the tensor operating on r vectors and s covectors are equal to the corresponding expansion coefficients.*

Proof. Let $\mathbf{T} = T_{i...j}{}^{k...l} e^i \otimes ... \otimes e^j \otimes e_k \otimes ... \otimes e_l$. Applying 18.2.2 and using the definition of the dual vectors we have:

$$
\begin{aligned}
\mathbf{T}(e_a, ..., e_b, e^m, ..., e^n) &= T_{i...j}{}^{k...l} e^i(e_a)...e^j(e_b) e_k(e^m)...e_l(e^n) \\
&= T_{i...j}{}^{k...l} \delta^i_a...\delta^j_b \delta^m_k...\delta^n_l \\
&= T_{a...b}{}^{m...n}
\end{aligned}
$$

This is exactly the same result as the one we get by applying the first definition. $\square$

Example 18.2.11. The scalars (elements of the base field) are in fact $(0,0)$ tensors. From the transformation rule for tensors it follows that scalars are invariant under any kind of transformation, which was allready a well known fact. Vectors and covectors are respectively $(1,0)$ and $(0,1)$ tensors.

Example 18.2.12. The linear operators on a vector space V are isomorphic to elements of the tensor space $V^* \otimes V$ or by property 18.9 to $\mathcal{T}^r_s(V)$:

$$
\begin{aligned}
\mathbf{T}(v) &= T_i{}^j e_j e^i (v^k e_k) \\
&= v^i T_i{}^j e_j
\end{aligned}
$$

The last line is exactly the componentwise transformation of vectors.

18.2.2 General operations

Definition 18.2.13 (Contraction). Let A be a tensor of type (n, m). Setting a sub- and superscript equal and summing over this index gives a new tensor of type $(n-1, m-1)$. This operation is called the contraction of A. It is given by the evaluation map

$$
V \otimes V^* : e_i \otimes e^j \mapsto e^j(e_i) \tag{18.12}
$$

Definition 18.2.14 (Direct product). Let A and B be two random tensors (both rank and co-/contravariancy). The tensor constructed by the componentwise multiplication of A and B is called the direct product of A and B.

Example 18.2.15. Let $A^i{}_k$ and $B^j{}_{lm}$ be two tensors. The direct product is equal to:

$$
C^i{}_k{}^j{}_{lm} = A^i{}_k B^j{}_{lm}
$$

Formula 18.2.16 (Operator product). It is also possible to combine operators working on different vector spaces so to make them work on the tensor product space. To do this we use following definition:

$$(\hat{A} \otimes \hat{B})(v \otimes w) = (\hat{A}v) \otimes (\hat{B}w) \tag{18.13}$$

Remark. Consider an operator $\hat{A}$ working on a space $\mathcal{H}_1$. When working with a combined space $\mathcal{H}_1 \otimes \mathcal{H}_2$ the corresponding operator is in fact $\hat{A} \otimes \mathbb{1}$ but it is often still denoted by $\hat{A}$ in physics.

18.2.3 Differentiation

Property 18.2.17.

$$\vec{\nabla} \cdot (\vec{A} \otimes \vec{B}) = (\vec{\nabla} \cdot \vec{A})\vec{B} + (\vec{A} \cdot \vec{\nabla})\vec{B} \tag{18.14}$$

18.2.4 Levi-Civita tensor

Definition 18.2.18 (Levi-Civita tensor). Let e^i be the dual vector to e_i. In n dimensions, we define the Levi-Civita tensor as follows:

$$\boldsymbol{\varepsilon} = \varepsilon_{12...n}e^1 \otimes e^2 \otimes ... \otimes e^n \tag{18.15}$$

where

$$\varepsilon_{i...n} = \begin{cases} 1 & \text{if } (i...n) \text{ is an even permutation of } (12...n) \\ -1 & \text{if } (i...n) \text{ is an odd permutation of } (12...n) \\ 0 & \text{if any of the indices occurs more than once} \end{cases}$$

Remark 18.2.19. The Levi-Civita symbol is not a tensor, but a pseudotensor. This means that the sign changes under reflections (or any transformation with determinant -1).

Formula 18.2.20 (Cross product). By using the Levi-Civita symbol, we can define the i-th component of the cross product[4] of two vectors $\vec{v}, \vec{w}$ as follows:

$$(\vec{v} \times \vec{w})_i = \varepsilon_{ijk}v_j w_k \tag{18.16}$$

18.3 (Anti)symmetric tensors

18.3.1 Symmetric tensors

Notation 18.3.1. The space of symmetric $(0, n)$ tensors is denoted by $S^n(V^*)$. The space of symmetric $(n, 0)$ tensors is denoted by $S^n(V)$.

---

[4]Following from remark 18.2.19 we can see that the cross product is in fact not a vector, but a pseudovector.

18.3.2 Antisymmetric tensors

Definition 18.3.2 (Antisymmetric tensor). Tensors that change sign under the interchange of any two indices.

Notation 18.3.3. The space of antisymmetric $(0, n)$ tensors is denoted by $\Lambda^n(V^*)$. The space of antisymmetric $(n, 0)$ tensors is denoted by $\Lambda^n(V)$.

Remark. Elements of $\Lambda^2(V)$ are also known as **bivectors**. Elements of $\Lambda^k(V)$ are generally known as k-**blades**.

Property 18.3.4. Let $n = \dim(V)$. $\Lambda^r(V)$ equals the null-space for all $r \geq n$.

18.3.3 Wedge product

Definition 18.3.5 (Wedge product).

$$\boxed{f \wedge g = f \otimes g - g \otimes f} \tag{18.17}$$

From this definition it immediately follows that the wedge product is antisymmetric.

Formula 18.3.6. Let $\{P_i\}_i$ be the set of all permutations of the sequence $(1, ..., k)$.

$$e_1 \wedge ... \wedge e_k = \sum_i \text{sgn}(P_i) e_{P_i(1)} \otimes ... \otimes e_{P_i(k)} \tag{18.18}$$

Construction 18.3.7. Let $\{e_i\}_{1 \leq i \leq n}$ be a basis for V. It is clear from the definition 18.17 that a basis for $\Lambda^r(V)$ is given by

$$\{e_{i_1} \wedge ... \wedge e_{i_r} \mid \forall k : 1 \leq i_k \leq \dim(V)\}$$

The dimension of this space is given by:

$$\dim \Lambda^k(V) = \binom{n}{k} \tag{18.19}$$

Remark 18.3.8. For $k = 0$, the above construction is not useful, so we just define $\Lambda^0(V) = \mathbb{R}$.

Formula 18.3.9 (Levi-Civita symbol). The Levi-Civita tensor in n dimensions as introduced in 18.15 can now be rewritten more concisely as:

$$\boldsymbol{\varepsilon} = e_1 \wedge ... \wedge e_n \tag{18.20}$$

Formula 18.3.10. In 3 dimensions there exists an important isomorphism $J : \Lambda^2(\mathbb{R}^3) \to \mathbb{R}^3$:

$$J(\lambda)^i = \frac{1}{2} \varepsilon^i_{\ jk} \lambda^{jk} \tag{18.21}$$

where $\lambda \in \Lambda^2(\mathbb{R}^3)$.

Looking at the definition of the cross product 14.67, we can see that $\vec{v} \times \vec{w}$ is actually the same as $J(\vec{v} \wedge \vec{w})$. One can thus use the wedge product to generalize the cross product to higher dimensions.

Example 18.3.11. Let A, B and C be three vectors in V. Now consider following expression:

$$(C \wedge B)(L(A), \cdot)$$

where $L(A)$ is the metric dual of A (see 14.24). Evaluating this formula using the properties of the wedge and tensor products leads to the well known BAC-CAB rule of triple cross products:

$$(C \cdot A)B - (B \cdot A)C$$

Remark 18.3.12. The wedge product can also be defined without prior knowledge of tensor products. The wedge product is then defined by letting $V \wedge W$ be the smallest space such that for all elements $v \in V, w \in W$ the following property holds:

$$v \wedge w = -w \wedge v \tag{18.22}$$

Now, let $\{e_i\}_{i \leq m}$ be an ordered basis for V and let $\{d_j\}_{j \leq n}$ be an ordered basis for W. The wedge product $V \wedge W$ is spanned by the basis $\{e_i \wedge d_j\}_{i \leq m, j \leq n}$.

18.3.4   Exterior algebra

Definition 18.3.13 (Exterior power). In the theory of exterior algebras, the space $\Lambda^k(V)$ is often called the kth exterior power of V.

Definition 18.3.14 (Exterior algebra). We can define a graded vector space[5] $\Lambda^*(V)$ as follows:

$$\Lambda^*(V) = \bigoplus_{k \geq 0} \Lambda^k(V)$$

Then we can turn this graded vector space into a graded algebra by taking the wedge product as the multiplication:

$$\wedge : \Lambda^k(V) \times \Lambda^l(V) \to \Lambda^{k+l}(V)$$

This algebra is called the exterior algebra or **Grassmann algebra** of V.

Alternative Definition 18.3.15 (†). Let $T(V)$ be the (free) tensor algebra over the vector space V, i.e.

$$T(V) = \bigoplus_{k \geq 0} T^{\otimes k}(V) \tag{18.23}$$

where $T^{\otimes k}(V)$ is the kth tensor power of V. The exterior algebra over V is generally defined as the quotient of $T(V)$ by the two-sided ideal I generated by $\{v \otimes v | v \in V\}$.

Property 18.3.16. The exterior algebra is both an associative algebra and a unital algebra with unit element $1 \in \mathbb{R}$. Furthermore it is also commutative in the graded sense (see 2.25).

---

[5]See definition 14.6.

18.3.5 Hodge star

It follows from equation 18.19 that the spaces $\Lambda^k(V)$ and $\Lambda^{n-k}(V)$ have the same dimension, so there exists an isomorphism between them. This map is given by the Hodge star $*$. However this map can only be defined independent of the choice of (ordered) basis if we restrict ourselves to vector spaces equipped with a non-degenerate Hermitian form 14.4.2.

Definition 18.3.17 (Volume element). Let V be an n-dimensional vector space with ordered basis $\{e_i\}_{i \leq n}$. The volume element on V is defined as:

$$\text{Vol}(V) := e_1 \wedge ... \wedge e_n \tag{18.24}$$

It is clear that this is an element of $\Lambda^n(V)$.

Definition 18.3.18 (Orientation). Let $\omega \in \Lambda(V)$ be an element of degree n. From the previous definition it follows that this k-blade is a scalar multiple of Vol because $\Lambda^n(V)$ is one-dimensional:

$$\omega = r\text{Vol}(V)$$

The k-blade ω induces an orientation on V in the following way. If the scalar $r > 0$ then the orientation is said to be **positive**. If $r < 0$ then the orientation is **negative**.

Formula 18.3.19 (Inner product). Let V be equipped with an inner product $\langle \cdot, \cdot \rangle$. Then we can define an inner product on $\Lambda^k(V)$ by:

$$\boxed{\langle v_1 \wedge ... \wedge v_k | w_1 \wedge ... \wedge w_k \rangle_k = \det(\langle v_i, w_j \rangle)} \tag{18.25}$$

For an orthogonal basis, this formula factorises into:

$$\langle v_1 \wedge ... \wedge v_k | w_1 \wedge ... \wedge w_k \rangle_k = \langle v_1 | w_1 \rangle \cdots \langle v_k | w_k \rangle \tag{18.26}$$

Definition 18.3.20 (Hodge star). The Hodge star $* : \Lambda^k(V) \to \Lambda^{n-k}(V)$ is defined as the isomorphism such that for all $\omega \in \Lambda^k(V)$ and $\rho \in \Lambda^{n-k}(V)$ we have the following equality:

$$\omega \wedge \rho = \langle *\omega, \rho \rangle_{n-k} \text{Vol}(V) \tag{18.27}$$

where $\langle \cdot, \cdot \rangle$ is the inner product 18.25 on $\Lambda^{n-k}(V)$. Furthermore, this isomorphism is unique.

Proof. Because $\omega \wedge \rho$ is an element of $\Lambda^n(V)$ it is a scalar multiple of $\text{Vol}(V)$. This implies that it can be written as

$$c(\rho)\text{Vol}(V)$$

The map $c : \Lambda^{n-k}(V) \to \mathbb{R} : \rho \mapsto c(\rho)$ is a linear map and thus a continuous map, so we can apply Riesz' representation theorem to identify c with a unique element $*\omega \in \Lambda^{n-k}(V)$ such that

$$c(\rho) = \langle *\omega, \rho \rangle_{n-k}$$

$\square$

Formula 18.3.21. Let $\{e_i\}_{i \leq n}$ be a positively oriented ordered orthonormal basis for V. An explicit formula for the Hodge star is given by the following construction. Let $\{i_1, ..., i_k\}$ and $\{j_1, ..., j_{n-k}\}$ be two complementary index sets with increasing subindices. Let $\omega = e_{i_1} \wedge ... \wedge e_{i_k}$.

$$\ast\omega = \operatorname{sgn}(\tau) \prod_{m=1}^{n-k} \langle e_{j_m} | e_{j_m} \rangle e_{j_1} \wedge ... \wedge e_{j_{n-k}} \tag{18.28}$$

where τ is the permutation that maps $e_{i_1} \wedge ... \wedge e_{i_k} \wedge e_{j_1} \wedge ... \wedge e_{j_{n-k}}$ to $\operatorname{Vol}(V)$

Corollary 18.3.22. Consider three vectors $u, v, w \in \mathbb{R}^3$.

$$\ast(v \wedge w) = v \times w \tag{18.29}$$
$$\ast(v \times w) = v \wedge w \tag{18.30}$$
$$\ast(u \wedge v \wedge w) = u \cdot (v \times w) \tag{18.31}$$

Remark 18.3.23. Formula 18.21 is an explicit evaluation of the first equation 18.29.

Proof. The sign $\operatorname{sgn}(\tau)$ can be written using the Levi-Civita symbol ε_{ijk} as defined in 18.15. The factor $\frac{1}{2}$ is introduced to correct for the double counting due to the contraction over both the indices j and k.

Property 18.3.24. Consider an inner product space V, then

$$\ast\ast\ \omega = (-1)^{k(n-k)}\omega \tag{18.32}$$

In $n = 4$ this leads to $\ast\ast\omega = \omega$ which means that the Hodge star is an involution in 4-dimensional inner product spaces.

Definition 18.3.25 (Self-dual). Let V be a 4-dimensional inner product space. Consider $\omega \in \Lambda^2(V)$. Then ω is said to be self-dual if $\ast\omega = \omega$. Furthermore every $v \in \Lambda^2(V)$ can be uniquely decomposed as the sum of a self-dual and an anti-self-dual 2-form.

18.3.6   Grassmann numbers

Although this section does not really belong to the chapter about tensors, we have included it here as it is an application of the concept of exterior algebras. The concept of Grassmann numbers (or variables) is used in QFT when performing calculations in the fermionic sector.

Definition 18.3.26 (Grassmann numbers). Let V be a complex vector space spanned by a set of generators θ_i. The Grassmann algebra with Grassmann variables θ_i is the exterior algebra over V. The wedge symbol of Grassmann variables is often ommitted when writing the product: $\theta_i \wedge \theta_j \equiv \theta_i\theta_j$.

Remark 18.3.27. Furthermore, from the anti-commutativity it follows that we can regard the Grassmann variables as being non-zero square-roots of zero.

Property 18.3.28. Consider a one-dimensional Grassmann algebra. When constructing the polynomial ring $\mathbb{C}[\theta]$ generated by θ, we see that, due to the anti-commutativity, $\mathbb{C}[\theta]$ 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-dimenisonal Grassmann algebra can be written as

$$p(\theta) = a + b\theta \tag{18.33}$$

Definition 18.3.29. We can equip the exterior algebra Λ with Grassmann variables θ_i with an involution similar to that on $\mathbb{C}$:

$$(\theta_i\theta_j...\theta_k)^* = \theta_k...\theta_j\theta_i \tag{18.34}$$

Elements $z \in \Lambda$ such that $z^* = z$ are called **(super)real**, elements such that $z^* = -z$ are called **(super)imaginary**. This convention is called the *DeWitt* convention.

Chapter 19

Clifford Algebra

19.1 Clifford algebra

Definition 19.1.1 (Clifford algebra). Let V be unital associative algebra. The Clifford algebra over V with quadratic form $Q : V \to K$ is the free algebra generated by V under the following condition:

$$v \cdot v = Q(v)1 \tag{19.1}$$

where 1 is the unit element in V. This condition implies that the square of a vector is a scalar.

Notation 19.1.2. The Clifford algebra corresponding to V and Q is denoted by $C\ell(V, Q)$.

Construction 19.1.3. The previous definition can be given an explicit construction. First we construct the tensor algebra of V:

$$T(V) = \bigoplus_{k \in \mathbb{N}} V^{\otimes k} \tag{19.2}$$

Then we construct a two-sided ideal I of V generated[1] by $\{v \otimes v - Q(v)1_V \mid v \in V\}$. The Clifford algebra $C\ell(V, Q)$ can then be constructed as the quotient algebra $T(V)/I$.

Remark 19.1.4. Looking at definition 18.3.15 we see that the exterior algebra $\Lambda^*(V)$ coincides with the Clifford algebra $C\ell(V, 0)$. If $Q \neq 0$ then the two algebras are still linearly isomorphic when $\text{char}(V) \neq 2$.

Property 19.1.5 (Dimension). If V has dimension n then $C\ell(V, Q)$ has dimension 2^n.

19.2 Geometric algebra

Definition 19.2.1 (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

---

[1]See definition 2.5.9.

the Clifford algebra $C\ell(V, g)$. If $\text{char}(V) \neq 2$ then the bilinear form uniquely determines a quadratic form $Q : v \mapsto g(v, v)$ as required in definition 19.1.1.

Definition 19.2.2 (Inner and exterior product). Analogous to the inner product in linear algebra and the wedge product in exterior algebras one can define an (a)symmetric product on the geometric algebra.

First of all we note that the product ab of two vectors a and b can be written as the sum of a symmetric and an antisymmetric part:

$$ab = \frac{1}{2}(ab + ba) + \frac{1}{2}(ab - ba) \tag{19.3}$$

We can then define the inner product as the symmetric part:

$$a \cdot b := \frac{1}{2}(ab + ba) = \frac{1}{2}\left((a + b)^2 - a^2 - b^2\right) = g(a, b) \tag{19.4}$$

Analogously we define the exterior (outer) product as the antisymmetric part:

$$a \wedge b := \frac{1}{2}(ab - ba) \tag{19.5}$$

These definitions allow us the rewrite formula 19.3 as:

$$\boxed{ab = a \cdot b + a \wedge b} \tag{19.6}$$

Looking at the last equality in the definition of the inner product 19.4 we see that condition 19.1 is satisfied when $a = b$. Furthermore we see that when g is fully degenerate, i.e. $g(v, v) = 0$ for all $v \in V$, the inner product is identically zero for all vectors and the geometric algebra coincides with exterior algebra[2] over V. For general forms g the exterior algebra is a subalgebra of the GA.

Definition 19.2.3 (Multivector). Any element of the GA over V is called a multivector. The simple multivectors of grade k, i.e. elements of the form $v_1 v_2 ... v_k$ with $v_i \in V$ for all i, are called k-blades. This generalizes the remark underneath 18.3.3. Sums of multivectors of different grades are called mixed multivectors[3].

Let $n = \dim(V)$. Multivectors of grade n are also called **pseudoscalars** and multivectors of grade $n - 1$ are also called **pseudovectors**.

Definition 19.2.4 (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 the grade operators we can extend the inner and exterior product to the complete GA as follows.

---

[2]See definition 18.3.14.

[3]These elements do not readily represent a geometric structure.

Formula 19.2.5. Let A, B be two multivectors of respectively grade m and n. Their inner product is defined as:

$$A \cdot B = \langle AB \rangle_{|m-n|} \tag{19.7}$$

Their exterior product is defined as:

$$A \wedge B = \langle AB \rangle_{m+n} \tag{19.8}$$

19.3   Pin group

19.3.1   Clifford group

Definition 19.3.1 (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 operator $\hat{\ }$ defined by:

$$\hat{v} = \begin{cases} v & v \in V_0 \\ -v & v \in V_1 \end{cases} \tag{19.9}$$

This operation can be generalized to all of $C\ell(V, Q)$ using linearity. The resulting operator is called the main involution or **inversion** on $C\ell(V, Q)$. Furthermore it turns the Clifford algebra into a superalgebra[4].

Formula 19.3.2 (Twisted conjugation). Let $v \in V$ be a vector and let $s \in C\ell(V, Q)$ be an invertible element of the Clifford algebra over V. The twisted conjugation of v by s is given by the map:

$$\chi : C\ell(V, Q) \times V : \chi(s)v = sv\hat{s}^{-1} \tag{19.10}$$

Definition 19.3.3 (Clifford group). The Clifford group $\Gamma(V, Q)$ is defined as follows:

$$\Gamma(V, Q) = \{s \in C\ell(V, Q) : sv\hat{s}^{-1} \in V, v \in V\} \tag{19.11}$$

It is the set of Clifford algebra elements that stabilize V under twisted conjugation. Furthermore this set is closed under multiplication and forms a group. It contains all invertible elements $s \in C\ell(V, Q)$.

Property 19.3.4. If V is finite-dimensional the map

$$\chi : \Gamma(V, Q) \to O(V, Q) : s \mapsto \chi(s) \tag{19.12}$$

defines a representation[5] called the *vectorial representation*. Furthermore, from the first isomorphism theorem 2.4.30 it follows that $O(V, Q)$ is isomorphic to $\Gamma(V, Q)/\ker \chi$ where $\ker \chi = \mathbb{R} \backslash \{0\}$. This isomorphism[6] also implies that the Clifford group is given by the the set of finite products of invertible elements $v \in V$:

$$\Gamma(V, Q) = \left\{ \prod_i^k s_i : s_i \text{ invertible in } V, n \in \mathbb{N} \right\} \tag{19.13}$$

---

[4]See definition 14.2.24.

[5]The surjectiveness of the map χ follows from the *Cartan-Dieudonné theorem*.

[6]Together with the *Cartan-Dieudonné theorem*.

Corollary 19.3.5. By nothing that pure rotations can be decomposed as an even number of reflections we find that:

$$\Gamma^+(V,Q)/\mathbb{R}_0 = SO(V,Q) \tag{19.14}$$

where Γ^+ is the intersection of the even Clifford algebra and the Clifford group.

19.3.2 Pin and Spin groups

Formula 19.3.6 (Spinor norm). On $\Gamma(V,Q)$ (and in fact on al of $C\ell(V,Q)$) one can define the spinor norm:

$$\mathcal{N}(x) = \hat{x}x \tag{19.15}$$

where $\hat{\ }$ is the main involution. The map $|\mathcal{N}|$ gives a group homomorphism from $\Gamma(V,Q)$ to $\mathbb{R}_0^+$.

Definition 19.3.7 (Pin and spin groups). Using the spinor norm $\mathcal{N}$ we can now define the pin and spins groups as follows:

Part V

Differential Geometry

Chapter 20

Curves and Surfaces

20.1 Curves

Property 20.1.1 (Regular curve). Let $\vec{c}(t)$ be a curve defined on an interval I. $\vec{c}(t)$ is said to be regular if $\frac{d\vec{c}}{dt} \neq \vec{0}$ for all $t \in I$.

Definition 20.1.2 (C^r-parameter transformation). A transformation $\varphi :]c,d[\to]a,b[:$ $u \mapsto t$ such that $\varphi(u)$ is a C^r-diffeomorphism[1].

Definition 20.1.3 (Geometric property). A geometric property is a property that is invariant under:

1. parameter transformations

2. positive orthonormal changes of basis

Theorem 20.1.4. *Let $\vec{c}(t), \vec{d}(t)$ be two curves with the same image. We then have the following relation:*
$$\vec{c}(t) \ regular \iff \vec{d}(t) \ regular \tag{20.1}$$

20.1.1 Arc length

Definition 20.1.5 (Natural parameter). Let $\vec{c}(t)$ be a curve. The parameter t is said to be a natural parameter if:
$$\left\|\frac{d\vec{c}}{dt}\right\| \equiv 1 \tag{20.2}$$

Formula 20.1.6 (Arc length). The following function $\phi(t)$ is a bijective map and a natural parameter of $\vec{c}(t)$:
$$\phi(t) = \int_{t_0}^{t} ||\dot{\vec{c}}(t)||dt \tag{20.3}$$

---

[1]See definition 14.3.12

Remark. The arc length as defined above is often denoted by 's'.

Theorem 20.1.7. *Let $\vec{c}(t)$ be a curve. Let u be an alternative parameter of $\vec{c}(t)$. It is a natural parameter if and only if there exists a constant α such that:*

$$u = \pm s + \alpha$$

where s is the integral as defined in equation 20.3.

Remark. As the last theorem implies, no unique natural parameter or arc length exists.

20.1.2 Frenet-Serret frame

Definition 20.1.8 (Tangent vector). Let $\vec{c}(s)$ be parametrized by arc length. The tangent vector $\vec{t}(s)$ is defined as:
$$\vec{t}(s) = \vec{c}\,'(s) \tag{20.4}$$

Property 20.1.9. From the definition of the natural parametrization 20.2 and the previous definition it follows that the tangent vector is a unit vector:

$$\vec{t}(s) \cdot \vec{t}(s) = \left\lVert \frac{d\vec{c}(s)}{ds} \right\rVert^2 = 1$$

Definition 20.1.10 (Principal normal vector). Let $\vec{c}(s)$ be parametrized by arc length. The principal normal vector is defined as:

$$\vec{n}(s) = \frac{\vec{t}\,'(s)}{\lVert \vec{t}\,'(s) \rVert} \tag{20.5}$$

Property 20.1.11. From property 20.1.9 and the definition of the principal normal vector it follows that the tangent vector and principal normal vector are orthogonal:

$$\vec{t}(s) \cdot \vec{t}(s) = 1 \implies \vec{t}(s) \cdot \vec{t}\,'(s) = 0 \implies \vec{t}(s) \cdot \vec{n}(s) = 0$$

Definition 20.1.12 (Binormal vector). Let $\vec{c}(s)$ be parametrized by arc length. The binormal vector is defined as:
$$\vec{b}(s) = \vec{t}(s) \times \vec{n}(s) \tag{20.6}$$

Definition 20.1.13 (Frenet-Serret frame). As the vectors $\vec{t}(s), \vec{n}(s)$ and $\vec{b}(s)$ are mutually orthonormal and linearly independent, we can use them to construct a positive orthonormal basis. The ordered basis $\left(\vec{t}(s), \vec{n}(s), \vec{b}(s) \right)$ is called the **Frenet-Serret** frame.

Remark. This basis does not have to be the same in every point of $\vec{c}(s)$.

Definition 20.1.14 (Curvature). Let $\vec{c}(s)$ be parametrized by arc length. The curvature of $\vec{c}(s)$ is defined as:
$$\frac{1}{\rho(s)} = \lVert \vec{t}\,'(s) \rVert \tag{20.7}$$

Definition 20.1.15 (Torsion). Let $\vec{c}(s)$ be a curve parametrized by arc length. The torsion of $\vec{c}(s)$ is defined as:

$$\tau(s) = \rho(s)^2 (\vec{t} \ \ \vec{t}' \ \ \vec{t}'') \tag{20.8}$$

Formula 20.1.16 (Frenet formulas). The derivatives of the tangent, principal normal and binormal vectors can be written as a linear combination of the those vectors themself as:

$$\begin{cases} \vec{t}'(s) & = & & \frac{1}{\rho(s)}\vec{n}(s) \\ \vec{n}'(s) & = & -\frac{1}{\rho(s)}\vec{t}(s) & + & \tau(s)\vec{b}(s) \\ \vec{b}'(s) & = & & -\tau(s)\vec{n}(s) \end{cases} \tag{20.9}$$

Theorem 20.1.17 (Fundamental theorem of curves). *Let $k(s), w(s) : U \to \mathbb{R}$ be two $\mathcal{C}^1$ functions with $k(s) \geq 0, \forall s$. There exists an interval $]-\varepsilon, \varepsilon[\subset U$ and a curve $\vec{c}(s) :$ $]-\varepsilon, \varepsilon[\to \mathbb{R}^3$ with natural parameter s such that $\vec{c}(s)$ has $k(s)$ as its curvature and $w(s)$ as its torsion.*

20.2   Surfaces

Notation 20.2.1. Let $\vec{\sigma}$ be a surface[2]. The derivative of $\vec{\sigma}$ with respect to the coordinate q^i is written as follows:

$$\frac{\partial \vec{\sigma}}{\partial q^i} = \vec{\sigma}_i \tag{20.10}$$

20.2.1   Tangent vectors

Definition 20.2.2 (Tangent plane). Let $P(q_0^1, q_0^2)$ be a point on the surface Σ. The tangent space $T_P\Sigma$ to $\vec{\sigma}$ in P is defined as follows:

$$\forall \vec{r} \in T_P\Sigma : \left[\vec{r} - \vec{\sigma}(q_0^1, q_0^2)\right] \cdot \left[\vec{\sigma}_1(q_0^1, q_0^2) \times \vec{\sigma}_2(q_0^1, q_0^2)\right] = 0 \tag{20.11}$$

Definition 20.2.3 (Normal vector). The cross product in equation 20.11 is closely related to the normal vector to Σ in P. The normal vector in the point (q_0^1, q_0^2) is defined as:

$$\vec{N}(q_0^1, q_0^2) = \frac{1}{||\vec{\sigma}_1 \times \vec{\sigma}_2||} (\vec{\sigma}_1 \times \vec{\sigma}_2) \tag{20.12}$$

20.2.2   First fundamental form

Definition 20.2.4 (Metric coefficients). Let $\vec{\sigma}$ be a surface. The metric coefficients g_{ij} are defined as follows:

$$\boxed{g_{ij} = \vec{\sigma}_i \cdot \vec{\sigma}_j} \tag{20.13}$$

---

[2]$\vec{\sigma}$ denotes the surface as a vector field. Σ denotes the geometric image of $\vec{\sigma}$.

Definition 20.2.5 (Scale factor). The following factors are often used in vector calculus:

$$g_{ii} = h_i^2 \tag{20.14}$$

Definition 20.2.6 (First fundamental form). Let $\vec{\sigma}$ be a surface. Define a bilinear form $I_P(\vec{v}, \vec{w}) : T_P\Sigma \times T_P\Sigma \to \mathbb{R}$ that restricts the inner product to $T_P\Sigma$:

$$I_P(\vec{v}, \vec{w}) = \vec{v} \cdot \vec{w} \tag{20.15}$$

This bilinear form is called the first fundamental form or **metric**.

Corollary 20.2.7. All $\vec{v}, \vec{w} \in T_P\Sigma$ are linear combinations of the tangent vectors $\vec{\sigma}_1, \vec{\sigma}_2$. This leads to the following relation between the first fundamental form and the metric coefficients 20.13:

$$I_P(\vec{v}, \vec{w}) = v^i \vec{\sigma}_i \cdot w^j \vec{\sigma}_j = g_{ij} v^i w^j$$

Notation 20.2.8. The length can be written as

$$s = \int \sqrt{||\dot{\vec{c}}(t)||} dt = \int \sqrt{ds^2}$$

where the second equality is formally defined. The two equalities together can be combined into the following notation for the metric:

$$\boxed{ds^2 = g_{ij} dq^i dq^j} \tag{20.16}$$

Formula 20.2.9. Let (g_{ij}) be the metric tensor. We define the matrix (g^{ij}) as its inverse:

$$(g^{ij}) = \frac{1}{\det(g_{ij})} \begin{pmatrix} g_{22} & -g_{12} \\ -g_{12} & g_{11} \end{pmatrix} \tag{20.17}$$

20.2.3   Isometries

Definition 20.2.10 (Isometry). An isometry is a distance-preserving map, i.e. a diffeomorphism $\Phi : \Sigma \to \Sigma'$ that maps arc segments in Σ to arc segments with the same length in Σ'.

Property 20.2.11. A diffeomorphism Φ is an isometry if and only if the metric coefficients of σ and σ' are the same.

Definition 20.2.12 (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 20.2.13. A diffeomorphism Φ is conformal if and only if the metric coefficients of σ and σ' are proportional.

Definition 20.2.14 (Surface preserving map). A diffeomorphism $\Phi : \Sigma \to \Sigma'$ is sadi to be surface-preserving if it maps a segment of Σ to a segment of Σ' with the same surface.

Property 20.2.15. A diffeomorphism Φ is surface-preserving if and only if the metric coefficients of σ and σ' satisfy:

$$g'_{11}g'_{22} - (g'_{12})^2 = g_{11}g_{22} - g_{12}^2 \tag{20.18}$$

for all points (q^1, q^2).

Corollary 20.2.16. A map that is surface-preserving and conformal is also isometric.

20.2.4 Second fundamental form

Definition 20.2.17 (Second fundamental form). Let $\vec{\sigma}(q^1, q^2)$ be a surface. The second fundamental form is a bilinear form $II_P(\vec{v}, \vec{w}) : T_P\Sigma \times T_P\Sigma \to \mathbb{R}$ defined as follows:

$$\boxed{II_P(\vec{v}, \vec{w}) = L_{ij}(q^1, q^2)v^i w^j} \tag{20.19}$$

where $L_{ij} = \vec{N} \cdot \vec{\sigma}_{ij}$.

Definition 20.2.18 (Normal curvature). Let $\vec{c}$ be a curve parametrized as

$$\vec{c}(s) = \vec{\sigma}\left(q^1(s), q^2(s)\right)$$

The normal curvature of $\vec{c}(s)$ at a point $(q^1(s), q^2(s))$ is defined as:

$$\boxed{\frac{1}{\rho_n(s)} = \vec{c}\,''(s) \cdot \vec{N}(s)} \tag{20.20}$$

From the definition of the second fundamental form it follows that the normal curvature can be written as:

$$\frac{1}{\rho_n(s)} = II(\vec{t}, \vec{t}) = \frac{II\left(\dot{\vec{c}}(t), \dot{\vec{c}}(t)\right)}{I\left(\dot{\vec{c}}(t), \dot{\vec{c}}(t)\right)} \tag{20.21}$$

where the last equality holds for any given parameter t.

Theorem 20.2.19 (Meusnier's theorem). *Let $\vec{c}, \vec{d}$ be two curves on a surface $\vec{\sigma}$. The curves have the same normal curvature in a point $(q^1(t_0), q^2(t_0))$ if $\vec{c}(t_0) = \vec{d}(t_0)$ and if $\dot{\vec{c}}(t_0) \parallel \dot{\vec{d}}(t_0)$. Furthermore, the osculating circles of all curves with the same normal curvature at a given point form a sphere.*

Property 20.2.20. The normal curvature of at a given point is equal to the curvature of the normal section, i.e. the intersection of the surface with a normal plane at the point.

Definition 20.2.21 (Geodesic curvature). Let $\vec{c}$ be a curve parametrized as $\vec{c}(s) = \vec{\sigma}\left(q^1(s), q^2(s)\right)$. The geodesic curvature of $\vec{c}(s)$ at a point $(q^1(s), q^2(s))$ is defined as:

$$\boxed{\frac{1}{\rho_g(s)} = \left(\vec{N}(s)\ \vec{t}(s)\ \vec{t}\,'(s)\right)} \tag{20.22}$$

Formula 20.2.22. Let $\vec{c}$ be a curve on a surface $\vec{\sigma}$. From the definitions of the normal and geodesic curvature it follows that:

$$\boxed{\frac{1}{\rho^2} = \frac{1}{\rho_n^2} + \frac{1}{\rho_g^2}} \tag{20.23}$$

20.2.5   Curvature of a surface

Definition 20.2.23 (Weingarten map). Let P be a point of a surface Σ. The Weingarten map $L_P : T_P\Sigma \to T_P\Sigma$ is a linear map defined as:

$$L_P(\vec{\boldsymbol{\sigma}}_1) = -\vec{\boldsymbol{N}}_1 \qquad \text{and} \qquad L_P(\vec{\boldsymbol{\sigma}}_2) = -\vec{\boldsymbol{N}}_2 \tag{20.24}$$

Formula 20.2.24. Let $\vec{v}, \vec{w} \in T_P\Sigma$. The following equalities hold:

$$L_P(\vec{v}) \cdot \vec{w} = L_P(\vec{w}) \cdot \vec{v} = II_P(\vec{v}, \vec{w}) \tag{20.25}$$

Formula 20.2.25 (Matrix elements of L_P). Let (g^{ij}) be the inverse of the metric tensor. The matrix elements of L_P are defined as:

$$L_j^k = g^{ki} L_{ij}$$

Formula 20.2.26 (Weingarten formulas).

$$\vec{\boldsymbol{N}}_j = -L_j^k \vec{\boldsymbol{\sigma}}_k \tag{20.26}$$

Theorem 20.2.27. *For every point P on the surface Σ there exists a basis $\{\vec{\boldsymbol{h}}_1, \vec{\boldsymbol{h}}_2\} \subset T_P\Sigma$ of eigenvectors of L_P. Furthermore, the corresponding eigenfunctions are given by $II_P(\vec{\boldsymbol{h}}_i, \vec{\boldsymbol{h}}_i)$ and these eigenvalues are the extreme values of the normal curvature at the point P.*

Definition 20.2.28 (Principal curvatures). The eigenvalues of the Weingarten map are called the principal curvatures of the surface and they are denoted by $\frac{1}{R_1}$ and $\frac{1}{R_2}$. The tangent vectors corresponding to these curvatures are called the **principal directions**.

Remark. If the principal curvatures are not equal, the principal directions are orthogonal. If they are equal, the point P is said to be an **umbilical point** or **umbilic**.

Property 20.2.29. For the principal directions we have $L_P(\vec{\boldsymbol{h}}_1) \cdot \vec{\boldsymbol{h}}_2 = 0$[3]. If P is an umbilic then every tangent vector in P is a principal direction and the equality is satisfied for every two tangent vectors.

Definition 20.2.30 (Gaussian curvature). The Gaussian curvature K of a surface is defined as the determinant of the Weingarten map, i.e.:

$$K = \frac{1}{R_1 R_2} \tag{20.27}$$

Definition 20.2.31 (Mean curvature). The mean curvature H of a surface is defined as the trace of the Weingarten map, i.e.:

$$H = \frac{1}{2}\left(\frac{1}{R_1} + \frac{1}{R_2}\right) \tag{20.28}$$

---

[3]Tangent vectors that satisfy this equation are called **adjoint** tangent vectors.

Property 20.2.32. The principal curvatures are the solutions of the following equation:

$$x^2 - 2Hx + K = 0$$

This is the characteristic equation (14.57) of the Weingarten map.

Definition 20.2.33. 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.

Remark. From previous definition it follows that a flat point is a special type of umbilic.

Theorem 20.2.34. *A surface Σ containing only umbilics is a part of a sphere or a part of a plane.*

Theorem 20.2.35. *In the neighbourhood of a point P of a surface with principal curvatures $1/R_1$ and $1/R_2$ is locally the same as the following quadric:*

$$x_3 = \frac{1}{2}\left(\frac{x_1^2}{R_1} + \frac{x_2^2}{R_2}\right) \tag{20.29}$$

if we ignore terms of order > 2.

Theorem 20.2.36 (Euler's formula). *The normal curvature of a couple $(P, \vec{e})$ where is $\vec{e} = \vec{h}_1 \cos\theta + \vec{h}_2 \sin\theta \in T_P\Sigma$ is given by:*

$$\frac{1}{\rho_n} = \frac{\cos^2\theta}{R_1} + \frac{\sin^2\theta}{R_2} \tag{20.30}$$

Definition 20.2.37 (Asymptotic curve). An asymptotic curve is a curve which is in every point P tangent to a direction with zero normal curvature.

Formula 20.2.38 (Differential equation for asymptotic curves).

$$L_{11}\left(\dot{q}^1(t)\right)^2 + 2L_{12}\dot{q}^1(t)\dot{q}^2(t) + L_{22}\left(\dot{q}^2(t)\right)^2 = 0 \tag{20.31}$$

Property 20.2.39. A curve on a surface is an asymptotic curve if and only if the tangent plane and the osculation plane coincide in every point P of the surface.

Definition 20.2.40 (Line of curvature). A curve is a line of curvature if the tangent vector in every point P is a principal direction of the surface in P.

Formula 20.2.41 (Rodrigues' formula). A curve is a line of curvature if and only if it satisfies the following formula:

$$\frac{d\vec{N}}{dt}(t) = -\frac{1}{R(t)}\frac{d\vec{c}}{dt}(t) \tag{20.32}$$

If the curve satisfies this formula, then the scalar function $1/R(t)$ coincides with the principal curvature along the curve.

Formula 20.2.42 (Differential equation for curvature lines).

$$\begin{vmatrix} (\dot{q}^2)^2 & -\dot{q}^1\dot{q}^2 & (\dot{q}^1)^2 \\ g_{11} & g_{12} & g_{22} \\ L_{11} & L_{12} & L_{22} \end{vmatrix} = 0 \tag{20.33}$$

Property 20.2.43. From theorem 20.2.27 we know that the principal directions are orthogonal vectors. It follows that on a surface containing no umbilics the curvature lines form an orthogonal web and in every point P we find 2 orthogonal curvature lines.

20.2.6 Christoffel symbols and geodesics

Formula 20.2.44 (Gauss' formulas).

$$\vec{\sigma}_{ij} = L_{ij}\vec{N} + \Gamma^k{}_{ij}\vec{\sigma}_k \tag{20.34}$$

where the **Christoffel symbols** $\Gamma^k{}_{ij}$ are defines as:

$$\boxed{\Gamma^k{}_{ij} = g^{kl}\vec{\sigma}_l \cdot \vec{\sigma}_{ij}} \tag{20.35}$$

Corollary 20.2.45. From the expression of the Christoffel symbols we can derive an alternative expression using only the metric tensor g_{ij}:

$$\Gamma^k{}_{ij} = \frac{1}{2}g^{kl}\left(\frac{\partial g_{il}}{\partial q^j} - \frac{\partial g_{ij}}{\partial q^l} + \frac{\partial g_{jl}}{\partial q^i}\right) \tag{20.36}$$

Definition 20.2.46 (Geodesic). A geodesic is a curve with zero geodesic curvature.

Theorem 20.2.47. *A curve on a surface is an geodesic if and only if the tangent plane and the osculation plane are orthogonal in every point P of the surface.*

Formula 20.2.48 (Differential equation for geodesic). If the curve is parametrized by arc length, then it is a geodesic if the functions $q^1(s)$ and $q^2(s)$ satisfy the following differential equation:

$$\boxed{q''^k + \Gamma^k{}_{ij}q'^i q'^j = 0} \tag{20.37}$$

20.2.7 Theorema Egregium

Formula 20.2.49 (Codazzi-Mainardi equations).

$$\frac{\partial L_{ij}}{\partial q^k} - \frac{\partial L_{ik}}{\partial q^j} = \Gamma^l{}_{ik}L_{lj} - \Gamma^l{}_{ij}L_{lk} \tag{20.38}$$

Definition 20.2.50 (Riemann curvature tensor).

$$\boxed{R^l{}_{ijk} = \frac{\partial \Gamma^l{}_{ik}}{\partial q^j} - \frac{\partial \Gamma^l{}_{ij}}{\partial q^k} + \Gamma^s{}_{ik}\Gamma^l{}_{sj} - \Gamma^s{}_{ij}\Gamma^l{}_{ks}} \tag{20.39}$$

Formula 20.2.51 (Gauss' equations).

$$R^l{}_{ijk} = L_{ik}L^l_j - L_{ij}L^l_k \tag{20.40}$$

Theorem 20.2.52 (Theorema Egregium). *The Gaussian curvature K (formula 20.27) is completely determined by the metric tensor g_{ij} and its derivatives:*

$$\boxed{K = \frac{R^l{}_{121}g_{l2}}{g_{11}g_{22} - g_{12}^2}} \tag{20.41}$$

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 tensor.

Property 20.2.53. From the condition of isometries 20.2.11 and the previous theorem it follows that if two surfaces are connected by an isometric map, the corresponding points in Σ and Σ' have the same Gaussian curvature.

Corollary 20.2.54. There exists no isometric projection from the sphere to the plane. This also implies that a perfect (read: isometric) map of the Earth can not be created.

Chapter 21

Manifolds

21.1 Charts

Definition 21.1.1 (Chart). Let M be a set. Let U be an open subset of M. Let $\varphi : U \to \mathbb{R}^n$ be at least a C^1-diffeomorphism. The pair (U, φ) is called a chart.

Definition 21.1.2 (Atlas). Let M be a set. Let $\{(U_i, \varphi_i)\}_i$ be a set of charts such that $\bigcup_i U_i = M$. This set of charts is called an atlas.

Definition 21.1.3 (Maximal Atlas). Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be two atlasses covering the same set M. If $\mathcal{A}_1 \cup \mathcal{A}_2 = \mathcal{A}$ is also an atlas, the atlasses are said to be equivalent. The largest such union is called a **maximal** atlas.

Definition 21.1.4 (Manifold). A set M equipped with a maximal atlas $\mathcal{A}$ is called a topological manifold. An alternative definition (often used in topology) is that of a locally Euclidean Hausdorff space.

Remark. In the previous definition second-countability is also often mentioned in the literature. This ensures that the space has certain 'desirable' properties.

Definition 21.1.5 (Transition map). Let $\mathcal{A}$ be an atlas for a set M. Let (U_1, φ_1) and (U_2, φ_2) be two charts in $\mathcal{A}$. The mapping $\varphi_1^{-1} \circ \varphi_2$ is called a transition map. If all transition maps are C^k-diffeomorphisms than the manifold is called a C^k-manifold.

Definition 21.1.6 (Smooth manifold). A C^∞-manifold is also called a smooth manifold.

Formula 21.1.7 (Smooth function). Let $f : M \to N$ be a function between two smooth manifolds. Let (U, φ) and (V, ψ) be charts for M and N such that $f(U) \subseteq V$. If the function

$$f_{\varphi\psi} = \psi \circ f \circ \varphi^{-1} \tag{21.1}$$

is of class C^∞ then the function f is also said to be smooth.

Remark. The function 21.1 is called the **local representation** of f.

Notation 21.1.8. The set of all C^∞ functions on a manifold M defined on a neighbourhood of $m \in M$ is denoted by $\mathcal{F}_m(M)$. This set forms a commutative ring when equipped with the usual sum and product (composition) of functions.

21.2 Tangent vectors

Definition 21.2.1 (Tangent vector). Let M be a smooth manifold and $p \in M$. Let $f, g : M \to \mathbb{R} \in \mathcal{F}_p(M)$. A tangent vector on M is a differential operator v_p satisfying the following properties:

1. Linearity: $v_p(af + g) = av_p(f) + v_p(g)$

2. Leibniz property: $v_p(fg) = f(p)v_p(g) + g(p)v_p(f)$

Maps with these properties are also called **derivations**[1].

Property 21.2.2. For every constant function $c : p \mapsto c$ we have:

$$v_p(c) = 0 \tag{21.2}$$

Definition 21.2.3 (Tangent space). Following from the previous definition, we can construct a tangent (vector) space T_pM in each point $p \in M$. First we construct a tangent vector:

$$\boxed{\left. \frac{\partial}{\partial q^i} \right|_p : \mathcal{F}_p(M, \mathbb{R}) \to \mathbb{R} : f \mapsto \frac{\partial}{\partial q^i} \left(f \circ \varphi^{-1} \right) (\varphi(p))} \tag{21.3}$$

where (U, φ) is a coordinate chart such that $p \in U$ and $(q^1, ..., q^n)$ are local coordinates. These tangent vectors $\frac{\partial}{\partial q^i}|_p$ form a basis for T_pM.

Remark 21.2.4. 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 will not be the same.

Definition 21.2.5 (Curve). A smooth function $\gamma : \mathbb{R} \to M$ with $\gamma(0) = m$ is called a smooth curve through $m \in M$.

Alternative Definition 21.2.6 (Tangent space). The alternative construction goes as follows. Two smooth curves γ_1, γ_2 through $p \in M$ are said to be tangent at p if:

$$\frac{d(\varphi \circ \gamma_1)}{dt}(0) = \frac{d(\varphi \circ \gamma_2)}{dt}(0) \tag{21.4}$$

equivalently, if there local representatives are tangent in 0. Furthermore, these formulae transform as follows:

$$\frac{d(\psi \circ \gamma)}{dt}(0) = D(\psi \circ \varphi^{-1})(\varphi(p))\frac{d(\varphi \circ \gamma)}{dt}(0) \tag{21.5}$$

This relation in fact imposes an equivalence relation on the set of smooth curves through p. One then defines the tangent space at p as the set of equivalence classes of tangent curves through p. Explicitly these equivalence classes are constructed as follows:

---

[1]Generally, every operation that satisfies the Leibniz property is called a derivation.

We can define the following tangent vector to the curve c through p:

$$v_p(f) = \left.\frac{d(f \circ c)}{dt}\right|_{t=0} \tag{21.6}$$

Applying the chain rule gives us

$$v_p(f) = \frac{\partial(f \circ \varphi^{-1})}{\partial q^i}(\varphi(p))\frac{dq^i}{dt}(0) \tag{21.7}$$

where $q^i = (\varphi \circ c)^i$. The first factor depends only on the point p and the second factor is equal for all tangent curves through p. We thus see that tangent curves define the same tangent vector.

Now we should prove that both definitions of the tangent space are in fact really equivalent.

Proof. Using 21.3 we can rewrite equation 21.7 as follows:

$$v_p(f) = \left.\frac{\partial f}{\partial q^i}\right|_p \frac{dq^i}{dt}(0) \tag{21.8}$$

Because the partial derivatives 21.3 are also basis vectors for the tangent space, we see that this equation is in fact an expansion in those basis vectors. So the representatives of the equivalence classes of tangent curves are indeed tangent vectors according to the first definition.

To prove the other direction we have to show that the partial derivative operators can be constructed as tangent vectors to curves.

A tangent vector (from the first definition) can be expanded as

$$v_p = v^i \left.\frac{\partial}{\partial q^i}\right|_p \tag{21.9}$$

with $v = (v^1, ..., v^n)$. By constructing the curve $\gamma : t \mapsto \varphi^{-1}(q_0 + vt)$ we see that this tangent vector v_p is indeed tangent to a curve. So we have an isomorphism between the tangent vectors from the first definition and the equivalence classes of tangent curves. Both definitions of the tangent space are thus equivalent. $\qquad\square$

Remark 21.2.7. From 21.9 it follows that the tangent vectors can be viewed as directional derivatives.

Definition 21.2.8 (Tangent bundle). The disjoint union[2] of all tangent spaces

$$TM = \bigsqcup_{p \in M} T_p M \tag{21.10}$$

is called the tangent bundle on M.

---

[2]See definition 4.1.3.

21.3   Curvature

Formula 21.3.1 (Riemann Curvature Tensor). Let $V \in TM$. Let D_μ be the covariant derivative.

$$\boxed{[D_\mu, D_\nu]V^\rho = R^\rho{}_{\kappa\mu\nu}V^\kappa} \tag{21.11}$$

Formula 21.3.2 (Ricci tensor).

$$R_{\mu\nu} = R^\lambda{}_{\mu\lambda\nu} \tag{21.12}$$

Formula 21.3.3 (Ricci scalar).

$$R = R^\mu{}_\mu \tag{21.13}$$

This scalar quantity is also called the **scalar curvature**.

Formula 21.3.4 (Einstein Tensor).

$$\boxed{G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R} \tag{21.14}$$

Theorem 21.3.5. *For 4-dimensional manifolds the Einstein tensor $G_{\mu\nu}$ is the only tensor containing at most second derivatives of the metric $g_{\mu\nu}$ and satisfying:*

$$\nabla_\mu G^{\mu\nu} = 0 \tag{21.15}$$

21.4   Submanifolds

Definition 21.4.1 (Immersion). Let $f : M \to N$ be a differentiable function between smooth manifolds. f is called an immersion if its derivative is everywhere injective, or equivalently if its derivative has maximal rank[3] everywhere:

$$\mathrm{rk}_p(f) = \dim(M), \forall p \in M \tag{21.16}$$

Definition 21.4.2 (Submersion). Let $f : M \to N$ be a differentiable map between smooth manifolds. A **regular point**[4] of f is a point $x \in M$ such that $T_x f$ is surjective. f is called a submersion if its derivative is everywhere surjective, or equivalently if

$$\mathrm{rk}_p(f) = \dim(N), \forall p \in M \tag{21.17}$$

Definition 21.4.3 (Embedding). A differentiable function between smooth manifolds is called a (smooth) embedding if its both an injective immersion and an embedding in the general sense 4.3.10. This implies that the submanifold topology coincides with the subspace (relative) topology 4.1.

---

[3]See definition 24.3.6.
[4]$f(x)$ is then called a **regular value**.

Definition 21.4.4 (Embedded submanifold). Let M be a manifold. A subset N is an embedded[5] submanifold if the inclusion map $f : M \hookrightarrow N$ is an embedding.

Definition 21.4.5 (Slice). Let $m < n$ be two positive integers. The space $\mathbb{R}^m$ can be viewed as a subspace of $\mathbb{R}^n$ by identifying them in the following way:

$$\mathbb{R}^m \cong \mathbb{R}^m \times \underbrace{\{0, ..., 0\}}_{n-m} \overset{\iota}{\hookrightarrow} \mathbb{R}^m \times \mathbb{R}^{n-m} \cong \mathbb{R}^n \tag{21.18}$$

where $\iota : (x_1, ..., x_m) \mapsto (x_1, ..., x_m, \underbrace{0, ..., 0}_{n-m})$ is the canonical inclusion map.

Alternative Definition 21.4.6. A k-dimensional embedded manifold N of M can now be defined equivalently as a subset of M such that there exists a positive integer k and such that for every point $p \in N$ there exists a chart (U, φ) with

$$\varphi(U \cap N) = \varphi(U) \cap (\mathbb{R}^k \times \underbrace{\{0, ..., 0\}}_{n-k}) \tag{21.19}$$

where $n = \dim(M)$. The set $U \cap N$ is called a slice of (U, φ) in analogy with the previous definition of a (standard) slice.

Theorem 21.4.7 (Submersion theorem[6]). *Consider a smooth map $f : M_1 \to M_2$ between smooth manifolds. Let $y \in M_2$ be a regular value. Then $N = f^{-1}(y)$ is a submanifold of M_1 with codimension $\dim(M_2)$.*

21.5 Manifolds with boundary

Definition 21.5.1 (Manifold with boundary). Let $\mathbb{H}^n$ denote the upper half space, i.e.:

$$\mathbb{H}^n = \{(x_1, ..., x_n) | x_n \geq 0\} \subset \mathbb{R}^n \tag{21.20}$$

An n-dimensional manifold with boundary is then given by a set M together 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 21.5.2 (Manifold 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 $\text{Int}(M) = M$ ∂M, in the sense of manifolds, should not be confused with the topological interior.

Property 21.5.3. Let M be an n-dimensional manifold with boundary. Let (U, φ) be a chart for $p \in \partial M$. Then

$$\varphi(p) \in \partial \mathbb{H}^n = \{(x_1, ..., x_n) | x_n = 0\} \tag{21.21}$$

---

[5] An immersed submanifold is defined analogously. The requirement of the inclusion map being an embedding is relaxed to it being an immersion. However the submanifold topology will no longer coincide with the relative topology.

[6] Also called the **regular value theorem**.

Chapter 22

Lie groups and Lie algebras

22.1 Lie groups

Definition 22.1.1 (Lie group). A Lie group is a group that is also a differentiable manifold such that both the multiplication and inversion are smooth functions.

Definition 22.1.2 (Lie subgroup). A subset of a Lie group is a Lie subgroup if it is both a subgroup and a closed submanifold.

Theorem 22.1.3 (Closed subgroup theorem[1]). *If H is a closed[2] subgroup of a Lie group G then H is a Lie subgroup of G.*

Definition 22.1.4 (Dimension). The dimension of a Lie group equals the number of parameters needed to parametrize the group.

Property 22.1.5. 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.

22.1.1 Left invariant vector fields

Definition 22.1.6 (Left Invariant Vector Field (LIVF)). Let G be a Lie group. Let X be a vector field on G. X is left invariant if the following equivariance relation holds for all $g \in G$:

$$L_{g,*}X(h) = X(g \cdot h) \tag{22.1}$$

where L_g denotes the left action map associated with g.

Property 22.1.7. The set $\mathcal{L}(G)$ of LIVF's on a Lie group G is a vector space over $\mathbb{R}$.

---

[1]Sometimes called *Cartan's theorem.*
[2]With respect to the group topology on G.

Property 22.1.8. 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 $v \in T_e(G)$, there exists a LIVF $X \in \mathcal{L}(G)$ such that $X(e) = v$ and this mapping is an isomorphism from $T_e(G)$ to $\mathcal{L}(G)$.

22.1.2 One-parameter subgroups

Definition 22.1.9 (One-parameter subgroup). A one-parameter (sub)group is a continuous group homomorphism $\Phi : \mathbb{R} \to G$ from the additive group of real numbers to a Lie group G.

Property 22.1.10. Let $\Phi : \mathbb{R} \to G$ be a one-parameter subgroup of G. Let $\Psi : G \to H$ be a continuous group homomorphism. Then $\Psi \circ \Phi : \mathbb{R} \to H$ is a one-parameter subgroup of H.

Property 22.1.11. Let X be a LIVF on a Lie group G. Let γ_X be the integral curve of X through $e \in G$. The maximal flow domain $D(X)$ is $]-\infty, +\infty[$ and the flow[3] σ_t determines a one-parameter subgroup on G. Furthermore, for every one-parameter subgroup $\phi(t)$ we can construct a LIVF $X = \phi'(0)$. This correspondence is a bijection.

22.1.3 Cocycles

Definition 22.1.12 (Cocycle). Let M be a smooth manifold and G a Lie group. A cocycle on M with values in G is a family of smooth functions $g_{ij} : U_i \cap U_j \to G$ that satisfy the following condition:

$$g_{ij} = g_{ik} \circ g_{kj} \tag{22.2}$$

Property 22.1.13. Let $\{g_{ij}\}_{i,j}$ be a cocycle on M. We have the following properties:

- $g_{ii}(x) = \mathbb{1}_M$
- $g_{ij}(x) = (g_{ji}(x))^{-1}$

for all $x \in M$.

22.2 Lie algebras

There are two ways to define a Lie algebra. The first one is a stand-alone definition using a vector space equipped with a multiplication operation. The second one establishes a direct relation between Lie groups (see 22.1.1) and real Lie algebras.

---

[3]See definition 24.27.

22.2.1   Definitions

Definition 22.2.1 (Lie algebra). Let V be a vector space equipped with a binary operation $[\cdot, \cdot] : V \times V \to V$ is a Lie algebra if the Lie bracket $[\cdot, \cdot]$ satisfies the following conditions:

1. Bilinearity: $[ax + y, z] = a[x, z] + [y, z]$

2. Alternativity: $[v, v] = 0$

3. Jacobi identity: $[a, [b, c]] + [b, [c, a]] + [c, [a, b]] = 0$

Remark 22.2.2 (Lie bracket). For the general linear group GL_n the Lie bracket is given by the commutator:

$$\boxed{[X, Y] = XY - YX} \tag{22.3}$$

This follows from the definition $[X, Y] = \frac{d}{dt}\mathrm{Ad}_{\gamma(t)}(Y)\big|_{t=0}$ with $\gamma(0) = e$ and $\gamma'(0) = X$.

Definition 22.2.3. Let G bie a Lie group. The tangent space $T_e G$ has the structure of a Lie algebra where the Lie bracket is given by the commutator of vector fields 24.32.

The following remark gives the equality of the set of LIVF's on a Lie group G and the Lie algebra $\mathfrak{g} := T_e G$:

Remark 22.2.4. From the second part of property 22.1.8 it follows that the Lie algebra $\mathfrak{g}$ associated to G is (isomorphically) equivalent to the set of LIVF's on G. Using property 24.32 we can show that the Lie bracket also defines a LIVF on G. It follows that $\mathfrak{g}$ is closed under Lie brackets.

Notation 22.2.5. Lie algebras are denoted by fraktur symbols. For example, the Lie algebra associated with the Lie group G is mostly denoted by $\mathfrak{g}$.

Theorem 22.2.6 (Ado's theorem). *Every finite-dimensional Lie algebra can be embedded as a subalgebra of $\mathfrak{gl}_n$.*

Definition 22.2.7 (Lie algebra homomorphism). A map $\Phi : \mathfrak{g} \to \mathfrak{h}$ is a Lie algebra homomorphism if it satisfies following condition

$$\Phi([X, Y]) = [\Phi(X), \Phi(Y)] \tag{22.4}$$

for all $X, Y \in \mathfrak{g}$.

Property 22.2.8. Let G, H be Lie groups with G simply-connected. A linear map $\Phi : \mathfrak{g} \to \mathfrak{h}$ is the differential of a Lie group homomorphism $\phi : G \to H$ if and only if Φ is a Lie algebra homomorphism.

22.2.2 Examples

Example 22.2.9. The cross product $\times : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ turns $\mathbb{R}^3$ into a Lie algebra.

Example 22.2.10. An interesting example is the Lie algebra associated to the Lie group of invertible complex[4] 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:

$$L_{A,*}(B) = L_A(B) \tag{22.5}$$

Following two examples of Lie algebras can be checked using condition 14.29:

Example 22.2.11 (Lie algebra of $O(3)$). The set of 3×3 anti-symmetric matrices. It is also important to note that $\mathfrak{o}(3) = \mathfrak{so}(3)$.

Example 22.2.12 (Lie algebra of $SU(2)$). The set of 2×2 traceless anti-Hermitian matrices. This result can be generalized to arbitrary $n \in \mathbb{N}$.

22.2.3 Exponential map

Formula 22.2.13 (Exponential map). Let $X \in \mathfrak{g}$ be a LIVF on G. We define the exponential map $\exp : \mathfrak{g} \to G$ as:

$$\boxed{\exp(X) := \gamma_X(1)} \tag{22.6}$$

where γ_X is the corresponding one-parameter subgroup.

Property 22.2.14. The exponential map is the unique map $\mathfrak{g} \to G$ such that $\exp(0) = e$, whose differential at the origin in $\mathfrak{g}$ is given by the identity $\mathbb{1}_\mathfrak{g}$ and for which the restrictions to the lines through the origin in $\mathfrak{g}$ are one-parameter subgroups of G.

Corollary 22.2.15. Because the identity element $\mathbb{1}_\mathfrak{g} = \exp_{e,*}$ is an isomorphism, the inverse function theorem 24.3.7 implies that the image of exp will contain a neighbourhood of the identity $e \in G$. If G is connected then property 22.1.5 implies that exp generates all of G.

Together with the property that $\psi \circ \exp = \exp \circ \psi_*$ for every Lie group homomorphism $\psi : G \to H$ it follows that *if G is connected, a Lie group homomorphism $\psi : G \to H$ is completely determined by its differential ψ_* at the identity $e \in G$.*

Example 22.2.16 (Matrix Lie groups). For matrix Lie groups we define the classic matrix exponential:

$$e^{tX} = \sum_{k=0}^{+\infty} \frac{(tX)^k}{k!} \tag{22.7}$$

This operation defines a curve $\gamma(t)$ which can be used as a one-parameter subgroup on G. It should be noted that this formula converges for every $X \in M_{m,n}$ and is invertible with the inverse given by $\exp(-X)$.

---

[4]As usual, this result is also valid for real matrices.

22.2.4   Structure

Definition 22.2.17 (Structure constants). As Lie algebras are closed under Lie brackets, every Lie bracket can be expanded in term of a chosen basis $\{X_k\}_{k \in I}$ as follows:

$$[X_i, X_j] = \sum_{k \in I} c_{ij}{}^k X_k \tag{22.8}$$

where the factors $c_{ij}{}^k$ are called the structure constants[5] of the Lie algebra.

An important use of these structure constants lies in the fact that 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 for all indices i, j and k.

Formula 22.2.18 (Baker-Campbell-Hausdorff formula). This formula is the solution of the equation

$$Z = \log(\exp(X)\exp(X)) \tag{22.9}$$

for $X, Y \in \mathfrak{g}$. The solution is given by following formula

$$e^X e^Y = \exp\left(X + Y + \frac{1}{2}[X,Y] + \frac{1}{12}[X,[X,Y]] - \frac{1}{12}[Y,[X,Y]] + \cdot\right) \tag{22.10}$$

However this formula will only converge if X, Y are sufficiently small (for matrix Lie algebras this means, using the Hilbert-Schmidt norm 14.40: $||X|| + ||Y|| < \frac{\ln(2)}{2}$). Due to the closure under commutators (see Lie algebra definition) the exponent in the BCH formula is also an element of the Lie algebra. So the formula gives an expression for Lie group multiplication in terms of Lie algebra elements (whenever the formula converges).

Corollary 22.2.19 (Lie product formula[6]). Let $\mathfrak{g}$ be a Lie algebra. Consider $X, Y \in \mathfrak{g}$. The following formula applies to any such X and Y:

$$e^{X+Y} = \lim_{n \to +\infty} \left(e^{\frac{X}{n}} e^{\frac{Y}{n}}\right)^n \tag{22.11}$$

22.2.5   Killing form

Definition 22.2.20 (Killing form). Let $\mathfrak{g}$ be a finite-dimensional Lie algebra. Define the symmetric bilinear form[7]

$$\boxed{K(X,Y) = \mathrm{tr}(\mathrm{ad}_X \mathrm{ad}_Y)} \tag{22.12}$$

Theorem 22.2.21 (Cartan's criterion). *A Lie algebra is semisimple if and only if its Killing form is non-degenerate.*

---

[5]Note that these constants are basis-dependent.
[6]Also called the Lie-Trotter formula.
[7]This is a symmetric $(0,2)$-tensor in $\mathfrak{g}^* \otimes \mathfrak{g}^*$. (See 18.2.8)

Corollary 22.2.22. If a Lie algebra is semisimple its Killing form induces a metric

$$g : (X, Y) \mapsto -\operatorname{tr}(\operatorname{Ad}_X, \operatorname{Ad}_Y) \tag{22.13}$$

which turns the corresponding Lie group G into a Riemannian manifold.

Property 22.2.23. The Killing-form is Ad-invariant, i.e.

$$K(\operatorname{Ad}_g(X), \operatorname{Ad}_g(Y)) = K(X, Y) \tag{22.14}$$

for all $g \in G$. From this it follows that $\operatorname{Ad} : G \to \operatorname{Isom}(\mathfrak{g})$.

22.2.6 Universal enveloping algebra

Definition 22.2.24 (Universal enveloping algebra). Let $\mathfrak{g}$ be a Lie algebra with Lie bracket $[\cdot, \cdot]$. First construct the tensor algebra $T(\mathfrak{g})$. The universal enveloping algebra $U(\mathfrak{g})$ is defined as quotient of $T(\mathfrak{g})$ by the two-sided ideal generated by the elements $g \otimes h - h \otimes g - [g, h]$.

Definition 22.2.25 (Casimir invariant[8]). Let $\mathfrak{g}$ be a Lie algebra. A Casimir invariant J is an element of the center of $U(\mathfrak{g})$, the universal enveloping algebra of $\mathfrak{g}$.

22.2.7 Poisson algebras and Lie superalgebras

Definition 22.2.26 (Poisson algebra). Let V be a vector space equipped with two bilinear operation $\star$ and $\{\cdot, \cdot\}$ that satisfy the following conditions:

- The couple $(V, \star)$ is an associative algebra.

- The couple $(V, \{\cdot, \cdot\})$ is a Lie algebra.

- the **Poisson bracket** acts as a derivation[9] with respect to the operation $\star$, i.e.

$$\{x, y \star z\} = \{x, y\} \star z + y \star \{x, z\}$$

Definition 22.2.27 (Gerstenhaber algebra).

---

[8]Also known as a *Casimir operator* or *Casimir element*.
[9]See definition 21.2.1.

Chapter 23

Representation Theory

23.1 Group representations

Definition 23.1.1 (Representation). A representation of a group G, acting on a vector space V, is a homomorphism $\rho : G \to GL(V)$ from G itself to the automorphism group[1] of V. This is a specific case of a group action[2].

Definition 23.1.2 (Subrepresentation). A subrepresentation of a representation V is a subspace of V invariant under the action of the group G.

Example 23.1.3 (Permutation representation). Consider a vector space V equipped with a basis $\{e_i\}_{i \in I}$ with $|I| = n$. Let $G = S^n$ be the symmetric group of dimension n. Based on remark 2.4.32 we can consider the action of G on the index set I. This representation is given by

$$\rho(g) : \sum_{i \in I} v_i e_i \mapsto \sum_{i \in I} v_i e_{g \cdot i} \tag{23.1}$$

Example 23.1.4. Consider a representation ρ on V. There exists a natural representation on the dual space V^*. The homomorphism $\rho^* : G \to GL(V^*)$ is given by:

$$\rho^*(g) = \rho^T(g^{-1}) : V^* \to V^* \tag{23.2}$$

where ρ^T is the transpose as defined in 14.3.38. This map satifies the following defining property:

$$\left\langle \rho^*(g)(v^*), \rho(g)(v) \right\rangle = \langle v^*, v \rangle \tag{23.3}$$

where $\langle \cdot, \cdot \rangle$ is the natural pairing of V and its dual.

Example 23.1.5. A representation ρ which acts on spaces V, W can also be extended to the tensor product $V \otimes W$ in the following way:

$$g(v \otimes w) = g(v) \otimes g(w) \tag{23.4}$$

---

[1]See definition 14.3.11.
[2]See definition 2.4.31.

23.2   Irreducible representations

Definition 23.2.1 (Irreducibility). A representation is said to be irreducible if there exist no proper non-zero subrepresentation.

Example 23.2.2 (Standard representation). Consider the action of S^n on a vector space V. The line generated by $v_1 + v_2 + ... + v_n$ 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

$$W = \{a_1 v_1 + a_2 v_2 + ... + a_n v_n | a_1 + a_2 + ... + a_n = 0\} \tag{23.5}$$

does form an irreducible representation when we restrict ρ to W. It is called the standard representation of S^n.

Theorem 23.2.3 (Schur's lemma). *Let V, W be two irreducible representations of a finite group G. Let $\varphi : V \to W$ be a G-module homomorphism. We then have:*

- *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 23.2.4. If W is a subrepresentation of V then there exists an invariant complementary subspace W' such that $V = W \oplus W'$.

This space can be found as follows: Choose an arbitrary complement U such that $V = W \oplus U$. From this we construct a projection map $\pi_0 : V \to W$. Averaging over G gives

$$\pi(v) = \sum_{g \in G} g \circ \pi_0(g^{-1}v) \tag{23.6}$$

which is a G-linear map $V \to W$. On W it is given by the multiplication of W by $|G|$. Its kernel is then an invariant subspace of V under the action of G and complementary to W.

Property 23.2.5. Let G be a finite group. A representation V can be uniquely decomposed as

$$V = V_1^{\oplus a_1} \oplus \cdots \oplus V_k^{\oplus a_k} \tag{23.7}$$

where all V_k's are distinct irreducible representations.

23.3   Lie group representations

For more information on Lie groups (and Lie algebras which will be considered in the next section) see chapter 22.

Definition 23.3.1 (Adjoint representation). Let G be a Lie group. Consider the conjugation map $\Psi_g : h \mapsto ghg^{-1}$. The adjoint representation of G is defined by its differential:

$$\mathrm{Ad}_g := T_e \Psi_g : T_e G \to T_e G : X \mapsto gXg^{-1} \tag{23.8}$$

It is a representation of G on its own tangent space $T_e G \equiv \mathfrak{g}$.

23.4 Lie algebra representations

Formula 23.4.1 (Adjoint representation on Lie algebras). Using the fact that the adjoint representation of Lie groups[3] is smooth we can define the adjoint representation of Lie algebras as:

$$\mathrm{ad}_g := T_e(\mathrm{Ad}_g) \tag{23.9}$$

Explicitly, let $\mathfrak{g}$ be a Lie algebra. For every element $X \in \mathfrak{g}$ we define the Lie bracket as follows:

$$[X, Y] := \mathrm{ad}_X(Y) \tag{23.10}$$

Property 23.4.2. Given the antisymmetry of the Lie bracket the Jacobi identity is equivalent to ad: $\mathfrak{g} \to \mathrm{Aut}(\mathfrak{g})$ being a Lie algebra homomorphism, i.e. $\mathrm{ad}_{[X,Y]} = [\mathrm{ad}_X, \mathrm{ad}_Y]$.

Using the exponential map we can give property 22.2.8 an explicit form:

Formula 23.4.3 (Induced homomorphism). Let $\phi : G \to H$ be a Lie group homomorphism[4] with G connected and simply-connected. This homomorphism induces a a Lie algebra homomorphism $\Phi : \mathfrak{g} \to \mathfrak{h}$ given by

$$\Phi(X) = \frac{d}{dt}\phi\left(e^{tX}\right)\bigg|_{t=0} \tag{23.11}$$

or equivalently

$$\phi\left(e^{tX}\right) = e^{t\Phi(X)} \tag{23.12}$$

Example 23.4.4. The homomorphism induced by $\mathrm{Ad} : G \to H$ is precisely $\mathrm{ad} : \mathfrak{g} \to \mathfrak{h}$. Informally we can thus say that the infinitesimal version of the similarity transformation is given by the commutator (in case of $G = \mathrm{GL}_n$).

---

[3]See definition 23.3.1.
[4]Continuity is needed to ensure that $\phi(e^{tX})$ is also a one-parameter subgroup (see 22.1.10). However, this condition is always satisfied as it is inherent to the definition of a Lie group homomorphism.

Chapter 24

Bundle theory

24.1 Fibre bundles

Definition 24.1.1 (Fibered manifold). A fibered manifold is a surjective submersion[1] $\pi : E \to B$ where E is called the **total space** and B the **base space**.

The most important example of a fibered manifold is a fibre bundle:

Definition 24.1.2 (Fibre bundle). A fibre bundle is a tuple (E, B, π, F, G) where E, B and F are topological spaces and G is a topological group (called the **structure group**), such that there exists a smooth surjective map $\pi : E \to B$ and an open cover $\{U_i\}_{i \in I}$ of B for which there exists a family of homeomorphisms $\{\varphi_i : \pi^{-1}(U_i) \to U_i \times F\}_{i \in I}$ that make the following diagram commute:

$$
\begin{array}{ccc}
\pi^{-1}(U_i) & \xrightarrow{\quad \varphi_i \quad} & U_i \times F \\
& \searrow{\scriptstyle \pi} \quad \swarrow{\scriptstyle \mathrm{pr}_1} & \\
& U_i &
\end{array}
$$

We call E and B the **total space** and **base space** respectively, F the **(typical) fibre**, φ_i a **local trivialization**[2], (U_i, φ_i) a **bundle chart**[3] and the set $\{(U_i, \varphi_i)\}_{i \in I}$ a **trivializing cover**.

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 the cocycle[4] $g_{ji} : U_i \cap U_j \to G$, associated to the (left) action, which we require to be faithful[5], of

---

[1]See definition 21.4.2.
[2]This name follows from the fact that locally we have $E \cong U \times \mathbb{R}^n$, which is the definition of a trivial bundle (see 24.1.11).
[3]This is due to the similarities with the charts as defined for manifolds.
[4]See definition 22.1.12.
[5]See definition 2.4.37.

G on every fibre, by the following relation:

$$\varphi_j \circ \varphi_i^{-1}(b, x) = (b, g_{ji}(b) \cdot x) \tag{24.1}$$

Remark 24.1.3. One should pay attention that the bundle charts are not coordinate charts in the original sense 21.1.1 because the image of φ_i is not an open subset of $\mathbb{R}^n$. However they serve the same purpose and we can still use them to locally inspect the total space P.

Notation 24.1.4. A fibre bundle (E, B, π, F, G) is often indicated by the following diagram:

$$
\begin{array}{ccc}
F & \longhookrightarrow & E \\
& & \downarrow{\scriptstyle \pi} \\
& & B
\end{array}
$$

or more compactly $F \hookrightarrow E \xrightarrow{\pi} B$. A drawback of these notations is that we do not immediately know what the structure group of the bundle is.

Definition 24.1.5 (Fibre). Let $F \hookrightarrow E \xrightarrow{\pi} B$ be a fibre bundle over a base space B. The fibre over $b \in B$ is defined as the set $\pi^{-1}(b)$.

Definition 24.1.6 (Smooth fibre bundle). A smooth fibre bundle is a fibre bundle (E, B, π, F, G) with the following constraints:

- The base space B and typical fibre F are smooth manifolds.

- The structure goup G is a Lie group.

- The projection map, trivializing maps and transition functions are diffeomorphisms.

Remark 24.1.7. A smooth fibre bundle is also a smooth manifold.

Definition 24.1.8 (Compatible[6] bundle charts). A bundle chart (U, φ) is compatible with a trivializing cover $\{(U_i, \varphi_i)\}_{i \in I}$ if whenever $U \cap U_i \neq \emptyset$ their exists a map $h_i : U \cap U_i \to G$ such that:

$$\varphi \circ \varphi_i^{-1}(b, x) = (b, h_i(b)x) \tag{24.2}$$

for all $b \in U \cap U_i$ and $x \in F$. Two trivializing covers are *equivalent* if all bundle charts are cross-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 eqipped with an equivalence class of G-atlases.

Definition 24.1.9 (Equivalent fibre bundles). Two fibre bundles $\pi_1 : F_1 \to B$ and $\pi_2 : F_2 \to B$ (over the same base space B) are equivalent if there exist trivializing covers[7] $\{(U_i, \varphi_i)\}_{i \in I}$ and $\{(U_i, \varphi_i')\}_{i \in I}$ and a family of smooth functions $\{\rho_i : U_i \to G\}_{i \in I}$ such that:

$$g_{ji}'(b) = \rho_j(b) \circ g_{ji}(b) \circ \rho_i^{-1}(b) \tag{24.3}$$

for every $b \in U_i \cap U_j$.

---

[6]Also called an **admissible chart**.

[7]Remark that the collection $\{U_i\}_{i \in I}$ is the same for both trivializing covers.

Property 24.1.10. Two fibre bundles over the same base space are equivalent if and only if they are isomorphic[8]. Furthermore, if there exists a bundle map between two fibre bundles over the same base space, then they are equivalent.

Definition 24.1.11 (Trivial bundle). A fibre bundle (E, B, π, F) is trivial if $E = B \times F$.

Definition 24.1.12 (Trivialization). A trivialization of a fibre bundle ξ is an equivalence $\xi \to B \times F$. Bundles for which a trivialization can be found are also called *trivial bundles*.

Definition 24.1.13 (Fibre product). Let (F_1, B, π_1) and (F_2, B, π_2) be two fibre bundles on a base space B. Their fibre product is defined as:

$$F_1 \diamond F_2 = \{f \times g \in F_1 \times F_2 : \pi_1(f) = \pi_2(g)\} \tag{24.4}$$

24.1.1   Sections

Definition 24.1.14 (Section). A **global** section on a fibre bundle $\pi : E \to B$ is a smooth function $s : B \to E$ such that $\pi \circ s = \mathbb{1}_B$. For any open subset $U \subset B$ we define a local section as a smooth function $s_U : U \to E$ such that $\pi \circ s_U(b) = b$ for all $b \in U$.

Notation 24.1.15. The set of all global sections on a bundle E is denoted by $\Gamma(E)$. The set of local sections on $U \subset E$ is similarly denoted by $\Gamma(U)$.

24.1.2   Jet bundles

Definition 24.1.16 (Jet). Consider a fibre bundle (E, B, π) with its sections $\Gamma(E)$. Two sections $\sigma, \xi \in \Gamma(E)$ with local coordinates (σ^i) and (ξ^i) define the same r-jet at a point $p \in B$ if and only if:

$$\left. \frac{\partial^\alpha \sigma^i}{\partial x^\alpha} \right|_p = \left. \frac{\partial^\alpha \xi^i}{\partial x^\alpha} \right|_p \tag{24.5}$$

for all $0 \leq i \leq \dim E$ and every multi-index α such that $0 \leq |\alpha| \leq r$. It is clear that this relation defines an equivalence relation. The r-jet at $p \in B$ with representative σ is denoted by $j_p^r \sigma$. The number r is called the **order** of the jet.

Definition 24.1.17 (Jet manifold). Consider a fibre bundle (E, B, π). The r-jet manifold $J^r(\pi)$ of the projection π is defined as:

$$J^r(\pi) = \{j_p^r \sigma : \sigma \in \Gamma(E), p \in B\} \tag{24.6}$$

The set $J^0(\pi)$ is identified with the total space E.

---

[8]Two fibre bundles F and G are isomorphic if there exist bundle maps $f : F \to G$ and $g : G \to F$ such that $f \circ g = \mathbb{1}_G$ and $g \circ f = \mathbb{1}_F$.

Definition 24.1.18 (Jet projections). Let (E, B, π) be a fibre bundle with r-jet manifolds $J^r(\pi)$. The **source projection** π_r and **target projection** $\pi_{r,0}$ are defined as the maps

$$\pi_r : J^r(\pi) \to B : j_p^r \sigma \mapsto p \tag{24.7}$$

$$\pi_r : J^r(\pi) \to E : j_p^r \sigma \mapsto \sigma(p) \tag{24.8}$$

These projections satisfy $\pi_r = \pi \circ \pi_{r,0}$. We can also define a k-**jet projection** $\pi_{r,k}$ as the map

$$\pi_{r,k} : J^r(\pi) \to J^k(\pi) : j_p^r \sigma \mapsto j_p^k \sigma \tag{24.9}$$

where $k \leq r$. The k-jet projections satisfy a transitivity property $j_{k,m} = j_{r,m} \circ j_{k,r}$.

Definition 24.1.19 (Jet prolongation). Let σ be a section on a fibre bundle (E, B, π). The r-jet prolongation $j^r \sigma$ corresponding to σ is defined as the following map:

$$j^r \sigma : B \to J^r(\pi) : p \mapsto j_p^r \sigma \tag{24.10}$$

Definition 24.1.20 (Jet bundle). The r-jet bundle corresponding to the projection π is then defined as the triple $(J^r(\pi), B, \pi_r)$. The bundle charts $(U_i, \varphi_i)^9$ on E define *induced* bundle charts on $J^r(\pi)$ in the following way:

$$U_i^r = \{j_p^r \sigma : \sigma(p) \in U_i\} \tag{24.11}$$

$$\varphi_i^r = \left(x^k, u^\alpha, \left.\frac{\partial^I u^\alpha}{\partial x^I}\right|_p \right) \tag{24.12}$$

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 $J^r(\pi)$.

24.2 Principal bundles

Definition 24.2.1 (Principal bundle). A principal bundle is a fibre bundle (E, B, π, G, G) such that the structure group and the typical fibre are the same, i.e. we identify the structure group with the group of left translations of G.

Remark 24.2.2. We remark that 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**. However it is possible to locally (i.e. in a neighbourhood of a point $p \in M$), but not globally, endow the fibres with a group structure by choosing an element of every fibre to be identity element.

Property 24.2.3. The dimension of P is given by:

$$\dim P = \dim M + \dim G \tag{24.13}$$

---

[9]Where $\varphi_i = (x^k, u^\alpha)$ with x^k the base space coordinates and u^α the total space coordinates.

Property 24.2.4. Let $\pi : P \to B$ be a principal G-bundle with local trivializations $\{(U_i, \varphi_i)\}_{i \in I}$. There exists a (faithful) right action of G on P given by:

$$z \cdot g = \varphi_i^{-1}(b, hg) \tag{24.14}$$

for all $g, h \in G$ and $z \in \pi(U_i)$. This action preserves fibres ($y \cdot g \in F_b$ for all $y \in F_b, g \in G$). Furthermore, it is free[10] and it is transitive. It follows that the fibres over B are exactly the orbits of the right action on P.

Every local trivialization φ_i is also G-equivariant:

$$\varphi_i(z \cdot g) = \varphi_i(z) \cdot g \tag{24.15}$$

Definition 24.2.5 (Bundle map). A bundle map $F : P_1 \to P_2$ between principal G-bundles is a pair of smooth maps (f_B, f_E) such that:

1. The diagram 24.1 below commutes.

2. f_E is G-equivariant[11].

The map f_E is said to **cover** f_B.

$$
\begin{array}{ccc}
E_1 & \xrightarrow{\ f_E\ } & E_2 \\
\downarrow{\scriptstyle \pi_1} & & \downarrow{\scriptstyle \pi_2} \\
B_1 & \xrightarrow[\ f_B\ ]{} & B_2
\end{array}
$$

Figure 24.1: Bundle map between principal G-bundles.

Example 24.2.6 (Associated principal bundle). For every fibre bundle (E, B, π, F, G) we can construct an associated principal G-bundle by replacing the fibre F by G itself.

Property 24.2.7. A fibre bundle ξ is trivial if and only if the associated principal bundle is trivial. More generally, two fibre bundles are isomorphic if and only if their associated principal bundles are isomorphic.

Example 24.2.8 (Frame bundle). Let V be an n-dimensional vector space. Denote the set of ordered bases, also called **frames**, of V by $F(V)$. This set is isomorphic to the group $GL(\mathbb{R}^n)$ which follows from the fact that every basis transformation is given by the action of an element of the general linear group. We can thus construct a principal bundle associated to the vector bundle E by replacing every fibre $\pi^{-1}(b)$ by $F(\pi^{-1}(b)) \cong \mathrm{GL}(\mathbb{R}^n)$.

---

[10]See definition 2.4.36.
[11]See definition 2.4.42.

24.2.1 Sections

Where every vector bundle has at least one global section, the **zero section**[12], a general principal bundle does not necessarily have a global section. This is made clear by the following property:

Property 24.2.9. A principal G-bundle P is trivial if and only if there exists a global section on P. Furthermore, there exists a bijection between the set of all global sections $\Gamma(P)$ and the set of trivializations $\mathrm{Triv}(P)$.

24.3 Vector bundles

The tangent space and tangent bundle, as introduced in subsection 21.2, can also be introduced in a more topological way:

24.3.1 Tangent bundles

Construction 24.3.1 (Tangent bundle). Let M be a manifold with atlas $\{(U_i, \varphi_i)\}_{i \leq n}$. Consider for every open set U an associated set $TU = U \times \mathbb{R}^n$. For every smooth function f we can define an associated smooth function on TU, called the differential of f, by:

$$Tf : U \times \mathbb{R}^n \to f(U) \times \mathbb{R}^n : (x, v) \mapsto (f(x), Df(x)v) \tag{24.16}$$

where $Df(x) : \mathbb{R}^n \to \mathbb{R}^n$ is the linear operator associated with the Jacobian matrix of f in x. Applying this definition to the transition functions ψ_{ji} we obtain a new set of functions $\widetilde{\psi}_{ji} := T\psi_{ji}$ given by:

$$\widetilde{\psi}_{ji}(\varphi_i(x), v) = \left(\varphi_j(x), (\varphi_j \circ \varphi_i^{-1})'(\varphi_i(x))v\right) \tag{24.17}$$

Now we take the disjoint union of all the sets $T(\varphi_i(U_i))$ (where we implicitly consider the restriction to $U_i \cap U_j$), equipped with the disjoint union topology (see 4.1.3). From this set, we construct a quotient space (see 4.4) by applying following equivalence relation $\approx$:

$$x \approx y \iff y = \widetilde{\psi}_{ji}(x) \tag{24.18}$$

This construction gives the tangent bundle:

$$TM = \left(\bigsqcup_{i \leq n} TU_i\right) / \approx \tag{24.19}$$

---

[12]This is the map $s : b \to \vec{\mathbf{0}}$ for all $b \in B$.

We can define an atlas on TM by remarking that the map $x \mapsto [x]$ is a homeomorphism[13] (and even a diffeomorphism). So we can use the image $\widetilde{U}_i$ of this map together with its inverse $\widetilde{\varphi}_i : TM \to \varphi_i(U_i) \times \mathbb{R}^n$ to construct a smooth atlas[14] $\{(\widetilde{U}_i, \widetilde{\varphi}_i)_{i \le n}\}$ with transition functions given by $\widetilde{\psi}_{ji}$.

Finally, we also define a projection map $\pi : TM \to M$ which is induced by the local projection $\varphi_i(U_i) \times \mathbb{R}^n \mapsto \varphi_i(U_i)$. This map is called the **tangent bundle map** and together with the *total space* TM we get the tangent bundle on M. Using the natural charts we can also see that TM is isomorphic to $\mathbb{R}^{2n}$ and thus has twice the dimension of M.

Definition 24.3.2 (Tangent space). Let $x \in M$. The topological definition of the tangent space is given by the fibre

$$T_x M := \tau_M^{-1}(x) \tag{24.20}$$

If we use the natural charts to map $T_x M$ to the set $\varphi_i(x) \times \mathbb{R}^n$, we see that $T_x M$ is isomorphic to $\mathbb{R}^n$ and thus also to M itself. Furthermore, we can equip every fibre with the following vector space structure:

$$(x, v_1) + (x, v_2) := (x, v_1 + v_2)$$
$$r(x, v) := (x, rv)$$

Remark 24.3.3. Now it is clear that the rule "*a vector is something that transforms like a vector*" stems from the fact that:

$$\text{a vector } v \in T_x M \text{ is tangent to } \varphi_i(x) \text{ in a chart } (U_i, \varphi_i)$$

if and only if

$$D(\varphi_j \circ \varphi_i^{-1})(\varphi_i(x))v \text{ is tangent to } \varphi_j(x) \text{ in a chart } (U_j, \varphi_j)$$

Comparing this property to 21.5, we see that tangent vectors defined through equivalence classes of tangent curves are indeed tangent vectors according to our new construction.

Definition 24.3.4 (Differential). The map T from 24.16 can be generalized to arbitrary smooth manifolds as the map $Tf : TM \to TN$. Furthermore, let $x \in U \subseteq M$ and let $V = f(U)$. By looking at the restriction of Tf to $T_x M$, denoted by $T_x f$, we see that it maps $T_x U$ to $T_{f(x)} V$ (where $V = f(U)$) linearly. So $T_x f$ is a linear map on fibres.

Property 24.3.5. The map $Tf : TM \to TN$ (see 24.16) has following properties[15]:

- $T(\mathbb{1}_M) = \mathbb{1}_{TM}$

- Let f, g be two smooth functions on smooth manifolds. Then $T(f \circ g) = Tf \circ Tg$.

---

[13]This follows from the fact that every $x \in TU_i$ is only equivalent to one point in every other set TU_j, because $\widetilde{\psi}_{ii} = \mathbb{1}, \forall i$.

[14]The charts in this atlas are called **natural charts**.

[15]This turns the map T into a functor on the category of smooth manifolds. We can view T as a functorial derivative.

Remark. We can also use a construction similar to that of the tangent bundle to reconstruct the original manifold M from the sets $\varphi_i(U_i)$.

Definition 24.3.6 (Rank). Let $f : M \to N$ be a differentiable map between smooth manifolds. Using the fact that Tf is a linear map of fibres[16], we define the rank of f at $p \in M$ as the rank (as in 14.3.16) of the differential $Tf : T_p M \to T_{f(p)} N$.

Theorem 24.3.7 (Inverse function theorem). *A C^∞ map $f : M \to N$ between smooth manifolds is locally homeomorphic (resp. locally diffeomorphic) if and only if its differential $Tf : T_p M \to T_p N$ is an isomorphism (resp. diffeomorphism) at p.*

24.3.2 Vector bundles

Instead of restricting ourselves by letting the typical fibre be a Euclidean space with the same dimension as the base manifold, we can generalize the construction of the tangent bundle in the following way:

Construction 24.3.8 (Vector bundle). Consider a smooth n-dimensional manifold M with atlas $\{(U_i, \varphi_i)\}_{i \leq n}$, a cocycle $\{g_{ji} : U_i \cap U_j \to G\}_{i,j \leq n}$ with values in a Lie group G and a smooth representation $\rho : G \to GL(V)$, where V is a vector space.

Now we can construct a new topological space E, similar to the construction of the tangent bundle, by taking the disjoint union of the sets $\varphi_i(U_i) \times V$ and quotienting out using the functions $\widetilde{g}_{ji} : (\varphi_i(x), v) \mapsto (\varphi_j(x), g_{ji}(x) \cdot v)$, where $g \cdot v \equiv \rho(g)v$. This gives us a set of natural charts[17] $\{(\widetilde{U}_i, \widetilde{\varphi}_i)\}_{i \leq n}$, a projection map $\pi : E \to M$ induced by the local projection $\varphi_i(U_i) \times V \to \varphi_i(U_i)$ and a naturally defined vector space on every fibre $V_x := \pi^{-1}(x)$. Furthermore every fibre V_x is (although not necessarilly canonically) isomorphic to V.

This set E is called a **smooth vector bundle** over M with *typical fibre V* and *projection map π*.

Remark 24.3.9. As is also the case for tangent bundles (which are specific cases of vector bundles where the typical fibre has the same dimension as the manifold) the choice of charts on E is not random. To preserve the structure of fibres, the use of the natural charts is imperative.

Remark 24.3.10. Vector bundles are smooth fibre bundles where the typical fibre is a vector space V and the structure group is given by $GL(V)$.

Definition 24.3.11 (Associated vector bundle). Consider a representation $\rho : GL(\mathbb{R}^n) \to GL(\mathbb{R}^l)$ and the cocycle $t_{ji} := D(\psi_{ji}) \circ \varphi_i$ as defined for tangent bundles. The composition $\rho \circ t_{ji} : U_i \cap U_j \xrightarrow{t_{ji}} GL(\mathbb{R}^n) \xrightarrow{\rho} GL(\mathbb{R}^l)$ is again a cocycle and can thus be used to define a new vector bundle on M. The vector bundle $E = \rho(TM)$ so obtained is called the associated bundle of the tangent bundle induced by ρ.

---

[16]See definition 24.3.4.

[17]We could instead use any other kind of topological space. The point is that a vector bundle is a fibre bundle 24.1.2 for which the typical fibres are vector spaces.

Remark 24.3.12. It should also be noted that every vector bundle is associated to a principle $GL(V)$-bundle where the cocycles g_{ji} now act by left multiplication on elements of $GL(V)$.

Example 24.3.13 (Contravariant vectors). By noting that the kth tensor power $\otimes^k$ induces a representation given by the tensor product of the representations, we can construct the bundle of kth order contravariant vectors $\otimes^k(TM)$ with the cocycle given by $x \mapsto t_{ji}(x) \otimes \cdots \otimes t_{ji}(x)$.

Example 24.3.14 (Cotangent bundle). Another (smooth) representation is given by $A \mapsto (A^*)^{-1} = (A^{-1})^*$ for every linear map A. The vector bundle constructed this way, where the cocycle is given by $(t_{ji}^*)^{-1}$, is called the cotangent bundle on M and is denoted by T^*M. Elements of the fibres are called covariant vectors or covectors.

Notation 24.3.15. A combination of the cocycle t_{ji} and its dual $(t_{ji}^*)^{-1}$ can also be used to define the bundle of kth order contravariant and l^{th} order covariant vectors on M. This bundle is denoted by $T^{(k,l)}M$.

Example 24.3.16 (Pseudovectors). If we consider the representation

$$\rho : A \mapsto \operatorname{sgn} \det(A) A \tag{24.21}$$

we can construct a bundle similar to the tangent bundle. The sign of the cocycle functions t_{ji} now has an influence on the fibres. Elements of these fibres are called **pseudovectors**.

Definition 24.3.17 (Subbundle). A subbundle of a vector bundle $\pi : E \to M$ is a collection of subspaces U_x of fibres E_x that make up a vector bundle on their own.

Definition 24.3.18 (Whitney sum). Consider two vector bundles W, W' with fibres E, E' respectively. Then we can construct a new vector bundle $W \oplus W'$ by defining the new typical fibre to be the direct sum $E \oplus E'$, i.e. the fibre above b is given by $E_b \oplus E_b'$. This operation is called the Whitney sum or direct sum of vector bundles.

24.3.3   Sections

Remark. Vector fields can be regarded as sections of the tangent bundle. Similarly, 1-forms can be regarded as sections of the cotangent bundle.

Definition 24.3.19 (Frame). A frame of a vector bundle E is a tuple $(s_1, ..., s_n)$ of smooth sections such that $(s_1(b), ..., s_n(b))$ is a basis of the fibre $\pi^{-1}(b)$ for all $b \in B$.

Property 24.3.20. A vector bundle is trivial if and only if there exists a frame of global sections.

24.4   Vector fields

Definition 24.4.1 (Vector field). A smooth section $s \in \Gamma(TM)$ of the tangent bundle is called a vector field.

Notation 24.4.2. The set of all vector fields on a manifold M is often denoted by $\mathfrak{X}(M)$.

Definition 24.4.3 (Pullback). Let X be vector field on M and let $\varphi : M \to N$ be a diffeomorphism between smooth manifolds. The pullback of X along φ is defined as:

$$(\varphi^* X)_p = T\varphi^{-1}(X_{\varphi(p)}) \tag{24.22}$$

Definition 24.4.4 (Pushforward). Let $X \in \mathfrak{X}(M)$ and let $\varphi : M \to N$ be a diffeomorphism between smooth manifolds. Using the differential $T\varphi$ we can define the pushforward of X along φ as:

$$(\varphi_* X)_{\varphi(p)} = T\varphi(X_p) \tag{24.23}$$

which we can rewrite using the pullback as:

$$\varphi_* X = \varphi^{-1*} X \tag{24.24}$$

Or equivalently we can define a vector field on N by:

$$(\varphi_* X)_q(f) = X_{\varphi^{-1}(q)}(f \circ \varphi) \tag{24.25}$$

for all smooth functions $f : N \to \mathbb{R}$ and points $q \in N$.

Remark. For both the pullback and pushforward, we need the map φ to be a diffeomorphism. For differential forms this is only necessary for the definition of pushforwards. (See definitions 24.37 and 24.38).

24.4.1 Integral curves

Definition 24.4.5 (Integral curve). Let $X \in \mathfrak{X}(M)$ and let $\gamma :]a, b[\to M$ be a smooth curve on M. γ is said to be an integral curve of X if:

$$\boxed{\gamma'(t) = X(\gamma(t))} \tag{24.26}$$

for all $t \in]a, b[$ where we defined $\gamma'(t) := T\gamma(t, 1)$ using the functorial derivative 24.16.

This equation can be seen as a system of ordinary differential equations in the second argument. Using Picard's existence theorem[18] together with the initial value condition $\gamma(0) = p$ we can find a unique curve on $]a, b[$ satisfying the defining equation 24.26. Furthermore we can extend the interval $]a, b[$ to a maximal interval such that the solution is still unique. This solution, denoted by γ_p, is called the **integral curve of X through** p.

Definition 24.4.6 (Flow). Let $X \in \mathfrak{X}(M)$. The function σ_t:

$$\sigma_t(p) = \gamma_p(t) \tag{24.27}$$

is called the flow of X at time t. The flow domain is defined as the set $D(X) = \{(t, p) \in \mathbb{R} \times M \mid t \in]a_p, b_p[\}$ where $]a_p, b_p[$ is the maximal interval on which $\gamma_p(t)$ is defined.

---

[18] Also Picard-Lindelöf theorem.

Property 24.4.7. Suppose that $D(X) = \mathbb{R} \times M$. The flow σ_t has following properties for all $s, t \in \mathbb{R}$:

- $\sigma_0 = \mathbb{1}_M$

- $\sigma_{s+t} = \sigma_s \circ \sigma_t$

- $\sigma_{-t} = (\sigma_t)^{-1}$

These three properties[19] say that σ_t is a bijective group action from M to the additive group of real numbers. This implies that σ_t is indeed a **flow** in the general mathematical sense.

Definition 24.4.8 (Complete vector field). A vector field X is called complete if the flow domain for every flow is whole $\mathbb{R}$.

Property 24.4.9. The flow σ_t of a vector field is of class C^∞. If X is complete it follows from previous definition that the flow is a diffeomorphism from M onto itself.

24.4.2 Lie derivative

Formula 24.4.10 (Lie derivative for smooth functions). Let $X \in \mathfrak{X}(M)$ and let $f : M \to \mathbb{R}$ be a smooth function. The Lie derivative of f with respect to X at $p \in M$ is defined as:

$$(\mathcal{L}_X f)(p) = \lim_{t \to 0} \frac{f(\gamma_p(t)) - f(p)}{t} \tag{24.28}$$

which closely resembles the standard derivative in Euclidean space.

Formula 24.4.11 (†). Working out previous formula and rewriting it as an operator equality gives:

$$\mathcal{L}_X = \sum_k X_k \frac{\partial}{\partial x^k} \tag{24.29}$$

It is clear that this is just the vector field X expanded in the basis 21.3. We also recover the behaviour of a tangent vector as a derivation. So for smooth functions $f : M \to \mathbb{R}$ we obtain:

$$\mathcal{L}_X f(p) = X_p(f) \tag{24.30}$$

Formula 24.4.12 (Lie derivative for vector fields†). Let $X, Y \in \mathfrak{X}(M)$

$$\mathcal{L}_X Y = \frac{d}{dt}(\sigma_t^* X)(\gamma_p(t)) \Big|_{t=0} \tag{24.31}$$

Property 24.4.13. Let $X, Y \in \mathfrak{X}(M)$ be vector fields of class C^k. The Lie derivative has following properties:

- $\mathcal{L}_X Y$ is a vector field.

---

[19]The third property follows from the other two.

- **Lie bracket**:

$$\mathcal{L}_X Y = [X, Y] \tag{24.32}$$

which is also a derivation on $C^{k-1}(M, \mathbb{R})$ due to the cancellation of second-order derivatives in the local representation.

- The Lie derivative is antisymmetric:

$$\mathcal{L}_X Y = -\mathcal{L}_Y X \tag{24.33}$$

This follows from the previous property.

24.4.3 Frobenius theorem

Looking at property 14.2.10 and noting that $\mathrm{GL}_n(\mathbb{C})$ is a Lie group, we can endow the Grassmannian $\mathrm{Gr}(k, \mathbb{R}^n)$ 14.2.9 with a differentiable structure, turning it into a smooth manifold. With this we can construct a new bundle[20] by applying the usual 'gluing' process:

Construction 24.4.14 (Grassman bundle). First define the transition functions:

$$\psi_{ji} : \varphi_i(U_i \cap U_j) \times \mathrm{Gr}(k, \mathbb{R}^n) \to \varphi_j(U_i \cap U_j) \times \mathrm{Gr}(k, \mathbb{R}^n) : (\varphi_i(x), V) \mapsto (\varphi_j(x), t_{ji}(x) \cdot V) \tag{24.34}$$

where $\{t_{ji}\}_{i,j \leq n}$ is the tangent bundle cocycle, but now with an action on the compact manifold $\mathrm{Gr}(k, \mathbb{R}^n)$ instead of the vector space $\mathbb{R}^n$. Using this set of transition functions we use a construction similar to 24.3.8 to create a new (fibre) bundle where every fibre is diffeomorphic to $\mathrm{Gr}(k, \mathbb{R}^n)$, namely it is the Grassmannian $\mathrm{Gr}(k, T_p M)$ associated to the tangent space in every point $p \in M$.

Notation 24.4.15. The Grassman k-plane bundle is denoted by $\mathrm{Gr}(k, TM)$.

Definition 24.4.16 (Distribution). A smooth section of the Grassman k-plane bundle is called a distribution of k-planes.

Definition 24.4.17 (Integrable). Let M be a smooth manifold and let $W \in \Gamma(\mathrm{Gr}(k, TM))$ be a distribution of k-planes. A submanifold $N \subseteq M$ is said to integrate W with initial condition $p_0 \in M$ if for every $p \in N$ we find that $W(p) = T_p N$ and $p_0 \in N$. W is said to be integrable if there exists such a submanifold N.

Definition 24.4.18 (Frobenius integrability condition). A distribution of k-planes W over a smooth manifold M is said to satisfy the Frobenius integrability condition in 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 $p \in U$, there Lie bracket $[X, Y](p)$ is also an element of $W(p)$ for all $p \in U$.

Theorem 24.4.19 (Frobenius' integrability theorem). *Let W be a distribution of k-planes over a smooth manifold M. Then W is integrable if and only if W satisfies the Frobenius integrability condition.*

---

[20]Due to the fact that the Grassmannian is not a vector space, we construct a general fibre bundle and not a vector bundle.

24.5 Differential k -forms

Definition 24.5.1 (Differential form). A differential k-form is a map

$$\boxed{\omega : T^{\diamond k} M \to \mathbb{R}} \tag{24.35}$$

such that the restriction of ω to each fibre of the fibre product[21] $T^{\diamond k} M$ is multilinear and antisymmetric.

The space of all differential k-forms on a manifold M is denoted by $\Omega^k(M)$. $\Omega^0(M)$ is defined as the space of smooth functions $C^\infty : M \to \mathbb{R}$.

Alternative Definition 24.5.2. An alternative definition goes as follows. Consider the representation

$$\rho_k : GL(R^{m*}) \to GL(\Lambda^k(\mathbb{R}^{m*})) : T \mapsto T \wedge ... \wedge T$$

where T is a linear map. This representation induces an associated vector bundle[22] $\rho_k(\tau_M^*)$ of the cotangent bundle on M. A differential k-form is then given by a section of $\rho_k(\tau_M^*)$. $\Omega^k(M)$ can then be defined as follows:

$$\Omega^k(M) = \Gamma(\rho_k(\tau_M^*))$$

Construction 24.5.3. We can construct a graded algebra by equipping the graded vector space

$$\Omega(M) = \bigoplus_{k \geq 0} \Omega^k(M) \tag{24.36}$$

with the wedge product of differential forms (which is induced by the wedge product on $\Lambda^k(\mathbb{R}^{m*})$ through the alternative definition). This graded algebra is associative, graded-commutative and unital with the constant function $1 \in C^\infty(M)$ as identity element.

Definition 24.5.4 (Pullback). Let $f : M \to N$ be a smooth function between smooth manifolds and let ω be a differential k-form on N. The pullback of ω by f is defined as:

$$\boxed{f^*(\omega) = \omega \circ Tf : TM \to \mathbb{R}} \tag{24.37}$$

So f^* can be seen as a map pulling elements from T^*N back to T^*M.

Definition 24.5.5 (Pushforward). Let $f : M \to N$ be a diffeomorphism between smooth manifolds and let ω be a differential k-form on M. The pushforward ω by f is defined as:

$$f_*(\omega) : \omega \circ Tf^{-1} : TN \to \mathbb{R} \tag{24.38}$$

or using the pullback:

$$f_*(\omega) = f^{-1*}(\omega) \tag{24.39}$$

---

[21]See definition 24.4.
[22]See definition 24.3.11.

Remark. Note that the pushforward of differential k-form is only defined for diffeomorphisms, in constrast to pullbacks which only require smooth functions. Furthermore this also explains why differential forms are the most valuable elements in differential geomeotry. Vector fields can't even be pulled back in general by smooth maps.

Formula 24.5.6 (Dual basis). Consider the basis $\{\frac{\partial}{\partial x_i}\big|_p\}_{i\le n}$ from definition 21.3 for the tangent space T_pM. From this set we can construct a natural dual basis for the cotangent space T_p^*M using the natural pairing of these spaces:

$$\left\langle \frac{\partial}{\partial x_i}, dx_j \right\rangle = \delta_{ij} \tag{24.40}$$

24.5.1   Exterior derivative

Definition 24.5.7 (Exterior derivative). The exterior derivative d_k is a map defined on the graded algebra of differential k-forms:

$$d_k : \Omega^k(M) \to \Omega^{k+1}(M)$$

For $k = 0$ it is given by[23]:

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i \tag{24.41}$$

where we remark that the 'infinitesimals' are in fact unit vectors with norm 1. This formula can be generalized to higher dimensions as follows:

$$\boxed{d(f\,dx_{i_1} \wedge ... \wedge dx_{i_k}) = df \wedge dx_{i_1} \wedge ... \wedge dx_{i_k}} \tag{24.42}$$

Corollary 24.5.8. It follows immediately from 24.42 that

$$d(dx_i) = 0 \tag{24.43}$$

for all $i \le n$.

Property 24.5.9. The exterior derivatives have following properties:

- For all $k \ge 0$, for all $\omega \in \Omega^k(M)$: $d_k \circ d_{k+1} = 0$, so $\text{im}(d_k) \subseteq \ker(d_{k+1})$.

- The exterior derivative is an $\mathbb{R}$-linear map.

- Graded Leibniz rule:

$$d(\omega_1 \wedge \omega_2) = d\omega_1 \wedge \omega_2 + (-1)^j \omega_1 \wedge d\omega_2 \tag{24.44}$$

  where $\omega_1 \in \Omega^j(M), \omega_2 \in \Omega^k(M)$.

---

[23]For $f \in \Omega^0(M)$, we call df the **differential** of f.

- Let $f \in C^\infty(M)$: $f^*(d\omega) = d(f^*\omega)$ where f^* denotes the pullback 24.37.

Remark 24.5.10 (†). The gradient, rotor (curl) and divergence from standard vector calculus can be rewritten using exterior derivatives as follows: Let $\vec{f} = (f_1, f_2, f_3)$ with f_i smooth for every i and let f be a smooth function. Denote the canonical isomorphism between $\mathbb{R}^3$ and $\mathbb{R}^{3*}$ by $\sim$.

$$\sim df = \nabla f \tag{24.45}$$
$$\sim (*d\alpha) = \nabla \times \vec{f} \tag{24.46}$$
$$*d\omega = \nabla \cdot \vec{f} \tag{24.47}$$

Example 24.5.11. Let $f \in C^\infty(M, \mathbb{R})$. Let γ be a curve on M. From the definition 24.40 of the basis $\{dx_k\}_{k \leq n}$ we obtain following result:

$$\langle df(x), \gamma'(t) \rangle = \sum_k \frac{\partial f}{\partial x_k}(x)\gamma_k'(t) = (f \circ \gamma')(t)$$

Example 24.5.12. An explicit formula for the exterior derivative of a 1-form Φ is:

$$d\Phi(X, Y) = X(\Phi(Y)) - Y(\Phi(X)) - \Phi([X, Y]) \tag{24.48}$$

24.5.2 Lie derivative

Formula 24.5.13 (Lie derivative of differential forms).

$$\mathcal{L}_X\omega(p) = \lim_{t \to 0} \frac{\sigma_t^*\omega - \omega}{t}(p) \tag{24.49}$$

Formula 24.5.14 (Lie derivative of smooth functions). Using the definition of the exterior derivative of smooth functions 24.41 and the definition of the dual (cotangent) basis 24.40 we can rewrite the Lie derivative 24.29 as:

$$Xf(p) = df_p(X(p)) \tag{24.50}$$

Formula 24.5.15. The Lie derivative also has following Leibniz-type property with respect to differential forms:

$$\mathcal{L}_X(\omega(Y)) = (\mathcal{L}_X\omega)(Y) + \omega(\mathcal{L}_XY) \tag{24.51}$$

where X, Y are two vector fields and ω is a 1-form.

Formula 24.5.16 (Lie derivative of tensor fields). By comparing the definitions of the Lie derivatives of vector fields 24.31 and differential forms 24.49 we can see that both definitions are identical upon replacing X by ω. This leads to the definition of a Lie derivative of a general tensor field $\mathcal{T} \in \Gamma(T^{(k,l)}M)$:

$$\mathcal{L}_X\mathcal{T}(p) = \frac{d}{dt}\sigma_t^*\mathcal{T}(\gamma_p(t))\Big|_{t=0} \tag{24.52}$$

24.5.3 Interior product

Definition 24.5.17 (Interior product). Aside from the differential (exterior derivative) we can also define another operation on the algebra of differential forms:

$$\iota_X : (\iota_X \omega)(v_1, ..., v_{k-1}) \mapsto \omega(X, v_1, ..., v_{k-1}) \tag{24.53}$$

This antiderivation (of degree -1) from $\Omega^k(M)$ to $\Omega^{k-1}(M)$ is called the **interior product** or **interior derivative**. This can be seen as a generalization of the contraction map 18.12.

Formula 24.5.18 (Cartan[24]). 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:

$$\mathcal{L}_X \omega = \iota_X(d\omega) + d(\iota_X \omega) \tag{24.54}$$

24.5.4 de Rham Cohomology

Definition 24.5.19 (Exact form). Let $\omega \in \Omega^k(M)$. If ω can be written as $\omega = d\chi + 0$ for some $\chi \in \Omega^{k-1}(M)$ and $0 \in \Omega^0(M)$ the zero function then ω is said to be exact. It follows that

$$\mathrm{im}(d_k) = \{\omega \in \Omega^{k+1}(M) : \omega \text{ is exact}\} \tag{24.55}$$

Definition 24.5.20 (Closed form). Let $\omega \in \Omega^k(M)$. If $d\omega = 0$ then ω i said to be closed. It follows that

$$\{\omega \in \Omega^k(M) : \omega \text{ is closed}\} \subseteq \ker(d_k) \tag{24.56}$$

Remark 24.5.21. From the first item of property 24.5.9 it follows that every exact form is closed. The converse however is not true[25].

Definition 24.5.22 (Cochain complex). Let $(A_k)_{k\in\mathbb{N}}$ be a sequence of Abelian groups or modules together with a sequence $(\partial_k)_{k\in\mathbb{N}}$ of homomorphisms, called the **boundary operators** or **differentials**, such that for all k:

$$\partial_k : A_k \to A_{k+1} \tag{24.57}$$

Furthermore let $\partial_k^2 = 0$ for every $k \in \mathbb{N}$. This structure is called a cochain complex[26]. Elements in $\mathrm{im}(\partial_k)$ are called **coboundaries** and elements in $\ker(\partial_k)$ are called **cocycles**.

Definition 24.5.23 (de Rham complex). The structure given by the chain

$$0 \to \Omega^0(M) \xrightarrow{d} \Omega^1(M) \xrightarrow{d} ...$$

together with the sequence of exterior derivatives d_k forms a cochain complex. This complex is called the de Rham complex.

---

[24] Sometimes called *Cartan's magic formula* or *Cartan's (infinitesimal) homotopy formula*.

[25] See result 24.5.26 for more information.

[26] A chain complex is constructed similarly. For this structure we consider a descending order, i.e.: $\partial_k : A_k \to A_{k-1}$.

The relation between closed and exact forms can be used to define the de Rham cohomology groups.

Definition 24.5.24 (de Rham cohomology). The kth de Rham cohomology group on M is defined as the following quotient space:

$$\boxed{H_{\mathrm{dr}}^k(M) = \frac{\ker(d_{k+1})}{\mathrm{im}(d_k)}} \tag{24.58}$$

This quotient space is a vector space. Two elements of the same equivalence class in $H_{dr}^k(M)$ are said to be **cohomologous**.

One can construct a graded ring 2.5.10 from these cohomology groups, called the cohomology ring H^*. The product is called the **cup product** $\smile$ and it is a graded-commutative product (see 2.25).

Definition 24.5.25 (Cup product). Let $[\nu] \in H_{\mathrm{dr}}^k$ and $[\omega] \in H_{\mathrm{dr}}^l$, where we used $[\cdot]$ to show that the elements are in fact equivalence relations belonging to differential forms ν and ω. The cup product is defined as follows: $[\nu] \smile [\omega] = [\nu \wedge \omega]$.

Theorem 24.5.26 (Poincaré's lemma[27]). *For every point $p \in M$ there exists a neighbourhood on which the de Rham cohomology is trivial:*

$$\forall p \in M : \exists U \subseteq M : H_{dr}^k(U) = 0 \tag{24.59}$$

This implies that every closed form is locally exact.

24.5.5   Vector-valued differential forms

Definition 24.5.27 (Vector-valued differential form). Let V be a vector space and E a vector bundle with V as typical fibre. A vector-valued differential form can be defined in two ways. Firstly we can define a vector-valued k-form as a map $\omega : \bigotimes^k TM \to V$. A more general definition is based on sections of a corresponding vector bundle:

$$\Omega^k(M, E) = \Gamma(E \otimes \Lambda^k T^*M) \tag{24.60}$$

Formula 24.5.28 (Wedge product). Let $\omega \in \Omega^k(M, E_1)$ and $\nu \in \Omega^p(M, E_2)$. The wedge product of these differential forms is defined as:

$$\omega \wedge \nu(v_1, ..., v_{k+p}) = \frac{1}{(k+p)!} \sum_{\sigma \in S_{k+p}} \mathrm{sgn}(\sigma)\omega\big(v_{\sigma(1)}, ..., v_{\sigma(k)}\big) \otimes \nu\big(v_{\sigma(k+1)}, ..., v_{\sigma(p)}\big) \tag{24.61}$$

This is a direct generalization of the formula for the wedge product of ordinary differential forms where we replaced the (scalar) product (product in the algebra $\mathbb{R}$) by the tensor product (product in the general tensor algebra). It should be noted that result of this operation is not an element of any of the original bundles E_1 or E_2 but of the product bundle $E_1 \otimes E_2$.

---

[27]The original theorem states that on a contractible space (see definition 4.7.5) every closed form is exact.

Definition 24.5.29 (Lie-algebra-valued differential form). A vector-valued differential form where the vector space V is equipped with a Lie algebra structure.

Formula 24.5.30 (Wedge product). Let $\omega \in \Omega^k(M, \mathfrak{g})$ and $\nu \in \Omega^p(M, \mathfrak{g})$. The wedge product of these differential forms is defined as:

$$[\omega \wedge \nu](v_1, ..., v_{k+p}) = \frac{1}{(k+p)!} \sum_{\sigma \in S_{k+p}} \mathrm{sgn}(\sigma)[\omega(v_{\sigma(1)}, ..., v_{\sigma(k)}), \nu(v_{\sigma(k+1)}, ..., v_{\sigma(p)})] \quad (24.62)$$

where $[\cdot, \cdot]$ is the Lie bracket in $\mathfrak{g}$.

24.6   Connections

24.6.1   Vertical bundle

Because smooth fibre bundles (which include smooth principal G-bundles) are also smooth manifolds we can define the traditional notions for them, such as the tangent bundle. We use this to construct the notions of horizontal and vertical bundles:

Definition 24.6.1 (Vertical vector). Let $\pi : E \to B$ be a smooth fibre bundle. The subbundle $\ker(T\pi)$ of TE is called the vertical bundle of E. Fibrewise this gives us $V_x = T_x(E_{\pi(x)})$.

For principal G-bundles we can use an equivalent definition:

Alternative Definition 24.6.2. Consider a smooth principal G-bundle $G \hookrightarrow P \xrightarrow{\pi} M$. We first construct a map ι_p for every element $p \in P$:

$$\iota_p : G \to P : g \mapsto p \cdot g \quad (24.63)$$

We then define a tangent vector $v \in T_pP$ to be vertical if it lies in the image of $T_e\iota_p$, i.e. $\mathrm{Vert}(T_pP) = \mathrm{im}(T_e\iota_p)$. This construction is supported by the exactness of following short sequence:

$$0 \to \mathfrak{g} \xrightarrow{T_e\iota_p} T_pP \xrightarrow{T_p\pi} T_xM \to 0 \quad (24.64)$$

Property 24.6.3 (Dimension). It follows from the second definition that the vertical vectors of a principal G-bundle are nothing but the pushforward of the Lie algebra $\mathfrak{g}$ under the right action of G on P. Furthermore, the exactness of the sequence implies that $T_e\iota_p : \mathfrak{g} \to \mathrm{Vert}(T_pP)$ is an isomorphism of vector spaces. In particular, it implies that

$$\dim \mathrm{Vert}(T_pP) = \dim \mathfrak{g} = \dim G \quad (24.65)$$

Definition 24.6.4 (Fundamental vector field). Consider a principle G-bundle. Let $A \in \mathfrak{g}$, where $\mathfrak{g}$ is the Lie algebra corresponding to G. The vertical vector field $A^\# : P \to TP$ given by

$$A^\#(p) = \iota_{p,*}(A) \in \mathrm{Vert}(T_pP) \quad (24.66)$$

is called the fundamental vector field associated to A.

Property 24.6.5. The map $(\cdot)^\# : \mathfrak{g} \to \Gamma(TP)$ is a Lie algebra morphism:

$$[A, B]^\# = [A^\#, B^\#] \tag{24.67}$$

where the Lie bracket on the left is that in $\mathfrak{g}$ and the Lie bracket on the right is that in $\mathfrak{X}(M)$ given by 24.32.

Property 24.6.6. The vertical bundle satisfies the following G-equivariance condition:

$$R_{g,*}(\mathrm{Vert}(T_pP)) = \mathrm{Vert}(T_{pg}P) \tag{24.68}$$

By differentiating the equality

$$R_g \circ \iota_p = \iota_{pg} \circ \mathrm{ad}_{g^{-1}}$$

and using 24.66 and 23.8 we obtain the following algebraic formulation of the G-equivariance condition:

$$R_{g,*}\left(A^\#(p)\right) = (\mathrm{Ad}_{g^{-1}}A)^\#(pg) \tag{24.69}$$

24.6.2   Horizontal bundle

Definition 24.6.7 (Connection). Consider a principal bundle P with structure group G. A connection on P is the selection of a subspace $\mathrm{Hor}(T_pP) \le T_pP$ for every $p \in P$ such that:

- $\mathrm{Vert}(T_pP) \oplus \mathrm{Hor}(T_pP) = T_pP$

- The selection depends smoothly on p.[28]

- The subspace $\mathrm{Hor}(T_pP)$ is G-equivariant:

$$R_{g,*}(\mathrm{Hor}(T_pP)) = \mathrm{Hor}(T_{pg}P) \tag{24.70}$$

The elements of $\mathrm{Hor}(T_pP)$ are said to be **horizontal vectors** with respect to the connection.

Remark 24.6.8. Remark that the G-invariance condition for vertical bundles is an intrinsic property while we have to require it by definition for the horizontal bundle.

Definition 24.6.9 (Horizontal bundle). The horizontal bundle $\mathrm{Hor}(TP)$ is defined as $\bigcup_{p \in P} \mathrm{Hor}(T_pP)$. It is a subbundle of TP. The G-invariance condition then implies that this subbundle is invariant under (the pushforward of) the right action of G.

Property 24.6.10 (Dimension). Properties 24.13, 24.65 and the direct sum decomposition of T_pP imply the following relation:

$$\dim \mathrm{Hor}(T_pP) = \dim M \tag{24.71}$$

---

[28]See the definiton of a (smooth) distribution 24.4.16.

Property 24.6.11. We take some time to summarize all dimensional relations between the components of a principal G-bundle over a base manifold M:

$$\dim P = \dim M + \dim G \tag{24.72}$$
$$\dim M = \dim \mathrm{Hor}(T_p P) \tag{24.73}$$
$$\dim G = \dim \mathrm{Vert}(T_p P) \tag{24.74}$$

for all $p \in P$.

Definition 24.6.12 (Dual connection). First we define the dual of the horizontal bundle:

$$\mathrm{Hor}(T_p^* P) = \{h^* \in T_p^* P | h^*(v) = 0, v \in \mathrm{Vert}(T_p P)\} \tag{24.75}$$

Equivalently, the horizontal covector bundle is defined as the set of linear functionals that annihilate vertical vectors. Just as with the vertical bundle this structure is independent of any connection on P.

A dual connection can then be defined as the selection of a vertical covector bundle $\mathrm{Vert}(T_p^* P)$ satisfying the conditions of definition 24.6.7 where Vert and Hor should be interchanged.

24.6.3   Ehresmann connection

Definition 24.6.13 (Ehresmann connection). Let (P, M, π, G) be a principal bundle. An Ehresmann connection is a $\mathfrak{g}$-valued 1-form $\omega : TP \to \mathfrak{g}$ that satisfies following 2 conditions:

1.
$$\omega \circ R_{g,*} = \mathrm{Ad}_{g^{-1}} \circ \omega \tag{24.76}$$

2.
$$\omega(A^{\#}) = A \tag{24.77}$$

The horizontal subspaces are then defined as $\mathrm{Hor}(T_p P) = \ker \omega|_p$.

Property 24.6.14. Consider two principal G-bundles ξ_1 and ξ_2. Let ω be an Ehresmann connection on ξ_1 and let $F; \xi_1 \to \xi_2$ be a bundle map covering a smooth map f. The map $F^* \omega$ defines an Ehresmann connection on ξ_2.

Example 24.6.15. Consider a principal G-bundle. An Ehresmann connection on this bundle is given by the following map:
$$\omega = (T_e \iota_p)^{-1} \circ \mathrm{pr}_V \tag{24.78}$$
where pr_V is the projection $TP \to \mathrm{Vert}(TP)$ associated to the decomposition from definition 24.6.7.

Definition 24.6.16 (Horizontal and vertical forms). Let ω be an Ehresmann connection on a principal bundle $G \hookrightarrow P \to M$. Let $\theta \in \Omega^k(P)$ be a k-form. We define following notions:

- θ is said to be horizontal if
$$\theta(v_1, ..., v_k) = 0 \tag{24.79}$$
whenever at least 1 of the v_i lies in $\mathrm{Vert}(T_pP)$.

- θ is said to be vertical if
$$\theta(v_1, ..., v_k) = 0 \tag{24.80}$$
whenever at least 1 of the v_i lies in $\mathrm{Hor}(T_pP)$.

For functions $f \in \Omega^0(P)$ it is vacuously true that they are both vertical and horizontal.

24.6.4 Maurer-Cartan form

Definition 24.6.17 (Maurer-Cartan form). For every $g \in G$ we have that the tangent space T_gG is isomorphic to $T_eG \cong \mathfrak{g}$. The isomorphism $T_gG \to \mathfrak{g}$ is given by the Maurer-Cartan form:
$$\boxed{\Omega := L_{g^{-1},*}} \tag{24.81}$$

Definition 24.6.18. Consider a manifold $M = \{x\}$. When constructing a principal G-bundle over M we see that the total space $P = \{x\} \times G$ can be identified with the structure group G. From the relations in property 24.6.11 we see that the horizontal spaces are null-spaces (which defines a smooth distribution and thus a connection according to 24.6.7) and that the vertical spaces are equal to the tangent spaces, i.e. $\mathrm{Vert}(T_gG) = T_gG$ (where we already made the association $P \cong G$).

The simplest way to define a connection form ω on this bundle would be the trivial projection $TP \to \mathrm{Vert}(TP) = \mathbb{1}_{TP}$. The image of this map would however be T_gG and not $\mathfrak{g}$ as required. This can be solved by using the Maurer-Cartan form $\Omega : T_gG \to \mathfrak{g}$, i.e. we define $\omega(v) = \Omega(v)$.

Property 24.6.19. The Maurer-Cartan form is the unique Ehresmann connection on the bundle $G \hookrightarrow G \to \{x\}$.

24.6.5 Horizontal lifts and parallel transport

Property 24.6.20. Consider a principal G-bundle $G \hookrightarrow P \to M$ and a curve $\gamma : [0, 1] \to M$. Let $p_0 \in \pi^{-1}(\gamma(0))$. There exists a unique curve $\widetilde{\gamma}_{p_0} : [0, 1] \to P$ satisfying the following conditions:

- $\widetilde{\gamma}_{p_0}(0) = p_0$

- $\pi \circ \widetilde{\gamma}_{p_0} = \gamma$

- $\widetilde{\gamma}'_{p_0}(t) \in \mathrm{Hor}(TP)$ for all $t \in [0, 1]$

The curve $\widetilde{\gamma}_{p_0}$ is said to be 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 ommited and we write $\widetilde{\gamma}$ instead of $\widetilde{\gamma}_{p_0}$.

Remark 24.6.21 (Horizontal curve). Curves satisfying the last condition are said to be horizontal.

Definition 24.6.22 (Parallel transport on principal bundles). The parallel transport map is defined as follows:

$$\operatorname{Par}_t^\gamma : \pi^{-1}(\gamma(0)) \to \pi^{-1}(\gamma(t)) : p_0 \mapsto \widetilde{\gamma}_{p_0}(t) \tag{24.82}$$

This map is G-equivariant and it is a diffeomorphism of fibres.

Formula 24.6.23. Consider a principal bundle $G \hookrightarrow P \to M$. Let $\gamma(t)$ be a curve in M and let ω be an Ehresmann connection on this bundle. The horizontal lift of $\gamma(t)$ can be parametrized as $(\gamma(t), g(t)$ where $g(t)$ is some unique curve in G. To determine $\widetilde{\gamma}(t)$ it is thus sufficient to find $g(t)$. The following uniquely (and globally) characterizes $g(t)$:

$$g'(t) = -\omega(\gamma(t), e, \gamma'(t), 0)g(t) \tag{24.83}$$

Using the trivial section $s : U \to U \times G$ where U is an open subset of M we can rewrite this formula. First we consider the action of the pullback $s^*\omega$ on the derivative $T\gamma(t, 1) = (\gamma(t), \gamma'(t))$. Using the fact that it is linear (because it is a 1-form) we can write

$$s^*\omega(\gamma(t), \gamma'(t)) = A(\gamma(t))\gamma'(t)$$

where $A : M \to \operatorname{Hom}(\mathbb{R}^m, \mathfrak{g})$ 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

$$s^*\omega(\gamma(t), \gamma'(t)) = \omega[s_*(\gamma(t), \gamma'(t)))] = \omega(\gamma(t), e, \gamma'(t), 0)$$

Combining these relations with the ODE for $g(t)$ gives

$$\left(\frac{d}{dt} + A(\gamma(t))\gamma'(t)\right)g(t) = 0 \tag{24.84}$$

where $\frac{d}{dt}$ is a matrix given by the scalar multiplication of the derivative $\frac{d}{dt}$ and the unit matrix I.

24.6.6 Koszul connections and covariant derivatives

Definition 24.6.24 (Parallel transport on vector bundles). Consider a principal bundle $G \hookrightarrow P \to M$ where we explicitly require P to be trivial, i.e. $P = M \times G$. Suppose that the Lie group G acts on a vector space V by a representation $\rho : G \to GL_m$. We can then construct an associated vector bundle $\pi_1 : M \times V \to M$.

Parallel transport on this vector bundle is then defined as follows. Let $\gamma(t)$ be a curve in M such that $\gamma(0) = x_0$ and $x_1 = \gamma(1)$. Furthemore, let the horizontal lift $\widetilde{\gamma}(t)$ have $\widetilde{\gamma}(0) = (x_0, h)$ as initial condition. The parallel transport of the point $(x_0, v_0) \in M \times V$ along γ is given by the following map:

$$\operatorname{Par}_t^\gamma : \pi_1^{-1}(x_0) \to \pi_1^{-1}(\gamma(t)) : (x_0, v_0) \mapsto \left(\gamma(t), \rho\big(g(t)h^{-1}\big)v_0\right) \tag{24.85}$$

It should be noted that this map is independent of the initial element $h \in G$. Furthermore, Par_t^γ is an isomorphism of vector spaces and can thus be used to identify distant fibers (as long as they lie in the same path-component).

Remark 24.6.25. Two remarks have to be made. First of all, although the previous construction explicitly used trivial bundles, it is also valid for general non-trivial vector bundles. Secondly, following remark 24.3.12 we can construct a principal bundle for any vector bundle and use the parallel transport on this bundle to define parallel transport of vectors. The previous construction is thus possible for every vector bundle.

Definition 24.6.26 (Covariant derivative). Consider a vector bundle with model fibre space V and its associated principal $GL(V)$-bundle with Ehresmann connection ω, both over a base manifold M. 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:

$$\nabla_X \sigma(x_0) = \lim_{t \to +\infty} \frac{(\mathrm{Par}_t^\gamma)^{-1} \sigma(\gamma(t)) - \sigma(x_0)}{t} \tag{24.86}$$

where $\gamma(t)$ is any curve such that $\gamma(0) = x_0$ and $\gamma'(0) = X(x_0)$.

Property 24.6.27. Let $\pi : E \to M$ be a vector bundle. Let σ, X and f be respectively a section on E, a vector field on M and a C^∞ function on M. The covariant derivative along X satisfies following properties:

- $\nabla_X \sigma$ is a smooth section on E.

- The map $(X, \sigma) \mapsto \nabla_X \sigma$ is bilinear over $\mathbb{R}$.

- $\nabla_{(fX)} \sigma = f \nabla_X \sigma$

- $\nabla_X (f\sigma) = f \nabla_X \sigma + X(f) \sigma$

Remark 24.6.28. The last two properties show the major difference between the Lie derivative and the covariant derivative when σ is a section of the tangent bundle, i.e. a vector field. 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 24.6.29 (Koszul connection). The map

$$\Gamma(TM) \times \Gamma(E) \to \Gamma(E) : (X, \sigma) \mapsto \nabla_X \sigma \tag{24.87}$$

is called a Koszul connection if the above properties hold. From the above constructions it also follows that every Ehresmann connection on a principal bundle induces a Koszul connection on all of its associated vector bundles.

Definition 24.6.30 (Exterior covariant derivative). Consider a principal bundle $G \hookrightarrow P \to M$ equipped with an Ehresmann connection ω. Let $\theta \in \Omega^k(P)$ be a differential k-form. The exterior covariant derivative $D\theta$ is defined as follows:

$$D\theta(v_0, ..., v_k) = d\theta(v_0^H, ..., v_k^H) \tag{24.88}$$

where d is the exterior derivative 24.5.7 and v_i^H is the projection of v_i on the horizontal subspace $\mathrm{Hor}(T_pP)$ associated to the Ehresmann connection ω. From the definition it follows that the covariant derivative $D\theta$ is a horizontal form 24.6.16.

Remark 24.6.31. The exterior covariant derivative can also be defined for general W-valued k-forms where W is a vector space. This can be done by defining it component-wise with respect to a given basis on W. Afterwards one can prove that the choice of basis plays no role.

Formula 24.6.32. Using the Koszul connection on the tangent bundle TP we can rewrite the action of the exterior covariant derivative as follows:

$$D\theta(v_0, ..., v_k) = \sum_i^k (-1)^i \nabla_{v_i} \theta(v_0, ..., \hat{v}_i, ..., v_k) + \sum_{i<j}^k (-1)^{i+j} \theta([v_i, v_j], v_0, ..., \hat{v}_i, ..., \hat{v}_j, ..., v_k)$$
(24.89)

where $\hat{v}_i$ means that this vector is omitted. As an example we explicitly give the formula for a 1-form Φ:

$$D\Phi(X, Y) = \nabla_X(\Phi(Y)) - \nabla_Y(\Phi(X)) - \Phi([X, Y])$$
(24.90)

which should remind the reader of the analogous formula for the ordinary exterior derivative 24.48.

24.6.7   Curvature of a connection

Definition 24.6.33 (Curvature). Let ω be an Ehresmann connection on a principal bundle $G \hookrightarrow P \to M$. The curvature of ω is defined as the exterior covariant derivative $D\omega$.

Definition 24.6.34 (Flat connection). An Ehresmann connection ω is said to be flat if its curvature $D\omega$ vanishes everywhere.

The following property is an immediate consequence of the Frobenius integrability theorem 24.4.19 and the fact that an Ehresmann connection vanishes on the horizontal subbundle.

Property 24.6.35. Let ω be an Ehresmann connection. The associated horizontal distribution[29]

$$p \mapsto \mathrm{Hor}(T_pP)$$

is integrable if and only if the connection ω is flat. The vertical distribution is always integrable.

24.7   Complex bundles

---

[29]See 24.4.16 for the definition of a distribution of vector spaces.

Chapter 25

Integration on manifolds

25.1 Orientation

Definition 25.1.1 (Orientation). Similar to definition 18.3.18 we can define an orientation on a differentiable manifold M. First we modify the definition of the volume element a little bit. A **volume form** on M is a nowhere-vanishing top-dimensional form $\text{Vol} \in \Omega^n(M)$ where $n = \dim(M)$. The definition of an orientation is then equivalent to that in 18.3.18.

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 means that the transition functions have a positive Jacobian determinant[1]. The existence of a volume form turns a differentiable manifold into an **orientable manifold**.

An orientable manifold with volume form ω is said to be **positively oriented** if $\omega(v_1, ..., v_n) > 0$ where $(v_1, ..., v_n)$ is a basis for T_pM.

Example 25.1.2. Let $M = \mathbb{R}^n$. The canonical Euclidean volume form is given by the determinant map

$$\det : (u_1, ..., u_n) \mapsto \det(u_1, ..., u_n) \tag{25.1}$$

where the u_n's are expressed in the canonical basis $(e_1, ..., e_n)$. The name 'volume form' is justified by noting that the determinant map gives the signed volume of the n-dimensional parallelotope spanned by the vectors $\{u_1, ..., u_n\}$.

Property 25.1.3. Let ω_1, ω_2 be two volume forms on M. Then 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.

---

[1]This is in fact an equivalent definition.

25.2 Integration of top-dimensional forms

Formula 25.2.1. Let θ be a top-dimensional form on M with compact support. Let $\{\varphi_i\}_i$ be a partition of unity[2] subordinate to an atlas $\{(U_i, \varphi_i)\}_i$.

$$\int_M \theta = \sum_i \int_{U_i} \varphi_i \theta \, dx_1 ... dx_n \qquad (25.2)$$

25.3 Stokes' theorem

Theorem 25.3.1 (Stokes' theorem). *Let Σ be an orientable smooth manifold. Denote the boundary of Σ by $\partial\Sigma$. Let ω be a differential k-form on Σ. We have the following equality:*

$$\boxed{\int_{\partial\Sigma} \omega = \int_\Sigma d\omega} \qquad (25.3)$$

Corollary 25.3.2. The Kelvin-Stokes theorem 15.20, the divergence theorem 15.21 and Green's identity 15.22 are immediate results of this (generalized) Stokes' theorem.

25.4 de Rham Cohomology

Now we can also give a little side note about why the de Rham cohomology groups 24.58 really form a cohomology theory. For this we need some concepts from homology which can be found in section 4.8. Let M be a compact differentiable manifold and let $\{\lambda_i : \Delta^k \to M\}$ be the set of singular k-simplexes on M.

Now suppose that we want to integrate over a singular k-chain C on M, i.e. $C = \sum_{i=0}^k a_i \lambda_i$. Formula 25.2 says that we can pair the k-form ω and the chain C such that they act as duals to each other (hence p-forms are also called p-**cochains**), producing a real number[3]:

$$\langle \cdot, C \rangle : \Omega^n(M) \to \mathbb{R} : \omega \mapsto \int_C \omega = \sum_{i=0}^k a_i \int_{\Delta_k} \lambda_i^* \omega \qquad (25.4)$$

where λ_i^* pulls back ω to Δ^k which is a subset of $\mathbb{R}^k$ as required. Now Stokes' theorem 25.3 tells us that

$$\int_C d\omega = \int_{\partial C} \omega \qquad (25.5)$$

---

[2]See definition 4.5.14.

[3]This requires the chain group to have real coefficients instead of integer coefficients as is mostly used in homology.

Using the paring $\langle \cdot, \cdot \rangle$ this becomes

$$\langle d\omega, C \rangle = \langle \omega, \partial C \rangle \tag{25.6}$$

The operators d and ∂ can thus be interpreted as formal adjoints. After checking (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[4] $\langle \omega, C \rangle$ we see that the singular homology groups and the de Rham cohomology groups on M are well defined dual groups. The name cohomology is thus wel chosen for 24.58.

---

[4]Suppose that $A, B \in [C]$ and $\phi, \chi \in [\omega]$ then $\langle \phi, A \rangle = \langle \chi, B \rangle$.

Chapter 26

Riemannian Geometry

26.1 Riemannian manifolds

26.1.1 Metric

Definition 26.1.1 (Bundle metric). Consider the bundle of second order covariant vectors. Following from 18.2.2 every section g of this bundle gives a bilinear map

$$g_x : T_x M \times T_x M \to \mathbb{R}$$

for all $x \in M$. If this map is symmetric and non-degenerate and if it depends smoothly on p it is called a **(Lorentzian)** metric.[1]

The maps $\{g_x\}_{x \in M}$ can be 'glued' together to form a global metric g, defined on the fibre product[2] $TM \diamond TM$. Defining this map on $TM \times TM$ is not possible as tangent vectors belonging to different points in M cannot be 'compared'. The collection $\{\langle \cdot | \cdot \rangle_x | x \in M\}$ is called a **bundle metric**.

A Riemannian metric also induces a duality between TM and T^*M. This is given by the *flat* and *sharp* isomorphisms:

Definition 26.1.2 (Musical isomorphisms). Let $g : TM \times TM \to \mathbb{R}$ be the Riemannian metric on M. The **flat** isomorphism is defined as:

$$\flat : v \mapsto g(v, \cdot) \tag{26.1}$$

The **sharp** isomorphism is defined as the inverse map:

$$\sharp : p \mapsto v \tag{26.2}$$

such that $p(\cdot) = g(v, \cdot)$. These 'musical' isomorphisms can be used to lower and raise tensor indices.

---

[1]See also the section about Hermitian forms and metric forms 14.4.
[2]See definition 24.4.

26.1.2 Riemannian manifold

Definition 26.1.3 (Pseudo-Riemannian manifold). Let M be a smooth manifold. This manifold is called pseudo-Riemannian if it is equipped with a pseudo-Riemannian metric. A **Riemannian manifold** is similarly defined.

Definition 26.1.4 (Riemannian isometry). Let (M, g_M) and (N, g_N) be two Riemannian manifolds. An isometry 20.2.10 $f : M \to N$ is said to be Riemannian if $F^* g_N = g_M$.

Property 26.1.5. Let M be a pseudo-Riemannian manifold. For every $p \in M$ there exists a splitting $T_p M = P \oplus N$ where P is a subspace on which the pseudometric is positive-definite and N is a subspace on which the pseudometriv is negative-definite. This splitting is however not unique, only the dimensions of the two subspaces are well-defined.

Due to the continuity of the pseudometric, the dimensions of this splitting wil be the same for points in the same neighbourhood. For connected manifolds this amounts to a global invariant:

Definition 26.1.6 (Index). Let M be a connected pseudo-Riemannian manifold. The dimension of the *negative* subspace N in the above splitting $T_p P = P \oplus N$ is called the index of the pseudo-Riemannian manifold.

Theorem 26.1.7 (Whitney). *Every smooth paracompact[3] manifold admits a Riemannian metric.*

26.2 Sphere bundle

Definition 26.2.1 (Unit sphere bundle). Let V be a normed vector space. Consider a vector bundle $V \hookrightarrow E \to B$. From this bundle we can derive a new bundle where we replace the typical fibre V by the unit sphere $\{v \in V : ||v|| = 1\}$. It should be noted that this new bundle is not a vector bundle as the unit sphere is not a vector space.

Remark 26.2.2 (Unit disk bundle). A similar construction can be made by replacing the unit sphere by the unit disk $\{v \in V : ||v|| \leq 1\}$.

26.3 Hilbert bundles

Definition 26.3.1 (Hilbert bundle). A vector bundle for which the typical fibre is a Hilbert space is called a Hilbert bundle.

---

[3]See definition 4.5.12.

Definition 26.3.2 (Compatible Hilbert bundle). Consider the isomorphisms

$$l_x : F_x \to \mathcal{H} : h \mapsto \varphi_i(x, h) \in \pi(x) \tag{26.3}$$

where $\mathcal{H}$ is the typical fibre and where $\{(U_i, \varphi_i)\}_{i \in I}$ is a trivializing cover. These maps l_x are called **point-trivializing maps**.

Using these maps we can extend the metric structure of the typical fibre $\mathcal{H}$ to the fibres F_x for all x by:

$$\langle v|w \rangle_x = \langle l_x(v)|l_x(w) \rangle_{\mathcal{H}} \tag{26.4}$$

The Hilbert bundle is said to be compatible (with the metric structure on $\mathcal{H}$) if the above extension is valid for all $v, w \in F_x$.

Remark. For compatible Hilbert bundles, the transition maps $l_{x \to y} = l_y^{-1} \circ l_x : \pi^{-1}(x) \to \pi^{-1}(y)$ are easily seen to be isometries.

Chapter 27

Symplectic Topology

27.1 Symplectic manifolds

Definition 27.1.1 (Symplectic form). Let $\omega \in \Omega^2(M)$ be a differential 2-form. ω is said to be a symplectic form if it satisfies following properties:

- Closed: $d\omega = 0$

- Non-degeneracy: if $\omega(u, v) = 0, \forall u \in TM$ then $v = 0$

Definition 27.1.2 (Symplectic manifold). A manifold M equipped with a symplectic 2-form ω is called a symplectic manifold. This structure is often denoted as a pair (M, ω).

Property 27.1.3. From the antisymmetry (valid for all differential k-forms) and the non-degeneracy of the symplectic form, it follows that M is even dimensional.

Theorem 27.1.4 (Darboux). *Let (M, ω) be a symplectic manifold. For every neighbour-hood Ω in T^*M there exists a fibered chart (x^i, y^i) such that*

$$\omega|_\Omega = \sum_i dx^i \wedge dy^i \tag{27.1}$$

The charts in Darboux's theorem are called **Darboux charts** and they form a cover for M.

27.2 Lagrangian submanifolds

Definition 27.2.1 (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 $T_p^\perp S$ at the point $p \in S$ is defined as:

$$T_p^\perp S = \{v \in T_pM : \omega(v, \iota_* w) = 0, \forall w \in T_pS\} \tag{27.2}$$

Definition 27.2.2 (Isotropic submanifold). Let (M, ω) be a symplectic manifold. An embedded submanifold $\iota : S \hookrightarrow M$ is called isotropic if $T_p S \subset T_p^\perp S$.

Definition 27.2.3 (Isotropic submanifold). Let (M, ω) be a symplectic manifold. An embedded submanifold $\iota : S \hookrightarrow M$ is called co-isotropic if $T_p^\perp S \subset T_p S$.

Definition 27.2.4 (Larangian submanifold). Let (M, ω) be a symplectic manifold. An embedded submanifold $\iota : S \hookrightarrow M$ is called Lagrangian if $T_p S = T_p^\perp S$. Therefore they are sometimes called maximal isotropic submanifolds.

Part VI

Propability Theory & Statistics

Chapter 28

Propability

28.1 Propability

Definition 28.1.1 (Axioms of propability).

- $P(E) \geq 0$

- $P(E_1 \text{ or } E_2) = P(E_1) + P(E_2)$ if E_1 and E_2 are exclusive.

- $\sum_S P(E_i) = 1$ where the summation runs over all exclusive events.

Remark 28.1.2. The second axiom can be defined more generally by saying that the propability P should be σ-additive. Together with the first axiom and the consequence that $P(\emptyset) = 0$ means that the propability is a measure 8.1.1.

Definition 28.1.3 (Sample space). Let X be a random variable. The set of all possible outcomes of X is called the sample space. The sample space is often denoted by Ω.

Definition 28.1.4 (Propability space). Let (Ω, Σ, P) be a measure space[1]. This measure space is called a propability space if $P(X) = 1$. Furthermore, the measure P is called a propability measure or simply propability.

Definition 28.1.5 (Event). Let (Ω, Σ, P) be a propability space. An element S of the σ-algebra Σ is called an event.

Remark. From the definition of an event 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 event can occur simultaneously.

Remark. When working with measure-theoretic propability spaces it is more convenient to use the σ-algebra (see 2.3.2) of events instead of the power set (see 2.0.1) of all events.

---

[1]See definition 8.1.2.

214

Intuitively this seems to mean that some possible outcomes are not treated as events. However we can make sure that the σ-algebra still contains all 'useful' events by using a 'nice' definition of the used propability space. Further information concerning propability spaces can be found in chapter 8.

Formula 28.1.6. Let A, B be two events.

$$P(A \cup B) = P(A) + P(B) + P(A \cap B) \tag{28.1}$$

Definition 28.1.7 (Disjoint events). Two events A and B are said to be disjoint if they cannot happen at the same time:

$$P(A \cap B) = 0 \tag{28.2}$$

Corollary 28.1.8. If A and B are disjoint, the propability that both A and B are true is just the sum of their individual propabilities.

Formula 28.1.9 (Complement). Let A be an event. The propability of A being false is denoted as $P\left(\overline{A}\right)$ and is given by:

$$P\left(\overline{A}\right) = 1 - P(A) \tag{28.3}$$

Corollary 28.1.10. From the previous equation and de Morgan's laws (equations 2.4 and 2.5) we derive the following formula[2] :

$$P\left(\overline{A} \cap \overline{B}\right) = 1 - P(A \cup B) \tag{28.4}$$

28.2   Conditional propability

Definition 28.2.1 (Conditional propability). Let A, B be two events. The propability of A given that B is true is denoted as $P(A|B)$.

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}} \tag{28.5}$$

Corollary 28.2.2. By interchanging A and B in previous equation and by remarking that this has no effect on the quantity $P(A \cap B)$ the following rsult can be deducued:

$$P(A|B)P(B) = P(B|A)P(A) \tag{28.6}$$

Theorem 28.2.3 (Bayes' theorem). *Let A, B be two events. From the conditional propability 28.5 it is possible to derive following important theorem:*

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}} \tag{28.7}$$

---

[2]Switching the union and intersection has no impact on the validity of the formula.

Definition 28.2.4 (Independent events). Let A, B be two events. A and B are said to be independent if they satisfy the following relation:

$$P(A \cap B) = P(A)P(B) \qquad (28.8)$$

Corollary 28.2.5. If A and B are two independent events, then equation 28.7 simplifies to:

$$P(A|B) = P(A) \qquad (28.9)$$

Property 28.2.6. The events $A_1, ..., A_n$ are independent if for all $k \leq n$ for each choice of k events the propability of their intersection is equal to the product of their indivudal propabilities.

Property 28.2.7. The σ-algebras $\mathcal{F}_1, ..., \mathcal{F}_n$ defined on propability space $(\Omega, \mathcal{F}, P)$ are independent if for all choices of distinct indices $i_1, ..., i_k$ from $\{1, ..., n\}$ and for all choices of sets $F_{i_n} \in \mathcal{F}_{i_n}$ the following equation holds:

$$P(F_{i_1} \cap ... \cap F_{i_k}) = P(F_{i_1})...P(F_{i_k}) \qquad (28.10)$$

Formula 28.2.8. Let $(B_i)_{i \in \mathbb{N}}$ be a sequence of pairwise disjoint events. If $\bigcup_{i=1}^{+\infty} B_i = \Omega$ then the total propability of a given event A can be calculated as follows:

$$P(A) = \sum_{i=1}^{+\infty} P(A|B_i)P(B_i) \qquad (28.11)$$

28.3 Random variables

Definition 28.3.1 (Random variable). Let (Ω, Σ, P) be a propability space. A function $X : \Omega \to \mathbb{R}$ is called a random variable if $\forall a \in \mathbb{R} : X^{-1}\big([a, +\infty[\big) \in \Sigma$.[3]

Definition 28.3.2 (σ-algebra generated by a random variable). Let X be a random variable defined on a propability space (Ω, Σ, P). The following family of sets is a σ-algebra:

$$X^{-1}(\mathcal{B}) = \{S \in \Sigma : S = X^{-1}(B \in \mathcal{B})\} \qquad (28.12)$$

This measure is called the propability distribution of X.

Notation 28.3.3. The σ-algebra generated by the random variable X is often denoted by $\mathcal{F}_X$, analogous to notation 2.3.7.

Theorem 28.3.4. *Let X, Y be two random variables. X and Y are independent if the σ-algebras generated by them are independent*[4].

---

[3]$X^{-1}\big([a, +\infty[\big) = \{\omega \in \Omega : X(\omega) \geq a\}$.
[4]See equation 28.10.

28.4   Propability distribution

Definition 28.4.1 (Propability distribution). Let X be a random variable defined on a propability space (Ω, Σ, P). The following function is a measure on the σ-algebra of Borel sets:

$$\boxed{P_X(B) = P(X^{-1}(B))} \tag{28.13}$$

Formula 28.4.2 (Change of variable). Let X be a random variable defined on a propability space (Ω, Σ, P).

$$\int_\Omega g(X(\omega))dP(\omega) = \int_\mathbb{R} g(x)dP_X(x) \tag{28.14}$$

Definition 28.4.3 (Density). Let f be a non-negative integrable function and recall theorem 8.2.20. The function f is called the **density** of P with respect to the Lebesgue measure m.

For P to be a propability, f should satisfy the following condition:

$$\int f\,dm = 1 \tag{28.15}$$

Definition 28.4.4 (Cumulative distribution function). Let f be a density. The c.d.f. corresponding to f is given by:

$$F(y) = \int_{-\infty}^y f(x)dx \tag{28.16}$$

Theorem 28.4.5 (Skorokhod's representation theorem). *Let $F : \mathbb{R} \to [0,1]$ be a function that satisfies following 3 properties:*

- *$F(x)$ is non-decreasing.*

- *$\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to+\infty} F(x) = 1$*

- *$F(x)$ is right-continuous: $y \geq y_0, y \to y_0 \implies F(y) \to F(y_0)$.*

There exists a random variable $X : [0,1] \to \mathbb{R}$ defined on the propability space $([0,1], \mathcal{B}, m_{[0,1]})$ such that $F = F_X$.

Formula 28.4.6. Let the absolutely continuous propability P_X be defined on the product space $\mathbb{R}^n$. Let f_X be the density associated with P_X. Let $g : \mathbb{R}^n \to \mathbb{R}$ be integrable with respect to P_X.

$$\int_{\mathbb{R}^\times} g(x)dP_X(x) = \int_{\mathbb{R}^n} f_X(x)g(x)dx \tag{28.17}$$

Corollary 28.4.7. Previous formula together with formula 28.14 gives rise to:

$$\int_\Omega g(X)dP = \int_{\mathbb{R}^\times} f_X(x)g(x)dx \tag{28.18}$$

28.5 Moments

28.5.1 Expectation value

Definition 28.5.1 (Expectation value). Let X be random variable defined on a propability space (Ω, Σ, P).

$$\boxed{E(X) = \int_{\Omega} X dP}$$

(28.19)

Notation 28.5.2. Other often used notations are $\langle X \rangle$ and μ.

Definition 28.5.3 (Moment of order r). The moment of order r is defined as the expectation value of the rth power of X and by equation 28.18 this becomes:

$$E(X^r) = \int x^r f_X(x) dx$$

(28.20)

Definition 28.5.4 (Central moment of order r).

$$E((X - \mu)^r) = \int (x - \mu)^r f_X(x) dx$$

(28.21)

Definition 28.5.5 (Variance). The central moment of order 2 is called the variance: $V(X) = E((X - \mu)^2)$.

Property 28.5.6. If $E(X^n)$ are finite for $n > 0$ then for all $k \leq n$, $E(X^k)$ are also finite. If $E(X^n)$ is infinite then for all $k \geq n$, $E(X^k)$ are also infinite.

Property 28.5.7. Moments of order n are determined by central moments of order $k \leq n$ and central moments of order n are determined by moments of order $k \leq n$.

Definition 28.5.8 (Characteristic function). Let X be a random variable. The characteristic function of X is defined as follows:

$$\varphi_X(t) = E(e^{itX})$$

(28.22)

Property 28.5.9. The characteristic function has the following properties:

- $\varphi_X(0) = 1$
- $|\varphi_X(t)| \leq 1$
- $\varphi_{aX+b}(t) = e^{itb} \varphi_X(at)$

Formula 28.5.10. If $\varphi_X(t)$ is k times continuously differentiable then X has finite k^{th} moment and

$$E(X^k) = \frac{1}{i^k} \frac{d^k}{dt^k} \varphi_X(0)$$

(28.23)

Conversely, if X has finite k^{th} moment then $\varphi_X(t)$ is k times continuously differentiable and the above formula holds.

Formula 28.5.11 (Inversion formula). Let X be a random varibale. If the c.d.f. of X is continuous at $a, b \in \mathbb{R}$ then

$$F_X(b) - F_X(a) = \lim_{c \to +\infty} \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t) dt \qquad (28.24)$$

Formula 28.5.12. If $\varphi_X(t)$ is integrable then the c.d.f. is given by:

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \varphi_X(t) dt \qquad (28.25)$$

Remark 28.5.13. From previous formula it is clear that the density function and the characteristic function are Fourier transformed quantities.

28.5.2 Correlation

Theorem 28.5.14. *Let X, Y be two random variables. They are independent if and only if $E(f(X)g(Y)) = E(f(X))E(g(Y))$ holds for all Borel measurable[5] bounded functions f, g.*

The value $E(XY)$ is equal to the inner product $\langle X | Y \rangle$ as defined in 8.36. It follows that independence of random variables implies orthogonality. To generalize this concept, we introduce following notions.

Definition 28.5.15 (Centred random variable). Let X be a random variable with finite expectation value $E(X)$. The centred random variable X_c is defined as $X_c = X - E(X)$.

Definition 28.5.16 (Covariance). Let X, Y be two random variables. The covariance of X and Y is defined as follows:

$$\mathrm{cov}(X, Y) = \langle X_c | Y_c \rangle = E\big((X - E(X))(Y - E(Y))\big) \qquad (28.26)$$

Some basic math gives:

$$\mathrm{cov}(X, Y) = E(XY) - E(X)E(Y) \qquad (28.27)$$

Definition 28.5.17 (Correlation). Let X, Y be two random variables. The correlation is defined as the cosine of the angle between X_c and Y_c:

$$\rho_{XY} = \frac{\mathrm{cov}(X, Y)}{||X||_2 ||Y||_2} \qquad (28.28)$$

Corollary 28.5.18. From theorem 28.5.14 it follows that independent random variables are also uncorrelated.

Corollary 28.5.19. Uncorrelated X and Y satisfy the following equality: $E(XY) = E(X)E(Y)$.

Property 28.5.20. Let $(X_i)_{i \in \mathbb{N}}$ be a sequence of independent random variables. Their variances satisfy the following equation:

$$V\left(\sum_{i=1}^{+\infty} X_i \right) = \sum_{i=1}^{+\infty} V(X_i) \qquad (28.29)$$

---

[5]See definition 8.1.33.

28.5.3   Conditional expectation

Let (Ω, Σ, P) be a propability space. Let the random variable $X \in L^2(\Omega, \Sigma, P)^6$. Consider the sub-$\sigma$-algebra $\mathcal{G} \subset \Sigma$, The spaces $L^2(\Sigma)$ and $L^2(\mathcal{G})$ are complete (see property 8.4.3). The projection theorem 16.2.22 can thus be applied, i.e. there exists for every X 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:

$$\forall Z \in L^2(\mathcal{G}) : \langle X - Y | Z \rangle = \int_\Omega (X - Y) Z dP = 0 \tag{28.30}$$

And since $\mathbb{1}_G \in L^2(\mathcal{G})$ for every $G \in \mathcal{G}$ we find by applying 8.25:

$$\int_G X dP = \int_G Y dP \tag{28.31}$$

This leads us to introducing the following notion of conditional expectations:

Definition 28.5.21 (Conditional expectation). Let (Ω, Σ, P) be a propability 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 28.31 for every $G \in \mathcal{G}$. This Y is called the conditional expectation of X given $\mathcal{G}$ and it is denoted by $Y = E(X|\mathcal{G})$:

$$\boxed{\int_G E(X|\mathcal{G}) dP = \int_G X dP} \tag{28.32}$$

Remark 28.5.22. Allthough our derivation was based on random variables from the L^2 class, it is also possible to construct (unique) conditional expectations for random variables from the L^1 class by using method 8.2.23.

28.6   Joint distributions

Definition 28.6.1 (Joint distribution). Let X, Y be two random variables defined on the same propability space (Ω, Σ, P). Consider the vector $(X, Y) : \Omega \to \mathbb{R}^2$. The distribution of (X, Y) is defined on the Borel sets of the plane $\mathbb{R}^2$ and it is given by the following measure:

$$P_{(X,Y)}(B) = P((X, Y) \in B) \tag{28.33}$$

Definition 28.6.2 (Joint density). If the propability measure from previous definition can be written as

$$P_{(X,Y)}(B) = \int_B f_{(X,Y)}(x, y) dm_2(x, y) \tag{28.34}$$

for some integrable $f_{(X,Y)}$ it is said that X and Y have a joint density.

---

[6]This vector space has the same interpretation as the one in section 8.4.2. The difference is that all sets are elements of Σ instead of $\mathcal{M}$, that the functions are Σ-measurable instead of $\mathcal{M}$-measurable and that the integral is calculated with respect to the measure P instead of the Lebesgue measure m.

Definition 28.6.3 (Marginal distribution). The distributions of one-dimensional random variables is determined by the joint distribution:

$$P_X(A) = P_{(X,Y)}(A \times \mathbb{R}) \tag{28.35}$$

$$P_Y(A) = P_{(X,Y)}(\mathbb{R} \times A) \tag{28.36}$$

where $A \subset \mathcal{B}$.

Corollary 28.6.4. If the joint density exists then the marginal distributions are absolutely continuous and given by

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x,y)dy \tag{28.37}$$

$$f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x,y)dx \tag{28.38}$$

The converse however is not always true. The one-dimensional densities can be absolutely continuous without the existence of the joint density.

28.6.1 Independence

Theorem 28.6.5. *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, i.e.:*

$$P_{(X,Y)} = P_X \times P_Y$$

Remark 28.6.6. If X and Y are absolutely continuous then the previous theorem also applies with the densities instead of the distributions.

28.6.2 Conditional propability

Formula 28.6.7 (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:

$$h(y|X \in A) = \frac{\int_A f_{(X,Y)}(x,y)dx}{\int_A f_X(x)dx} \tag{28.39}$$

For $X = \{a\}$ this equation fails as the denominator would become 0. However it is possible to avoid this problem by formally putting

$$h(y|A = a) = \frac{f_{(X,Y)}(a,y)}{f_X(a)} \tag{28.40}$$

with $f_X(a) \neq 0$ which is non-restrictive because the propability of having a measurement $(X, Y) \in \{(x, y) : f_X(x) = 0\}$ is 0. We can thus define the conditional propability of Y given $X = a$:

$$P(Y \in B | X = a) = \int_B h(y | X = a) dy \tag{28.41}$$

Formula 28.6.8 (Conditional expectation).

$$E(Y | X)(\omega) = \int_{\mathbb{R}} y h(y | X(\omega)) dy \tag{28.42}$$

Furthermore, let $\mathcal{F}_X$ denote the σ-algebra generated by the random variable X. Using Fubini's theorem we can prove that for all sets $A \in \mathcal{F}_X$ the following equality, which should be compared with equation 28.32, holds:

$$\int_A E(Y | X) dP = \int_A Y dP \tag{28.43}$$

Remark 28.6.9. Following from previous two equations we can say that the conditional expectation $E(Y | X)$ is the best representation of the random variable Y as a function of X (i.e. measurable with respect to $\mathcal{F}_X$).

Property 28.6.10. As mentioned above, applying Fubini's theorem gives:

$$\boxed{E(E(Y | X)) = E(Y)} \tag{28.44}$$

Chapter 29

Statistics

In this chapter, all (or most) definitions and formulas will be based on either a standard calculus approach or a data-driven approach. For a measure-theory based approach see chapter 28.

29.1 Data sample

29.1.1 Moment

Formula 29.1.1 (r^{th} sample moment).

$$\overline{x^r} = \frac{1}{N} \sum_{i=1}^{N} x_i^r \tag{29.1}$$

Formula 29.1.2 (r^{th} central sample moment).

$$m_r = \frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x})^r \tag{29.2}$$

29.1.2 Average

Definition 29.1.3 (Arithmetic mean). The arithmetic mean is used to average out differences between measurements. It is also equal to the 1st moment:

$$\boxed{\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i} \tag{29.3}$$

Formula 29.1.4 (Weighted mean). Let $f : \mathbb{R} \to \mathbb{R}$ be a weight function. The weighted mean is given by:

$$\overline{x} = \frac{\sum\limits_i f(x_i)x_i}{\sum\limits_i f(x_i)} \tag{29.4}$$

Corollary 29.1.5. If the data has been grouped in bins, the weight function is given by the number of elements in each bin. Knowing this the mean becomes:

$$\overline{x} = \frac{1}{N} \sum_{i=1}^{N} n_i x_i \tag{29.5}$$

Remark 29.1.6. The measurements x_i can be replaced by function values $f(x_i)$ to calculate the mean of the function $f(x)$.

Remark 29.1.7. It is also important to notice that $\overline{f}(x) \neq f(\overline{x})$. The equality only holds for linear functions.

Definition 29.1.8 (Geometric mean). The geometric mean is used to average out *normalized* measurements, i.e. ratios with respect to a reference value.

$$g = \left(\prod_{i=1}^{N} x_i \right)^{1/N} \tag{29.6}$$

The following relation exists between the arithmetic and geometic mean:

$$\ln g = \overline{\ln x} \tag{29.7}$$

Definition 29.1.9 (Harmonic mean).

$$h = \left(\frac{1}{N} \prod_{i=1}^{N} x_i^{-1} \right)^{-1} \tag{29.8}$$

The following relation exists between the arithmetic and harmonic mean:

$$\frac{1}{h} = \overline{x^{-1}} \tag{29.9}$$

Definition 29.1.10 (Mode). The most occurring value in a dataset.

Definition 29.1.11 (Median). The median of dataset is the value x_i such that half of the values is greater than x_i and the other half is smaller than x_i.

29.1.3 Dispersion

Definition 29.1.12 (Range). The simplest indicator for statistical dispersion. It is however very sensitive for extreme values.

$$R = x_{max} - x_{min} \tag{29.10}$$

Definition 29.1.13 (Mean absolute difference).

$$\text{MD} = \frac{1}{N} \sum_{i=1}^{N} |x_i - \overline{x}| \tag{29.11}$$

Definition 29.1.14 (Sample variance).

$$V(X) = \frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x})^2 \tag{29.12}$$

Definition 29.1.15 (Standard deviation).

$$\sigma(X) = \sqrt{V(x)} \tag{29.13}$$

Corollary 29.1.16. The variance can also be written in the following way:

$$\boxed{V(X) = \overline{x^2} - \overline{x}^2} \tag{29.14}$$

Remark 29.1.17. A better estimator for the variance of a sample is the following formula:

$$\hat{s} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \overline{x})^2 \tag{29.15}$$

When using equation 29.12 it is more correct to exchange the sample mean $\overline{x}$ by the "true" mean μ. Otherwise one should use the estimator 29.15.

Definition 29.1.18 (Skewness). The skewness γ describes the asymmetry of a distribution. It is defined in relation to the third central moment m_3;

$$m_3 = \gamma \sigma^3 \tag{29.16}$$

where σ is the standard deviation.

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 29.1.19 (Pearson's mode skewness).

$$\gamma_P = \frac{\overline{x} - \text{mode}}{\sigma} \tag{29.17}$$

Definition 29.1.20 (Kurtosis). The kurtosis c is an indicator for the "tailedness". It is defined in relation to the fourth central moment m_4:

$$m_4 = c\sigma^4 \tag{29.18}$$

Remark 29.1.21 (Excess kurtosis). The excess kurtosis is defined as $c-3$. All univariate normal distributions have 0 excess. A positive excess is an indicator for long "fat" tails, a negative excess indicates short "thin" tails.

Definition 29.1.22 (Percentile). The pth percentile c_p is defined as the number such that it is larger than $p\%$ of the measurements.

Remark. From the previous definition it follows that the median is the 50th percentile.

Definition 29.1.23 (Interquartile range). The interquartile range is the difference between the upper and lower quartile (75th and 25th percentile respectively).

Definition 29.1.24 (FWHM). The Full Width at Half Maximum is the difference between the two values of the independent variable where the dependent variable is half of its maximum.

Property 29.1.25. For Gaussian distributions the following relation exists between the FWHM and the standard deviation σ:

$$\text{FWHM} = 2.35\sigma \tag{29.19}$$

29.1.4 Multivariate datasets

When working with bivariate (or even multivariate) distributions it is useful to describe the relationship between the different random variables. The following two definitions are often used.

Definition 29.1.26 (Covariance). Let X, Y be two random variables. The covariance of X and Y is defined as follows:

$$\text{cov}(X, Y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y}) = \overline{xy} - \overline{x}\,\overline{y} \tag{29.20}$$

Remark. The covariance of X and Y is often denoted as σ_{XY}. Furthermore it is important to observe the following equality:

$$\sigma_X^2 = \sigma_{XX}$$

Definition 29.1.27 (Correlation coefficent).

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \tag{29.21}$$

The correlation coefficient is bounded to the interval $[-1, 1]$. Furthermore its magnitude is an indicator only for the linear dependence.

Remark 29.1.28. For multivariate distributions these two definitions can be generalized using the following matrices:

$$V_{ij} = \text{cov}(x_{(i)}, x_{(j)}) \tag{29.22}$$

$$\rho_{ij} = \rho_{(i)(j)} \tag{29.23}$$

where $\text{cov}(x_{(i)}, x_{(j)})$ and $\rho_{(i)(j)}$ are defined as in equations 29.20 and 29.21.

Using these notations we get the following general equality:

$$\boxed{V_{ij} = \rho_{ij}\sigma_i\sigma_j} \tag{29.24}$$

Remark 29.1.29. From equation 29.24 it is clear that the diagonal elemets ρ_{ii} are equal to 1.

29.2   Law of large numbers

Theorem 29.2.1 (Law of large numbers). *If the size N of a sample tends towards infinity, then the observed frequencies tend towards the theoretical propabilities.*

Corollary 29.2.2 (Frequentist propability[1]).

$$P(X) = \lim_{n\to\infty} \frac{f_n(X)}{n} \tag{29.25}$$

29.3   Propability densities

Remark. In the following sections and subsections, all distributions will be taken to be continuous. The formulas can be modified for use with discrete distributions by replacing the integral with a summation.

Definition 29.3.1 (Propability density functions p.d.f). Let X be a random variable and $P(X)$ the associated propability distribution. The p.d.f. $f(X)$ is defined as follows:

$$\boxed{P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(X)dX} \tag{29.26}$$

An alternative definition[2] is the following:

$$f(X) = \lim_{\delta x \to 0} \frac{P(x \leq X \leq x + \delta x)}{\delta x} \tag{29.27}$$

---

[1] Also called the empirical propability.

Definition 29.3.2 (Cumulative distribution function c.d.f.). Let X be a random variable and $f(X)$ the associated p.d.f. The cumulative distribution function $F(X)$ is defined as follows:

$$F(x) = \int_{-\infty}^{x} f(X)dX \tag{29.28}$$

Theorem 29.3.3. *Let X be a random variable. Let $P(X)$ and $F(X)$ be the associated propability and c.d.f.*

$$P(x_1 \leq X \leq x_2) = F(x_2) - F(x_1) \tag{29.29}$$

Theorem 29.3.4. *$F(X)$ is continuous if and only if $P_X(\{y\}) = 0$ for every $y \in \mathbb{R}$.*

Remark 29.3.5 (Normalization).

$$F(\infty) = 1 \tag{29.30}$$

Formula 29.3.6. The pth percentile c_p can be computed as follows[3]:

$$c_p = F^{-1}(p) \tag{29.31}$$

Definition 29.3.7 (Parametric family). A parametric family of propability densities $f(X; \vec{\theta})$ is a set of densities described by one or more parameters $\vec{\theta}$.

29.3.1 Theoretical moments

Formula 29.3.8 (Moment generating function).

$$M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tX} P(X)dX \tag{29.32}$$

Theorem 29.3.9. *If the above function exists we can derive the following useful result[4] by using the series expansion 5.15:*

$$E[X^n] = \left. \frac{d^n M_X(t)}{dt^n} \right|_{t=0} \tag{29.33}$$

29.3.2 Function of a random variable

Formula 29.3.10. Let X be random variable and $f(X)$ the associated p.d.f. Let $a(X)$ be a function of X. The random variable $A = a(X)$ has an associated p.d.f. $g(A)$. If the function $a(x)$ can be inverted, then $g(A)$ can be computed as follows:

$$\boxed{g(a) = f(x(a)) \left| \frac{dx}{da} \right|} \tag{29.34}$$

---

[2]A more formal definition uses measure-theory and the Radon-Nikodym derivative.

[3]This is clear from the definition of a percentile, as this implies that $F(c_p) = p$.

[4]This property is the reason why 29.32 is called the moment generating function.

29.3.3 Multivariate distributions

Remark. In this section all defintions and thereoms will be given for bivariate distributions, but can be easily generalized to more random variables.

Definition 29.3.11 (Joint density). Let X, Y be two random variables. The joint p.d.f. $f_{XY}(x, y)$ is defined as follows:

$$f_{XY}(x, y)dxdy = \begin{cases} f_x(x \in [x, x + dx]) \\ f_y(y \in [y, y + dy]) \end{cases} \tag{29.35}$$

Remark 29.3.12. As f_{XY} is a propability density, the normalization condition 29.30 should be fulfilled.

Definition 29.3.13 (Conditional density). The conditional p.d.f. of X when Y has the value y is given by the following formula:

$$g(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)} \tag{29.36}$$

where we should pay attention to the remark made when we defined 28.40.

Corollary 29.3.14. If X and Y are independent, then by remark 28.6.6 the marginal p.d.f is equal to the conditional p.d.f.

Theorem 29.3.15 (Bayes' theorem). *The conditional p.d.f. can be computed without prior knowledge of the joint p.d.f:*

$$\boxed{g(x|y) = \frac{h(y|x)f_X(x)}{f_Y(y)}} \tag{29.37}$$

Remark. This theorem is the statistical (random variable) analogon of theorem 28.7.

Formula 29.3.16. Let $Z = XY$ with X, Y two independent random variables. The distribution $f(z)$ is given by

$$f(z) = \int_{-\infty}^{+\infty} g(x)h(z/x)\frac{dx}{|x|} = \int_{-\infty}^{+\infty} g(z/y)h(y)\frac{dy}{|y|} \tag{29.38}$$

Corollary 29.3.17. Taking the Mellin transform 9.17 of both the positive and negative part of the above integrand (to be able to handle the absolute value) gives following relation

$$\mathcal{M}\{f\} = \mathcal{M}\{g\}\mathcal{M}\{h\} \tag{29.39}$$

Formula 29.3.18. Let $Z = X + Y$ with X, Y two independent random variables. The distribution $f(z)$ is given by the convolution of $g(x)$ and $h(y)$:

$$f(z) = \int_{-\infty}^{+\infty} g(x)h(z - x)dx = \int_{-\infty}^{+\infty} g(z - y)h(y)dy \tag{29.40}$$

29.3.4   Important distributions

Formula 29.3.19 (Uniform distribution).

$$P(x; a, b) = \begin{cases} \dfrac{1}{b-a} & a \leq x \leq b \\ 0 & \text{elsewhere} \end{cases} \tag{29.41}$$

$$E(x) = \frac{a+b}{2} \tag{29.42}$$

$$V(x) = \frac{(b-a)^2}{12} \tag{29.43}$$

Formula 29.3.20 (Normal distribution). Also called the Gaussian distribution.

$$\mathcal{G}(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{29.44}$$

Formula 29.3.21 (Standard normal distribution).

$$\mathcal{N}(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tag{29.45}$$

Remark 29.3.22. Every Gaussian distribution can be rewritten as a standard normal distribution by setting $Z = \dfrac{X - \mu}{\sigma}$.

Remark 29.3.23. The c.d.f. of the standard normal distribution is given by the error function: $F(z) = \text{Erf}(z)$.

Formula 29.3.24 (Exponential distribution).

$$P(x; \tau) = \frac{1}{\tau} e^{-\frac{x}{\tau}} \tag{29.46}$$

$$E(x) = \tau \tag{29.47}$$

$$V(x) = \tau^2 \tag{29.48}$$

Theorem 29.3.25. *The exponential distribution is memoryless:*

$$P(X > x_1 + x_2 | X > x_2) = P(X > x_1) \tag{29.49}$$

Formula 29.3.26 (Bernoulli distribution). A radnom 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 p and $1 - p$, the distribution is given by:

$$\boxed{P(x; p) = p^x(1 - p)^{1-x}} \tag{29.50}$$

$$E(x) = p \tag{29.51}$$

$$V(x) = p(1 - p) \tag{29.52}$$

Formula 29.3.27 (Binomial distribution). A process with n identical independent trials, all Bernoulli processes $P(x; p)$, is described by a binomial distribution:

$$\boxed{P(r; p, n) = p^r(1 - p)^{n-r}\frac{n!}{r!(n - r)!}} \tag{29.53}$$

$$E(r) = np \tag{29.54}$$

$$V(r) = np(1 - p) \tag{29.55}$$

Formula 29.3.28 (Poisson distribution). A process with known possible outcomes but an unknown number of events is described by a Poisson distribution $P(r; \lambda)$ with λ the average expected number of events.

$$\boxed{P(r; \lambda) = \frac{e^{-\lambda}\lambda^r}{r!}} \tag{29.56}$$

$$E(r) = \lambda \tag{29.57}$$

$$V(r) = \lambda \tag{29.58}$$

Theorem 29.3.29. *If two Poisson processes $P(r; \lambda_a)$ and $P(r; \lambda_b)$ occur simultaneously and if there is no distinction between the two, then the propability of r events is also described by a Poisson distribution with average $\lambda_a + \lambda_b$.*

Corollary 29.3.30. The number of events coming from A is given by a binomial distribution $P(r_a; \Lambda_a, r)$ where $\Lambda_a = \frac{\lambda_a}{\lambda_a + \lambda_b}$.

Remark 29.3.31. For large values of λ ($\lambda \to \infty$), the Poisson distribution $P(r; \lambda)$ can be approximated by a Gaussian distribution $\mathcal{G}(x; \lambda, \sqrt{\lambda})$.

Formula 29.3.32 (χ^2 distribution). The sum of k squared independent (standard) normally distributed random variables Y_i defines the random variable:

$$\chi_k^2 = \sum_{i=1}^{k} Y_i^2 \tag{29.59}$$

where k is said to be the number of **degrees of freedom**.

$$P(\chi^2; n) = \frac{\chi^{n-2} e^{-\frac{\chi^2}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \tag{29.60}$$

Remark 29.3.33. Due to the CLT the χ^2 distribution approximates a Guassian distribution for large k: $P(\chi^2; k) \xrightarrow{k > 30} \mathcal{G}(\sqrt{2\chi^2}; \sqrt{2k-1}, 1)$

Formula 29.3.34 (Student-t distribution). The student-t distribution describes a sample with estimated standard deviation $\hat{\sigma}$.

$$P(t; n) = \frac{\Gamma\left(\dfrac{n+1}{2}\right)}{\sqrt{n\pi}\ \Gamma\left(\dfrac{n}{2}\right)\left(1 + \dfrac{t^2}{n}\right)^{\frac{n+1}{2}}} \tag{29.61}$$

where

$$t = \frac{(x - \mu)/\sigma}{\hat{\sigma}/\sigma} = \frac{z}{\sqrt{\chi^2/n}} \tag{29.62}$$

Remark. The significance of a difference between the sample mean $\overline{x}$ and the true mean μ is smaller due to the (extra) uncertainty of the estimated standard deviation.

Formula 29.3.35 (Cauchy distribution[5]). The general form $f(x; x_0, \gamma)$ is given by:

$$f(x; x_0, \gamma) = \frac{1}{\pi} \frac{\gamma}{(x - x_0)^2 + \gamma^2} \tag{29.63}$$

The characteristic function 28.22 is given by:

$$E\left(e^{itx}\right) = e^{ix_0 t - \gamma|t|} \tag{29.64}$$

Property 29.3.36. Both the mean and variance of the Cauchy distribution are undefined.

---

[5] Also known (especially in particle physics) as the **Breit-Wigner** distribution.

29.4   Central limit theorem (CLT)

Theorem 29.4.1 (Central limit theorem). *A sum of n independent random variables* X_i *has the following properties:*

1. $\mu = \sum_i \mu_i$

2. $V(X) = \sum_i V_i$

3. *The sum will be approximately (!!) normally distributed.*

Remark 29.4.2. If the random variables are not independent, property 2 will not be fulfilled.

Remark 29.4.3. The sum of Gaussians will be Gaussian to.

29.4.1   Distribution of sample mean

The difference between a sample mean $\bar{x}$ and the true mean μ is described by a distribution with following mean and variance:

Property 29.4.4.

$$< \bar{x} >= \mu \tag{29.65}$$

Property 29.4.5.

$$V(\bar{x}) = \frac{\sigma^2}{N} \tag{29.66}$$

29.5   Errors

29.5.1   Different measurement types

When performing a sequence of measurements x_i with different variances σ_i^2, it is impossible to use the arithmetic mean 29.3 in a meaningful way because the measurements are not of the same type. Therefore it is also impossible to apply the CLT 29.4.1.

Definition 29.5.1 (Weighted mean). The appropriate alternative is the *weighted mean*:

$$\boxed{\bar{x} = \frac{\sum\limits_i \dfrac{x_i}{\sigma_i^2}}{\sum\limits_i \dfrac{1}{\sigma_i^2}}} \tag{29.67}$$

The resolution of the weighted mean is given by:

$$V(\bar{x}) = \frac{1}{\sum\limits_i \sigma_i^{-2}} \tag{29.68}$$

29.5.2   Propagation of errors

Formula 29.5.2. Let X be random variable with variance $V(x)$. The variance of a linear function $f(X) = aX + b$ is given by:

$$V(f) = a^2 V(x) \tag{29.69}$$

Formula 29.5.3. Let X be random variable with **small** (!!) variance $V(x)$. The variance of a general function $f(X)$ is given by:

$$V(f) \approx \left(\frac{df}{dx}\right)^2 V(x) \tag{29.70}$$

Corollary 29.5.4. The correlation coefficient ρ (29.21) of a random variable X and a **linear** function of X is independent of σ_x and is always equal to ± 1.

Formula 29.5.5 (Law of error propagation). Let $\vec{X}$ be a set of random variables with **small** variances. The variance of a general function $f(\vec{X})$ is given by:

$$\boxed{V(f) = \sum_p \left(\frac{\partial f}{\partial X_{(p)}}\right)^2 V(X_{(p)}) + \sum_p \sum_{q \neq p} \left(\frac{\partial f}{\partial X_{(p)}}\right)\left(\frac{\partial f}{\partial X_{(q)}}\right) \mathrm{cov}(X_{(p)}, X_{(q)})} \tag{29.71}$$

Definition 29.5.6 (Fractional error). Let X, Y be two **independent** random variables. The standard deviation of $f(x, y) = xy$ is given by the fractional error:

$$\left(\frac{\sigma_f}{f}\right)^2 = \left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_y}{y}\right)^2 \tag{29.72}$$

Remark 29.5.7. The fractional error of quantity is equal to the fractional error of the reciprocal of that quantity.

Property 29.5.8. Let X be a random variable. The error of the logarithm of X is equal to the fractional error of X.

Definition 29.5.9 (Covariance of functions).

$$\mathrm{cov}(f_1, f_2) = \sum_p \sum_q \left(\frac{\partial f_1}{\partial X_{(p)}}\right)\left(\frac{\partial f_2}{\partial X_{(q)}}\right) \mathrm{cov}(X_{(p)}, X_{(q)}) \tag{29.73}$$

Corollary 29.5.10. Let $\vec{f} = \{f_1, ..., f_k\}$. The covariance matrix V_f of the k functions is given by:

$$V_f = G \ V_X \ G^T \tag{29.74}$$

where G is the Jacobian matrix of $\vec{f}$.

29.5.3   Systematic errors

Systematic errors are errors that always have the same influence (they shift all values in the same way), that are not independent of eachother and that cannot be directly inferred from the measurements.

29.6   Estimators

Definition 29.6.1 (Estimator). An estimator is a procedure that, given a sample, produces a numerical value for a property of the parent population.

29.6.1   General properties

Property 29.6.2 (Consistency).

$$\boxed{\lim_{N \to \infty} \hat{a} = a} \tag{29.75}$$

Property 29.6.3 (Unbiased estimator).

$$\boxed{< \hat{a} >= a} \tag{29.76}$$

Definition 29.6.4 (Bias).

$$B(\hat{a}) = | < \hat{a} > -a| \tag{29.77}$$

Property 29.6.5 (Efficiency). An estimator $\hat{a}$ is said to be efficient if its variance $V(\hat{a})$ is equal to the minimum variance bound 29.86.

Definition 29.6.6 (Mean squared error).

$$\Upsilon(\hat{a}) = B(\hat{a})^2 + V(\hat{a}) \tag{29.78}$$

Remark 29.6.7. If an estimator is unbiased, the MSE is equal to the variance of the estimator.

29.6.2   Fundamental estimators

Property 29.6.8 (Mean estimator). The sample mean $\bar{x}$ is a consistent and unbiased estimator for the true mean μ due to the CLT. The variance $V(\bar{x})$ of the estimator is given by equation 29.66.

Property 29.6.9 (Variance estimator for known mean). If the true mean μ is known then a consistent and unbiased estimator for the variance is given by:

$$\widehat{V(x)} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2 \tag{29.79}$$

Property 29.6.10 (Variance estimator for unknown mean). If the true mean μ is unknown and the sample mean has been used to estimate μ, then a consistent and unbiased estimator is given by[6]:

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2 \tag{29.80}$$

29.6.3   Estimation error

Formula 29.6.11 (Variance of estimator of variance).

$$V(\widehat{V(x)}) = \frac{(N-1)^2}{N^3} < (x- <x>)^4 > -\frac{(N-1)(N-3)}{N^3} < (x- <x>)^2 >^2 \tag{29.81}$$

Formula 29.6.12 (Variance of estimator of standard deviation).

$$V(\hat{\sigma}) = \frac{1}{4\sigma^2} V(\widehat{V(x)}) \tag{29.82}$$

Remark 29.6.13. 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. Or a value (hopefully close to the real one) inferred from the sample is used as an estimator or a theoretical one is used in the design phase of an experiment to see what the possible outcomes are.

29.6.4   Likelihood function

Definition 29.6.14 (Likelihood). The likelihood $\mathcal{L}(a; \vec{x})$ is the propability to find a set of measurements $\vec{x} = \{x_1, ..., x_N\}$ given a distribution $P(X; a)$:

$$\mathcal{L}(a; \vec{x}) = \prod_{i=1}^{N} P(x_i; a) \tag{29.83}$$

Definition 29.6.15 (Log-likelihood).

$$\log \mathcal{L}(a; \vec{x}) = \sum_i \ln P(x_i; a) \tag{29.84}$$

Property 29.6.16. The expectation value of an estimator ˆ is given by:

$$<\hat{a}> = \int \hat{a}\mathcal{L}(\hat{a}; X) dX \tag{29.85}$$

---

[6]The modified factor $\frac{1}{N-1}$ is called the Bessel correction. It corrects the bias of the estimator given by the sample variance 29.12. The consistency however is guaranteed by the CLT.

Theorem 29.6.17 (Minimum variance bound). *The variance of an **unbiased** estimator has a lower bound: the minimum variance bound[7] (MVB).*

$$\boxed{V(\hat{a}) \geq \frac{1}{\left\langle \left(\dfrac{d\ln \mathcal{L}}{da}\right)^2 \right\rangle}} \tag{29.86}$$

For a biased estimator with bias b the MVB takes on the following form:

$$V(\hat{a}) \geq \frac{\left(1 + \dfrac{db}{da}\right)^2}{\left\langle \left(\dfrac{d\ln \mathcal{L}}{da}\right)^2 \right\rangle} \tag{29.87}$$

Remark 29.6.18.

$$\left\langle \left(\frac{d\ln \mathcal{L}}{da}\right)^2 \right\rangle = -\left\langle \frac{d^2 \ln \mathcal{L}}{da^2} \right\rangle \tag{29.88}$$

Definition 29.6.19 (Fisher information).

$$I_X(a) = \left\langle \left(\frac{d\ln \mathcal{L}}{da}\right)^2 \right\rangle = N \int \left(\frac{d\ln P}{da}\right)^2 P dX \tag{29.89}$$

29.6.5 Maximum likelihood estimator

Following from definition 29.6.14 it follows that the best estimator $\hat{a}$ is the value for which the likelihood function is maximal. It is the value that makes the measurements the most propable, but it is therefore not the most propable estimator.

Method 29.6.20 (Maximum likelihood estimator). The maximum likelihood estimator $\hat{a}$ is obtained by solving following equation:

$$\left. \frac{d\ln \mathcal{L}}{da} \right|_{a=\hat{a}} = 0 \tag{29.90}$$

Remark 29.6.21. MLH estimators are mostly consistent but often biased.

Property 29.6.22. MLH estimators are invariant under parameter transformations.

Corollary 29.6.23. The invariance implies that the two estimators $\hat{a}$ and $\widehat{f(a)}$ cannot both be unbiased at the same time.

Property 29.6.24. Asymptotically ($N \to \infty$) every **consistent** estimator becomes unbiased and efficient.

---

[7]It is also known as the Cramer-Rao bound.

29.6.6   Least squares

Method 29.6.25 (Least squares).

1. Fitting a function $y = f(x; a)$ to a set of 2 variables (x, y) where the x values are exact and the y valiues have an uncertainty σ_i to estimate the value a.

2. For every event (x_i, y_i) define the residual $d_i = y_i - f(x_i; a)$.

3. Determine (analytically) the χ^2 value: $\chi^2 = \sum_i \left[\frac{d_i}{\sigma_i} \right]^2$

4. Find the most propably value of $\hat{a}$ by solving the equation $\frac{d\chi^2}{da} = 0$.

Property 29.6.26. The optimalized (minimized) χ^2 is distributed according to a χ^2 dsitribution 29.60 $P(\chi^2; n)$. The number of degrees of freedom n is equal to the number of events N minus the number of fitted parameters k.

Formula 29.6.27 (Linear fit). When all uncertainties σ_i are equal, the slope $\hat{m}$ and intercept $\hat{c}$ are given by following formulas:

$$\hat{m} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{\mathrm{cov}(x, y)}{V(x)} \tag{29.91}$$

$$\hat{c} = \overline{y} - \hat{m}\overline{x} = \frac{\overline{x^2} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2} \tag{29.92}$$

Remark 29.6.28. The equation $\overline{y} = \hat{c} + \hat{m}\overline{x}$ means that the linear fit passes through the center of mass $(\overline{x}, \overline{y})$.

Formula 29.6.29 (Errors of linear fit).

$$V(\hat{m}) = \frac{1}{N(\overline{x^2} - \overline{x}^2)} \sigma^2 \tag{29.93}$$

$$V(\hat{c}) = \frac{\overline{x^2}}{N(\overline{x^2} - \overline{x}^2)} \sigma^2 \tag{29.94}$$

$$\mathrm{cov}(\hat{m}, \hat{c}) = \frac{-\overline{x}}{N(\overline{x^2} - \overline{x}^2)} \sigma^2 \tag{29.95}$$

Remark 29.6.30. When there are different uncertainties σ_i, the arithmetic means have to be replaced with weighted means, but the expressions remain the same. The quantity σ^2 has to be replaced by its weighted variant:

$$\overline{\sigma^2} = \frac{\sum \sigma_i^2 / \sigma_i^2}{\sum \sigma_i^{-2}} = \frac{N}{\sum \sigma_i^{-2}}$$

29.6.7   Binned least squares

The least squares method is very useful to fit data which has been grouped in bins (histograms).

Method 29.6.31 (Binned least squares).

1. N events with distributions $P(X; a)$ divided in N_B intervals. 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 jth interval: $f_j = NW_j P(x_j; a)$

3. The real number of events has a Poisson distribution: $\bar{n}_j = \sigma_j^2 = f_j$

4. Define the binned χ^2 as: $\chi^2 = \sum\limits_{i}^{N_B} \dfrac{(n_i - f_i)^2}{f_i^2}$

29.7   Confidence

The real value of a parameter ε can never be known exactly. But it is possible to construct an interval I in which the real value should lie with a certain confidence C.

Example 29.7.1. Let X be a random variable with distribution $\mathcal{G}(x; \mu, \sigma)$. The measurement x lies in the interval $[\mu - 2\sigma; \mu + 2\sigma]$ with 95% **chance**. The real value μ lies in the interval $[x - 2\sigma; x + 2\sigma]$ with 95% **confidence**.

Remark. In the previous example there are some Bayesian assumptions: all possible values (left or right side of peak) are given the same possiblity due to the Gaussian distribution, but if one removes the symmetry it is mandatory to use a more careful approach. The symmetry between uncertainties σ and confidence levels is only valid for Gaussian distributions.

29.7.1   Interval types

Definition 29.7.2 (Two-sided confidence interval).

$$\boxed{P(x_- \le X \le x_+) = \int_{x_-}^{x_+} P(x)dx = C} \tag{29.96}$$

There are three possible (often used) two-sided intervals:

- Symmetric interval: $x_+ - \mu = \mu - x_-$

- Shortest interval: $|x_+ - x_-|$ is minimal

- Central interval: $\int_{-\infty}^{x_-} P(x)dx = \int_{x_+}^{\infty} P(x)dx = \dfrac{1 - C}{2}$

Remark 29.7.3. For Gaussian distributions these three definitions are equivalent.

Remark. The central interval is the (best and) most widely used confidence interval.

Definition 29.7.4 (One-sided confidence interval).

$$\boxed{P(x \geq x_-) = \int_{x_-}^{+\infty} P(x)dx = C}$$

(29.97)

$$\boxed{P(x \leq x_+) = \int_{-\infty}^{x_+} P(x)dx = C}$$

(29.98)

Remark 29.7.5. 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 $[x_-; x_+]$.

Definition 29.7.6 (Discrete central confidence interval).

$$x_- = \max_\theta \left[\sum_{x=0}^{\theta-1} P(x; X) \right] \leq \frac{1-C}{2}$$

(29.99)

$$x_+ = \min_\theta \left[\sum_{x=\theta+1}^{+\infty} P(x; X) \right] \leq \frac{1-C}{2}$$

(29.100)

29.7.2 General construction

For every value of the true parameter X it is possible to construct a confidence interval. This leads to the construction of the 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.

Theorem 29.7.7. *Let x_0 be the measured value of a parameter X. From the confidence region, it is possible to infere a confidence interval $[X_-(x); X_+(x)]$. 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 it is 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.*

29.7.3 Extra conditions

Method 29.7.8 (Bayesian statistics).

$$p(\text{theory}|\text{result}) = p(\text{result}|\text{theory})\frac{p(\text{theory})}{p(\text{result})}$$

(29.101)

or more mathematically:

$$\boxed{p(X|x) = p(x|X)\frac{p(X)}{p(x)}}$$

(29.102)

- The denominator $p(\text{result})$ does not play a real role, it is a normalization constant.

- The propability $p(x|X)$ to have measurement x when the true parameter is X is a Gaussian distribution $\mathcal{G}(x; X, \sigma)$

Remark 29.7.9. If nothing is known about the theory, $p(X)$ is (exaggerated assumption) a uniform propability 29.41.

29.7.4   Interval for a sample mean

Formula 29.7.10 (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 $P(-z_{\alpha/2} < Z < z_{\alpha/2})$ with $Z = \dfrac{\overline{X} - \mu}{\sigma/\sqrt{N}}$ is given by:

$$\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{N}}; \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{N}}\right] \tag{29.103}$$

Remark 29.7.11. If the sample size is not sufficiently large, the measured quantity must follow a normal distribution.

Formula 29.7.12 (Interval with unknown variance). To account for the uncertainty of the estimated standard deviation $\hat{\sigma}$, the student-t distribution 29.61 is used instead of a Gaussian distribution to describe the sample mean $\overline{X}$. The α-level confidence interval is given by:

$$\left[\overline{X} - t_{\alpha/2;(n-1)}\frac{s}{\sqrt{N}}; \overline{X} + t_{\alpha/2;(n-1)}\frac{s}{\sqrt{N}}\right] \tag{29.104}$$

where s is the estimated standard deviation 29.80.

Formula 29.7.13 (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:

$$\left[\frac{(2N\hat{P} + z_{\alpha/2}^2) - z_{\alpha/2}\sqrt{z_{\alpha/2}^2 + 4N\hat{P}(1 - \hat{P})}}{2(N + z_{\alpha/2}^2)}; \frac{(2N\hat{P} + z_{\alpha/2}^2) + z_{\alpha/2}\sqrt{z_{\alpha/2}^2 + 4N\hat{P}(1 - \hat{P})}}{2(N + z_{\alpha/2}^2)}\right]$$
$$\tag{29.105}$$

Remark. The expectation value and variance are these of a binomial distribution 29.53 with $r = X/N$.

29.7.5   Confidence region

29.8   Hypotheses and testing

29.8.1   Hypothesis

Definition 29.8.1 (Simple hypothesis). A hypothesis is called simple if the distribution is fully specified.

Definition 29.8.2 (Composite hypothesis). A hypothesis is called composite if the distribution is given relative to some parameter values.

Definition 29.8.3 (Null hypothesis H_0).

Definition 29.8.4 (Alternative hypothesis H_1).

29.8.2   Testing

Definition 29.8.5 (Type I error). Rejecting a true null hypothesis.

Definition 29.8.6 (Type II error). Accepting/retaining a false null hypothesis.

Definition 29.8.7 (Significance). The propability of making a type I error:

$$\boxed{\alpha = \int P_{H_0}(x)dx} \tag{29.106}$$

Property 29.8.8. Let $a_1 > a_2$. An a_2-level test is also significant at the a_1-level.

Remark 29.8.9. 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, they should however be controlled.

Definition 29.8.10 (Power). The propability of not making a type II error:

$$\boxed{\beta = \int P_{H_1}(x)dx \qquad \rightarrow \qquad \text{power: } 1 - \beta} \tag{29.107}$$

Theorem 29.8.11. *A good test is a test with a small significance and a large power. The propabilities $P_{H_0}(x)$ and $P_{H_1}(x)$ should be as different as possible.*

Theorem 29.8.12 (Neyman-Pearson test). *The following test is the most powerful test at significance level α for a threshold η:*

The null hypothesis H_0 is rejected in favour of the alternative hypothesis H_1 if the likelihood ratio Λ satisfies the following condition:

$$\Lambda(x) = \frac{L(x|H_0)}{L(x|H_1)} \le \eta \tag{29.108}$$

where $P(\Lambda(x) \le \eta|H_0) = \alpha$

Remark. In some references the reciprocal of $\Lambda(x)$ is used as the definition of the likelihood ratio.

29.8.3   Confindence intervals and decisions

29.9   Goodness of fit

Let $f(x|\vec{\theta})$ be the fitted function with N measurements.

29.9.1 χ^2-test

Formula 29.9.1.

$$\chi^2 = \sum_{i=1}^{N} \frac{[y_i - f(x_i)]^2}{\sigma_i^2} \tag{29.109}$$

Property 29.9.2. If there are $N - n$ fitted parameters we have:

$$\int_{\chi^2}^{\infty} f(\chi^2|n)d\chi^2 \approx 1 \implies \begin{cases} \circ \text{ good fit} \\ \circ \text{ errors were overestimated} \\ \circ \text{ selected measurements} \\ \circ \text{ lucky shot} \end{cases} \tag{29.110}$$

Property 29.9.3 (Reduced chi-squared $\chi^2_{\mathbf{red}}$). Define the reduced chi-squared value as follows: $\chi^2_{\text{red}} = \chi^2/n$ where n is the number of degrees of freedom.

- $\chi^2_{\text{red}} >> 1$: Poor modelling.
- $\chi^2_{\text{red}} > 1$: Bad modelling or underestimation of the uncertainties.
- $\chi^2_{\text{red}} = 1$: Good fit.
- $\chi^2_{\text{red}} < 1$: Impropable, overestimation of the uncertainties.

29.9.2   Runs test

A good χ^2-test does not mean that the fit is good. As mentioned in property 29.110 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. (It is a result of the 'randomness' of a data sample') This condition is tested with a runs test 29.111/29.112.

Remark 29.9.4. 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 29.9.5 (Runs distribution).

$$\boxed{P(r_{even}) = 2\frac{C_{\frac{r}{2}-1}^{N_B-1}C_{\frac{r}{2}-1}^{N_O-1}}{C_{N_B}^N}} \tag{29.111}$$

$$\boxed{P(r_{odd}) = \frac{C_{\frac{r-3}{2}}^{N_B-1}C_{\frac{r-1}{2}}^{N_O-1} + C_{\frac{r-3}{2}}^{N_O-1}C_{\frac{r-1}{2}}^{N_B-1}}{C_{N_B}^N}} \tag{29.112}$$

$$E(r) = 1 + 2\frac{N_B N_O}{N} \tag{29.113}$$

$$V(r) = 2\frac{N_B N_O}{N}\frac{2N_B N_O - 1}{N(N-1)} \tag{29.114}$$

Remark 29.9.6. For $r > 10 - 15$ the runs distribution approximates a Gaussian distribution.

29.9.3   Kolmogorov test

Definition 29.9.7 (Empirical distribution function).

$$F_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{]-\infty,x]}(x_i) \tag{29.115}$$

where $\mathbb{1}_A(x)$ is the indicator function 8.19.

Definition 29.9.8 (Kolmogorov-Smirnov statistic). Let $F(x)$ be a given cumulative distribution function. The nth Kolmogorov-Smirnov statistic is defined as:

$$D_n = \sup_x |F_n(x) - F(x)| \tag{29.116}$$

Definition 29.9.9 (Kolmogorov distribution).

$$P(K \le x) = 1 - 2\sum_{i=1}^{+\infty}(-1)^{i-1}e^{-2i^2x^2} = \frac{\sqrt{2\pi}}{x}\sum_{i=1}^{+\infty}e^{-(2i-1)^2\pi^2/(8x^2)} \tag{29.117}$$

Property 29.9.10 (Kolmogorov-Smirnov test). Let the null hypothesis H_0 state that a given data sample is described by a continuous distribution $P(x)$ with cumulative distribution function $F(x)$. The null hypothesis is rejected at significance level α if:

$$D_n\sqrt{n} > K_\alpha \tag{29.118}$$

where K_α is defined by using the Kolmogorov distribution: $P(K \leq K_\alpha) = 1 - \alpha$

Chapter 30

Stochastic calculus

Part VII

Classical Mechanics

Chapter 31

Equations of motion

31.1 General quantities

31.1.1 Linear quantities

Formula 31.1.1 (Force).

$$\vec{F} = \frac{d\vec{p}}{dt} \tag{31.1}$$

Remark. In classical mechanics, this formula is given by Newton's second law.

Formula 31.1.2 (Work).

$$W = \int \vec{F} \cdot d\vec{l} \tag{31.2}$$

Definition 31.1.3 (Conservative force). If the work done by a force is independent of the path taken, the force is said to be **conservative**.

$$\oint_C \vec{F} \cdot d\vec{l} = 0 \tag{31.3}$$

According to 15.13 together with Stokes' theorem 15.20 it can be written as the gradient of a scalar field.

$$\vec{F} = -\nabla V \tag{31.4}$$

Formula 31.1.4 (Kinetic energy).

$$E_{kin} = \frac{1}{2}mv^2 = \frac{p^2}{2m} \tag{31.5}$$

31.1.2 Angular quantities

Formula 31.1.5 (Angular velocity).

$$\omega = \frac{v}{r} \tag{31.6}$$

Formula 31.1.6 (Angular frequency).

$$\nu = \frac{\omega}{2\pi} \tag{31.7}$$

Formula 31.1.7 (Moment of inertia).

$$I = \int_V r^2 \rho(r) dV \tag{31.8}$$

or more generally we can define the moment of inertia tensor:

$$\boxed{\mathcal{I} = \int_V \rho(\vec{r}) \left(r^2 \mathbb{1} - \vec{r} \otimes \vec{r} \right) dV} \tag{31.9}$$

Definition 31.1.8 (Principal axes of inertia). Let $[I]$ be the matrix of inertia. This is a real symmetric matrix, which means that it admits an eigendecomposition of the form (see 14.6.10):

$$[I] = [Q][\Lambda][Q]^T$$

The columns of $[Q]$ are called the principal axes of inertia. The eigenvalues are called the **principal moments of inertia**.

Example 31.1.9 (Objects with azimuthal symmetry†). Let r denote the radius of the object.

- Solid disk: $I = \frac{1}{2}mr^2$
- Cylindrical shell: $I = mr^2$
- Hollow sphere: $I = \frac{2}{3}mr^2$
- Solid sphere: $I = \frac{2}{5}mr^2$

Theorem 31.1.10 (Parallel axis theorem). *Consider a rotation about an axis ω through a point A. Let ω_{CM} be a parallel axis through the center of mass. The moment of inertia about ω is related to the moment of inertia about ω_{CM} in the following way:*

$$\boxed{I_A = I_{CM} + M||\vec{r}_A - \vec{r}_{CM}||^2} \tag{31.10}$$

where M is the mass of the rotating body.

Formula 31.1.11 (Angular momentum).

$$\boxed{\vec{L} = \vec{r} \times \vec{p}} \tag{31.11}$$

Given the angular velocity vector we can compute the angular momentum as follows:

$$\vec{L} = \mathcal{I}(\vec{\omega}) \tag{31.12}$$

where $\mathcal{I}$ is the moment of inertia tensor. If $\vec{\omega}$ is parallel to a principal axis, then the formula reduces to:

$$\vec{L} = I\vec{\omega} \tag{31.13}$$

Formula 31.1.12 (Torque).

$$\vec{\tau} = \frac{d\vec{L}}{dt} \tag{31.14}$$

For constant bodies, this formula can be rewritten as follows:

$$\vec{\tau} = I\vec{\alpha} = \vec{r} \times \vec{F} \tag{31.15}$$

Remark 31.1.13. From the previous definitions it follows that both the angular momentum and torque vectors are in fact pseudo-vectors and thus change sign under coordinate transforms with $\det = -1$.

Formula 31.1.14 (Rotational energy).

$$E_{\text{rot}} = \frac{1}{2}I\omega^2 \tag{31.16}$$

31.2   Central force

Definition 31.2.1 (Central force). A central force is a force that only depends on the relative position of two objects:

$$\vec{F}_c \equiv F\left(||\vec{r}_2 - \vec{r}_1||\right)\hat{e}_r \tag{31.17}$$

31.3   Kepler problem

31.3.1   Gravity

Formula 31.3.1 (Potential for a point mass).

$$\boxed{V = -G\frac{M}{r}} \tag{31.18}$$

where $G = 6.67 \times 10^{-11} \frac{Nm^2}{\text{kg}^2}$ is the **gravitational constant**.

31.4 Harmonic oscillator

Formula 31.4.1 (Harmonic potential).

$$\boxed{V = \frac{1}{2}kx^2}$$
(31.19)

or
(31.20)

$$V = \frac{1}{2}m\omega^2 x^2$$
(31.21)

where we have used: $\omega = \sqrt{\frac{k}{m}}$.

Formula 31.4.2 (Solution).

$$x(t) = A\sin\omega t + B\cos\omega t$$
(31.22)
$$= Ce^{i\omega t} + De^{-i\omega t}$$
(31.23)

Chapter 32

Lagrangian and Hamiltonian formalism

Definition 32.0.1 (Generalized coordinates). The generalized coordinates q_k are independent coordinates that completely describe the current configuration of a system relative to a reference configuration.

When a system has N degrees of freedom and n_c constraints, there are $(N - n_c)$ generalized coordinates. Furthermore, every set of generalized coordinates, describing the same system, should contain exactly $(N - n_c)$ coordinates.

Definition 32.0.2 (Generalized velocities). The generalized velocities $\dot{q}_k$ are the derivatives of the generalized coordinates with respect to time.

Notation 32.0.3.

$$L\left(\vec{q}(t), \dot{\vec{q}}(t), t\right) \equiv L\left(q_1(t), ..., q_n(t), \dot{q}_1(t), ..., \dot{q}_n(t), t\right) \tag{32.1}$$

Definition 32.0.4 (Action).

$$\boxed{S = \int_{t_1}^{t_2} L\left(\vec{q}(t), \dot{\vec{q}}(t), t\right) dt} \tag{32.2}$$

32.1 Euler-Lagrange equations[†]

Formula 32.1.1 (Euler-Lagrange equation of the first kind).

$$\boxed{\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}^k}\right) - \frac{\partial T}{\partial q^k} = Q_k} \tag{32.3}$$

where T is the total kinetic energy.

Formula 32.1.2 (Euler-Lagrange equation of the second kind).

$$\boxed{\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^k}\right) - \frac{\partial L}{\partial q^k} = 0} \tag{32.4}$$

32.2 Conservation laws and symmetry properties

Definition 32.2.1 (Conjugate momentum). Also called the **canonically conjugate momentum**.

$$p_k = \frac{\partial L}{\partial \dot{q}^k} \tag{32.5}$$

Definition 32.2.2 (Cyclic coordinate). If the lagrangian L does not explicitly depend on a coordinate q_k, the coordinate is called a cyclic coordinate.

Property 32.2.3. The conjugate momentum of a cyclic coordinate is a conserved quantity.

$$\dot{p}_k \overset{32.5}{=} \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^k}\right) \overset{32.4}{=} \frac{\partial L}{\partial q^k} \overset{\text{cyclic}}{\underset{\text{coord.}}{=}} 0 \tag{32.6}$$

32.3 Hamilton's equations

Definition 32.3.1 (Canonical coordinates). Consider the generalized coordinates $(q, \dot{p}, t)$ from the Lagrangian formalism. Using these we can define a new set of coordinates, called canonical coordinates, by exchanging the time-derivatives $\dot{q}^i$ in favour of the conjugate momenta p_i (see definition 32.5) and leaving the coordinates q^i and t invariant.

Definition 32.3.2 (Hamiltonian function). The (classical) Hamiltonian function is defined as follows:

$$H(q, p, t) = \sum_i p_i \dot{q}^i - L(q, p, t) \tag{32.7}$$

Formula 32.3.3 (Hamilton's equations[1]).

$$\boxed{\begin{aligned} \dot{q}^i &= \frac{\partial H}{\partial p_i} \\[2mm] -\dot{p}_i &= \frac{\partial H}{\partial q^i} \\[2mm] -\frac{\partial L}{\partial t} &= \frac{\partial H}{\partial t} \end{aligned}}$$

$$\tag{32.8}$$
$$\tag{32.9}$$
$$\tag{32.10}$$

The formula to obtain the Hamiltonian from the Lagrangian is an application of the following more general Legendre transformation:

---

[1]Also known as the *canonical equations of Hamilton.*

Definition 32.3.4 (Legendre transformation). Consider an equation of the following form:

$$df = udx + vdy \tag{32.11}$$

where $u = \frac{\partial f}{\partial x}$ and $v = \frac{\partial f}{\partial y}$.

Suppose we want to perform a coordinate transformation $(x, y) \to (u, y)$ while preserving the general form of 32.11 for differential quantities. To do this we consider the function

$$g = f - ux \tag{32.12}$$

Differentiating gives

$$\begin{aligned}
dg &= df - udx - xdu \\
&= (udx + vdy) - udx - xdu \\
&= vdy - xdu
\end{aligned}$$

which has the form of 32.11 as desired. The quantities v and x are now given by

$$x = -\frac{\partial g}{\partial u} \qquad \text{and} \qquad v = \frac{\partial g}{\partial y} \tag{32.13}$$

The transition $f \to g$ defined by equations 32.11 and 32.12 is called a Legendre transformation.

Remark 32.3.5. Although the previous derivation used only 2 coordinates, the definition of Legendre transformations can easily be generalized to more coordinates.

32.4 Hamilton-Jacobi equation

32.4.1 Canonical transformations

Definition 32.4.1 (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:

$$\sum_i \dot{q}^i p_i - H(q, p, t) = \sum_i \dot{Q}^i P_i - K(Q, P, t) - \frac{dG}{dt}(Q, P, t) \tag{32.14}$$

The function G is called the generating function of the canonical transformation. The choice of G uniquely determines the transformation.

Formula 32.4.2 (Hamilton-Jacobi equation). Sufficient conditions for the generating function S are given by:

$$P_i = \frac{\partial S}{\partial Q^i}$$
$$Q^i = \frac{\partial S}{\partial P_i}$$

and
$$K = H + \frac{\partial S}{\partial t}$$
Choosing the new Hamiltonian function K to be 0 gives the Hamilton-Jacobi equation:

$$\boxed{H\left(q, \frac{\partial S}{\partial q}\right) + \frac{\partial S}{\partial t} = 0} \tag{32.15}$$

The function S is called **Hamilton's principal function**.

Property 32.4.3. The new coordinates P_i and Q^i are all constants of motion. This follows immediately from the choice $K = 0$.

Definition 32.4.4 (Hamilton's characteristic function). If the system is time-independent it follows from the HJE that the principal function is of the form

$$S(q, p, t) = W(q, p) - Et \tag{32.16}$$

where E is a constant. The time-independent function W is called Hamilton's characteristic function.

Substituting this result in the HJE results in

$$H\left(q, \frac{\partial S}{\partial q}\right) = E \tag{32.17}$$

In time-independent systems the Hamiltonian function is thus a constant of motion and we call it the **energy** of the system.

32.4.2   Stäckel potential

Remark 32.4.5. If the principal function can be separated into n equations, the HJE splits up into n equations of the form

$$h_i\left(q^i, \frac{dS}{dq^i}, \alpha_i\right) = 0 \tag{32.18}$$

The partial differential equation for S can thus be rewritten as a system of n ordinary differential equations.

Theorem 32.4.6 (Stäckel condition). *Using an orthogonal coordinate system, the Hamilton-Jacobi equation is separable if and only if the potential is of the following form:*

$$V(q) = \sum_{i=1}^{n} \frac{1}{G_i^2(q)} W_i(q^i) \tag{32.19}$$

whenever the Hamiltonian function can be written as

$$H(q, p) = \frac{1}{2} \sum_i \frac{p_i^2}{G_i^2(q)} + V(q) \tag{32.20}$$

*These potentials are called **Stäckel potentials**.*

Chapter 33

Phase space

33.1 Phase space

Definition 33.1.1 (Phase space). The set of all possible n-tuples[1] (q^i, p_i) of generalized coordinates and associated momenta is called the phase space of the system.

Definition 33.1.2 (Rotation). A rotation is the change of a coordinate for which every possible value is allowed.

Definition 33.1.3 (Libration). A libration is the change of coordinate for which only a subset of the total range is allowed. It is the generalization of an oscillation.

33.2 Material derivative

Definition 33.2.1 (Lagrangian derivative[2]). Let $a(\vec{r}, \vec{v}, t)$ be a property of a system defined at every point of the system. The Lagrangian derivative along a path $(\vec{r}(t), \vec{v}(t))$ in phase space is given by:

$$
\begin{aligned}
\frac{Da}{Dt} &= \lim_{\Delta t \to 0} \frac{a(\vec{r} + \Delta\vec{r}, \vec{v} + \Delta\vec{v}, t + \Delta t) - a(\vec{r}, \vec{v}, t)}{\Delta t} \\
&= \frac{\partial a}{\partial t} + \frac{d\vec{r}}{dt} \cdot \frac{\partial a}{\partial \vec{r}} + \frac{d\vec{v}}{dt} \cdot \frac{\partial a}{\partial \vec{v}} \\
&= \frac{\partial a}{\partial t} + \vec{v} \cdot \nabla a + \frac{d\vec{v}}{dt} \cdot \frac{\partial a}{\partial \vec{v}}
\end{aligned}
\tag{33.1}
$$

The second term $\vec{v} \cdot \nabla a$ in this equation is called the **advective** term.

Remark 33.2.2. In the case that $a(\vec{r}, \vec{v}, t)$ is a tensor field the gradient ∇ has to be replaced by the covariant derivative. The advective term is then called the **convective** term.

---

[1] Not only those as given by the equations of motion.
[2] Also known as the **material derivative**, especially when applied to fluidum mechanics.

Corollary 33.2.3. If we take $a(\vec{r}, \vec{v}, t) = \vec{r}$ we obtain:

$$\frac{D\vec{r}}{Dt} = \vec{v} \tag{33.2}$$

33.3 Liouville's theorem

Formula 33.3.1 (Liouville's lemma). Consider a phase space volume element dV_0 moving along a path $(\vec{r}(t), \vec{v}(t)) \equiv (\vec{x}(t))$. The Jacobian $J(\vec{x}, t)$ associated with this motion is given by:

$$J(\vec{x}, t) = \frac{dV}{dV_0} = \det\left(\frac{\partial \vec{x}}{\partial \vec{x}_0}\right) = \sum_{ijklmn} \varepsilon_{ijklmn} \frac{\partial x^1}{\partial x_0^i} \frac{\partial x^2}{\partial x_0^j} \frac{\partial x^3}{\partial x_0^k} \frac{\partial x^4}{\partial x_0^l} \frac{\partial x^5}{\partial x_0^m} \frac{\partial x^6}{\partial x_0^n} \tag{33.3}$$

The Lagrangian derivative of this Jacobian then becomes:

$$\frac{DJ}{Dt} = (\nabla \cdot \vec{x}) J \tag{33.4}$$

Furthermore using the Hamiltonian equations 32.8 it is easy to prove that

$$\nabla \cdot \vec{x} = 0 \tag{33.5}$$

Theorem 33.3.2 (Liouville's theorem). *Let $V(t)$ be a phase space volume containing a fixed set of particles. Application of Liouville's lemma gives:*

$$\frac{DV}{Dt} = \frac{D}{Dt} \int_{\Omega(t)} d^6 x = \frac{D}{Dt} \int_{\Omega_0} J(\vec{x}, t) d^6 x_0 = 0 \tag{33.6}$$

It follows that the phase space volume of a Hamiltonian system[3] is invariant with respect to time-evolution.

Formula 33.3.3 (Boltzmann's transport equation). Let $F(\vec{r}, \vec{v}, t)$ be the mass distribution function:

$$M_{tot} = \int_{\Omega(t)} F(\vec{x}, t) d^6 x \tag{33.7}$$

From the conservation of mass we can derive the following formula:

$$\frac{DF}{Dt} = \frac{\partial F}{\partial t} + \frac{d\vec{r}}{dt} \cdot \frac{\partial F}{\partial \vec{r}} - \nabla V \cdot \frac{\partial F}{\partial \vec{v}} = 0 \tag{33.8}$$

This formula is a partial differential equation in 7 variables which can be solved to obtain $F(\vec{x}, t)$.

Theorem 33.3.4 (Poincaré recurrence theorem). *Consider a Hamiltonian system with a finite phase space $\mathcal{V}$ (for example when the system is trapped in a potential well). By Liouville's theorem, the phase flow generated by the equations of motion is a volume preserving map $g : \mathcal{V} \to \mathcal{V}$. Let $\mathcal{V}_0$ be the 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 $n \in \mathbb{N}$.*

---

[3] A system that satisfies Hamilton's equations of motion.

33.4 Continuity equation

Formula 33.4.1 (Reynolds transport theorem[4]). Consider a quantity

$$F = \int_{V(t)} f(\vec{r}, \vec{v}, t) dV$$

Using equation 33.4 and the divergence theorem 15.21 we can obtain:

$$\boxed{\frac{DF}{Dt} = \int_V \frac{\partial f}{\partial t} dV + \oint_S f\vec{v} \cdot d\vec{S}} \tag{33.9}$$

Formula 33.4.2 (Continuity equations). For a conserved quantity the equation above becomes:

$$\frac{Df}{Dt} + (\nabla \cdot \vec{v})f = 0 \tag{33.10}$$

$$\frac{\partial f}{\partial t} + \nabla \cdot (f\vec{v}) = 0 \tag{33.11}$$

If we set $f = \rho$ (mass density) then 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 33.9.

Corollary 33.4.3. Combining the Reynolds transport theorem with the Lagrangian continuity equation gives the following identity for an arbitrary function f:

$$\frac{D}{Dt} \int_V \rho f dV = \int_V \rho \frac{Df}{Dt} dV \tag{33.12}$$

---

[4]This is a 3D extension of the *Leibniz integral rule*.

Chapter 34

Fluid mechanics

34.1 Cauchy stress tensor

Theorem 34.1.1 (Cauchy's stress theorem[1]). *Knowing the stress vectors acting on the coordinate planes through a point A is sufficient to calculate the stress vector acting on an arbitrary plane passing through A.*

The *Cauchy stress theorem* is equivalent to the existence of the following tensor:

Definition 34.1.2 (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 $\vec{\boldsymbol{n}}$ to that plane:

$$\vec{\boldsymbol{t}}_{(\vec{\boldsymbol{n}})} = \mathbf{T}(\vec{\boldsymbol{n}}) \tag{34.1}$$

Example 34.1.3. For identical particles, the stress tensor is given by:

$$\mathbf{T} = -\rho \langle \vec{\boldsymbol{w}} \otimes \vec{\boldsymbol{w}} \rangle \tag{34.2}$$

where $\vec{\boldsymbol{w}}$ is the random component of the velocity vector and $\langle \cdot \rangle$ denotes the expectation value (see 28.19).

Theorem 34.1.4 (Cauchy's lemma). *The stress vectors acting on opposite planes are equal in magnitude but opposite in direction:*

$$\vec{\boldsymbol{t}}_{(-\vec{\boldsymbol{n}})} = -\vec{\boldsymbol{t}}_{(\vec{\boldsymbol{n}})} \tag{34.3}$$

Formula 34.1.5 (Cauchy momentum equation). From Newton's second law 31.1 it follows that:

$$\frac{D\vec{\boldsymbol{P}}}{Dt} = \int_V \vec{\boldsymbol{f}}(\vec{\boldsymbol{x}}, t)dV + \oint_S \vec{\boldsymbol{t}}(\vec{\boldsymbol{x}}, t)dS \tag{34.4}$$

---

[1]Also known as **Cauchy's fundamental theorem**.

where $\vec{P}$ is the momentum density, $\vec{f}$ are body forces and $\vec{t}$ are surface forces (such as shear stress). Using Cauchy's stress theorem and the divergence theorem 15.21 we get

$$\frac{D\vec{P}}{Dt} = \int_V \left[\vec{f}(\vec{x},t) + \nabla \cdot \mathbf{T}(\vec{x},t) \right] dV \tag{34.5}$$

The left-hand side can be rewritten using 33.12 as

$$\int_V \rho \frac{D\vec{v}}{Dt} dV = \int_V \left[\vec{f}(\vec{x},t) + \nabla \cdot \mathbf{T}(\vec{x},t) \right] dV \tag{34.6}$$

Chapter 35

Optics

35.1 General

35.1.1 Conservation of energy

From the law of conservation of energy we can derive the following formula:

$$\boxed{T + R + A = 1} \tag{35.1}$$

where

T : Transmission coefficient

R : Reflection coefficient

A : Absorption coefficient

35.1.2 Photon

Formula 35.1.1 (Energy).

$$E = h\nu = \hbar\omega = \frac{hc}{\lambda} \tag{35.2}$$

Formula 35.1.2 (Momentum).

$$p = \frac{h}{\lambda} = \hbar k \tag{35.3}$$

where formula 35.4 was used in the last step.

Remark. These formulas can also be (approximately) used for particles for which the rest mass (energy) is negligible.

35.2 Plane wave

Formula 35.2.1 (Wave number).

$$k = \frac{2\pi}{\lambda} \tag{35.4}$$

Formula 35.2.2 (Plane wave). Following equations represent a plane wave moving in the x-direction and polarized in the xy-plane:

$$\vec{E}(x,t) = \mathrm{Re}\left\{A \exp\left[i\left(kx - \omega t + \phi\right)\right]\right\} \vec{e}_y \tag{35.5}$$

$$\vec{E}(x,t) = \mathrm{Re}\left\{A \exp\left[2\pi i\left(\frac{x}{\lambda} - \frac{t}{T} + \frac{\phi}{2\pi}\right)\right]\right\} \vec{e}_y \tag{35.6}$$

35.3 Refraction

Formula 35.3.1 (Refraction).

$$v_2 = \frac{v_1}{n} \tag{35.7}$$

Formula 35.3.2 (Diëlectric function). In the case of non-magnetic materials ($\mu_r \approx 1$), we can write the diëlectric function as following:

$$\epsilon = \epsilon_r + i\epsilon_i = \widetilde{n}^2 = (n + ik)^2 \tag{35.8}$$

Where $\widetilde{n}$ is the **complex refractive index** and k is the **extinction coefficient**.

35.4 Absorption

Theorem 35.4.1 (Law of Lambert-Beer[†]).

$$\frac{I(x)}{I(0)} = exp\left(-\frac{4\pi\nu k}{c}x\right) \tag{35.9}$$

Definition 35.4.2 (Absorption coefficient). The constant factor in the Lambert-Beer law is called the absorption coefficient.

$$\alpha = \frac{4\pi\nu k}{c} \tag{35.10}$$

35.5 Diffraction

Chapter 36

Astronomy

36.1 Ellipsoidal coordinates

We start from folowing parametrized equation:

$$f(\tau) = \frac{x^2}{\tau + \alpha} + \frac{y^2}{\tau + \beta} + \frac{z^2}{\tau + \gamma} - 1 \tag{36.1}$$

where $\alpha < \beta < \gamma < 0$. By multiplying away the denominators and setting $f(\tau) = 0$ we obtain a polynomial equation of degree 3 in τ. This polynomial can be formally factorised as

$$-(\tau - \lambda)(\tau - \mu)(\tau - \nu) \tag{36.2}$$

such that the solutions (λ, μ, ν) obey following rules:

$$\begin{cases} \nu & \in \quad]-\gamma, -\beta[\\ \mu & \in \quad]-\beta, -\alpha[\\ \lambda & \in \quad]-\alpha, +\infty[\end{cases}$$

From previous two equations we can find a solution for x^2 by multiplying by $(\tau + \alpha)$ and letting $\tau \to -\alpha$. Solutions for y^2 and z^2 can be found in a similar way:

$$\begin{cases} x^2 & = \quad \dfrac{(\lambda + \alpha)(\mu + \alpha)(\nu + \alpha)}{(\beta - \alpha)(\gamma - \alpha)} \\ y^2 & = \quad \dfrac{(\lambda + \beta)(\mu + \beta)(\nu + \beta)}{(\beta - \alpha)(\beta - \gamma)} \\ z^2 & = \quad \dfrac{(\lambda + \gamma)(\mu + \gamma)(\nu + \gamma)}{(\alpha - \gamma)(\beta - \gamma)} \end{cases} \tag{36.3}$$

For these solutions multiple cases can be considered. We can define different surfaces by fixing τ at different values.

36.1.1 Ellipsoid: $\tau = \lambda$

First we look at the surfaces defined by fixing $\tau = \lambda$ in equation 36.1. By noting that all denominators are positive in this case, we see that the obtained surface is an ellipsoid with the x-axis as the shortest axis. By letting $\lambda \to +\infty$ we obtain the equation of a sphere with radius $\sqrt{\lambda}$. If $\lambda \to -\alpha$ we get an ellipse in the yz-plane. This ellipse is called the **focal ellipse**.

36.1.2 One-sheet hyperboloid: $\tau = \mu$

By fixing $\tau = \mu$ in 36.1 we obtain the equation of one-sheet hyperboloid (also called a **hyperbolic hyperboloid**) around the x-axis. By letting $\mu \to -\alpha$ the hyperboloid collapses in the yz-plane and we obtain the surface outside the focal ellipse. If $\mu \to -\beta$ the hyperboloid becomes degenerate and we get the surface inside the **focal hyperbola** defined by

$$\frac{x^2}{\alpha - \beta} + \frac{z^2}{\gamma - \beta} = 1 \tag{36.4}$$

This hyperbola intersects the z-plane in the foci of the focal ellipse.

36.1.3 Two-sheet hyperboloid: $\tau = \nu$

By fixing $\tau = \nu$ in 36.1 we obtain the equation of two-sheet hyperboloid (also called an **elliptic hyperboloid**) around the z-axis. By letting $\nu \to -\beta$ the hyperboloid becomes degenerate and we obtain the surface outside the focal hyperbola 36.4. If $\nu \to -\gamma$ the two sheets coincide in the xy-plane.

36.1.4 Hamiltonian function

When wrting out the kinetic energy in ellipsoidal coordinates by applying the chain rule for differentiation to the Cartesian kinetic energy while noting that mixed terms of the form $\frac{\partial x^a}{\partial \lambda^i} \frac{\partial x^a}{\partial \lambda^j}$ cancel out when writing them out using 36.3 it is clear that the Hamiltonian function can be spearated as follows:

$$H = \frac{1}{2} \left(\frac{p_\lambda^2}{Q_\lambda^2} + \frac{p_\mu^2}{Q_\mu^2} + \frac{p_\nu^2}{Q_\nu^2} \right) + V \tag{36.5}$$

where $Q_j^2 = \sum_i \left(\frac{\partial x^i}{\partial \lambda^j} \right)^2$ are the metric coefficients in ellipsoidal coordinates.

These coefficients can be calculated by noting that $\frac{\partial x^i}{\partial \lambda} = \frac{1}{x^i} \frac{\partial (x^i)^2}{\partial \lambda}$ and putting $\frac{1}{(\lambda+\alpha)(\lambda+\beta)(\lambda+\gamma)}$ in the front. Furthermore the coefficient belonging to λ^2, μ^2, ν^2, mixed terms and others can

be calculated easily. By doing so we obtain following result

$$Q_\lambda^2 = \frac{1}{4} \frac{(\lambda - \mu)(\lambda - \nu)}{(\lambda + \alpha)(\lambda + \beta)(\lambda + \gamma)} \tag{36.6}$$

which is also valid for μ and ν by applying cyclic permutation to the coordinates.

Following from the Stäckel conditions 32.19 the potential must be of the form

$$V = \sum_i \frac{W_i(\lambda^i)}{Q_i^2} \tag{36.7}$$

if we want to obtain a seperable Hamilton-Jacobi equation. Due to the disjunct nature of λ, μ and ν we can consider W_λ, W_μ and W_ν as three parts of a single function $G(\tau)$ given by:

$$G(\tau) = -4(\tau + \beta)W_\tau(\tau) \tag{36.8}$$

The 3D potential is thus completely determined by a 1D function $G(\tau)$.

36.1.5 Hamilton-Jacobi equation

If we consider a time-independent system we can use 32.17 as our starting point. If we multiply this equation by $(\lambda - \mu)(\lambda - \nu)(\mu - \nu)$ we obtain

$$(\mu - \nu)\left[2(\lambda + \alpha)(\lambda + \beta)(\lambda + \gamma)\left(\frac{dS^\lambda(\lambda)^2}{d\lambda}\right)\right.$$
$$\left. - (\lambda + \alpha)(\lambda + \gamma)G(\lambda) - \lambda^2 E\right] + \text{cyclic permutations} = 0 \quad (36.9)$$

where we rewrote the multiplication factor in the form $a\lambda^2 + b\mu^2 + c\nu^2$ before multiplying the RHS of 32.17. This equation can be elegantly rewritten as

$$(\mu - \nu)U(\lambda) + (\lambda - \mu)U(\nu) + (\nu - \lambda)U(\mu) = 0 \tag{36.10}$$

Differentiating twice with respect to any λ^i gives $U''(\tau) = 0$ or equivalently

$$U(\tau) = I_3 - I_2\tau \tag{36.11}$$

where I_2 and I_3 are two new first integrals of motion.

From the Hamiltonian-Jacobi equations of motion one can calculate the conjugate momenta $p_\tau = \frac{dS^\tau}{d\tau}$. After a lengthy calculation we obtain

$$p_\tau^2 = \frac{1}{2(\tau + \beta)} \left[E - V_{\text{eff}}(\tau)\right] \tag{36.12}$$

where the effective potential is given by

$$\boxed{V_{\text{eff}} = \frac{J}{\tau + \alpha} + \frac{K}{\tau + \gamma} - G(\tau)} \tag{36.13}$$

where J and K are two conserved quantities given by

$$J = \frac{\alpha^2 E + \alpha I_2 + I_3}{\alpha - \gamma} \quad \text{and} \quad K = \frac{\gamma^2 E + \gamma I_2 + I_3}{\gamma - \alpha}$$

To be physically acceptable, p_τ^2 should be positive. This leads to following conditions on the energy:

$$\begin{cases} E \geq V_{\text{eff}}(\lambda) \\ E \geq V_{\text{eff}}(\mu) \\ E \leq V_{\text{eff}}(\nu) \end{cases} \tag{36.14}$$

The generating $G(\tau)$ function should also satisfy some conditions. First we note that we can rewrite our Stäckel potential $V(\lambda, \mu, \nu)$ as

$$V = -\frac{1}{\lambda - \nu} \left(\frac{F(\lambda) - F(\mu)}{\lambda - \mu} - \frac{F(\mu) - F(\nu)}{\mu - \nu} \right) \leq 0 \tag{36.15}$$

where $F(\tau) = (\tau + \alpha)(\tau + \gamma)G(\tau)$.

For $\lambda \to +\infty$ (or $r^2 \to +\infty$) we get $V \approx -\frac{F(\lambda)}{\lambda^2} \approx -G(\lambda)$. Because $V \sim \lambda^{-1}$ it is clear that $G(\tau)$ cannot decay faster than $\lambda^{-1/2}$ at infity. Furthermore we can interpret 36.15 as an approximation of $-F''(\tau)$. So it follows that $F(\tau)$ should be convex. For $\tau \to -\gamma$ we get

$$\begin{cases} \alpha + \tau < 0 \\ \tau + \gamma \to 0 \end{cases}$$

So if $G(\tau)$ decays faster than $\dfrac{1}{\tau + \gamma}$ then $F(\tau) \to -\infty$ which is not possible for a convex function.

To fullfil these conditions we assume that the generating function can be written as

$$\boxed{G(\tau) = \frac{GM}{\sqrt{\gamma_0 + \tau}}} \tag{36.16}$$

where G is the gravitational constant and M is the galactic mass.

Theorem 36.1.1 (Kuzmin's theorem). *The spatial mass density function generated by a Stäckel potential is completely determined by a function of the form $\rho(z)$.*

Corollary 36.1.2. For triaxial mass models in ellipsoidal coordinates the axial ratios are inversely proportional to the axial ratios of the coordinate system.

Part VIII

Electromagnetism

Chapter 37

Electricity

37.1 Resistance R

37.1.1 Conductivity

Definition 37.1.1 (Drift velocity). The average speed of the independent charge carriers is the drift velocity $\vec{v_d}$. It is important to remark that v_d is not equal to the propagation speed of the electricity[1].

Formula 37.1.2 (Conductivity).

$$\sigma = nq\mu \tag{37.1}$$

Formula 37.1.3 (Resistivity).

$$\rho = \frac{1}{\sigma} \tag{37.2}$$

Formula 37.1.4 (Mobility).

$$\mu = \frac{v_d}{E} \tag{37.3}$$

37.1.2 Current density

Formula 37.1.5. Let A be the cross section of a conductor. Let $\vec{J}$ be the current density though A. The current trough A is then given by:

$$I = \iint_A \vec{J} \cdot \hat{n} dS \tag{37.4}$$

Formula 37.1.6 (Free current). The current density generated by free charges is given by:

$$\vec{J} = nq\vec{v}_d \tag{37.5}$$

---

[1]It is several orders of magnitude smaller.

37.1.3 Pouillet's law

$$R = \rho \, \frac{l}{A} \tag{37.6}$$

where:

ρ : resistivity of the material

l : length of the resistor

A : cross-sectional area

37.2 Ohm's law

Formula 37.2.1 (Ohm's law).

$$\boxed{\vec{J} = \overset{\leftrightarrow}{\sigma} \cdot \vec{E}} \tag{37.7}$$

where $\overset{\leftrightarrow}{\sigma}$ is the conductivity tensor.

Formula 37.2.2 (Ohm's law in wires). The following formula can be found by combining equations 37.1, 37.2,37.4 and 37.7 and by assuming that the conductivity tensor can be simplified to a scalar (follows from the isotropic behaviour of normal resistors):

$$U = RI \tag{37.8}$$

37.3 Capacitance C

Definition 37.3.1 (Capacitance). The capacitance is a (geometrical) value that reflects the amount of charge a certain body can store.

$$C = \frac{q}{V} \tag{37.9}$$

37.4 Electric dipole

Formula 37.4.1 (Electric dipole).

$$\vec{p} = q\vec{a} \tag{37.10}$$

Where:

q : charge of the positive particle

$\vec{a}$: vector pointing from the negative to the positive particle

Formula 37.4.2 (Energy). If an electric dipole is placed in an electric field, its potential energy is:

$$U = -\vec{p} \cdot \vec{E} \tag{37.11}$$

Formula 37.4.3 (Torque). If an electric dipole is placed in an electric field, a torque is generated:

$$\vec{\tau} = \vec{p} \times \vec{E} \tag{37.12}$$

Chapter 38

Magnetism

38.1 Magnetic field

38.1.1 Magnetizing field $\vec{H}$

The magnetizing field $\vec{H}$ is the field resulting from all exterior sources.

38.1.2 Magnetization $\vec{M}$

$$\vec{M} = \chi\vec{H} \tag{38.1}$$

where χ is the magnetic susceptibility.

38.1.3 Magnetic induction $\vec{B}$

The magnetic induction $\vec{B}$ is the field resulting from exterior sources and interior magnetization. (It is the 'real', detectable field.) In vacuum we have the following relation between the magnetic induction B, the magnetizing field H and the magnetization M:

$$\vec{B} = \mu_0 \left(\vec{H} + \vec{M} \right) \tag{38.2}$$

By combining this formula with formula 38.1 we get[1]:

$$\vec{B} = \mu_0 \left(1 + \chi \right) \vec{H} \tag{38.3}$$

---

[1]This equation is only valid in linear media.

Definition 38.1.1 (Magnetic permeability). The proportionality constant in formula 38.3 is called the magnetic permeability:

$$\mu = \mu_0(1 + \chi) \tag{38.4}$$

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.

38.1.4   Tensorial formulation

In anistropic materials we have to use the tensorial formulation.

$$B_i = \sum_j \mu_{ij} H_j \tag{38.5}$$

$$M_i = \sum_j \chi_{ij} H_j \tag{38.6}$$

Both μ and χ are tensors of rank 2.

38.2   Magnetic multipoles

38.2.1   Dipole

$$\vec{\boldsymbol{m}} = IS\vec{\boldsymbol{u}}_n \tag{38.7}$$

38.3   Electric charges in a magnetic field

38.3.1   Cyclotron

Formula 38.3.1 (Gyroradius).
$$r = \frac{mv_\perp}{|q|B} \tag{38.8}$$

Formula 38.3.2 (Gyrofrequency[2]).

$$\omega = \frac{|q|B}{m} \tag{38.9}$$

---

[2]Also called the Larmor frequency.

Chapter 39

Maxwell equations

39.1 Lorentz force

Formula 39.1.1 (Lorentz force).

$$\vec{F} = q\left(\vec{E} + \vec{v} \times \vec{B}\right) \tag{39.1}$$

Formula 39.1.2 (Lorentz force density).

$$\vec{f} = \rho\vec{E} + \vec{J} \times \vec{B} \tag{39.2}$$

39.2 Differential Maxwell equations

Formula 39.2.1 (Gauss' law for electricity).

$$\nabla \cdot \vec{E} = \frac{\rho}{\varepsilon} \tag{39.3}$$

Formula 39.2.2 (Gauss' law for magnetism).

$$\nabla \cdot \vec{B} = 0 \tag{39.4}$$

Formula 39.2.3 (Faraday's law).

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \tag{39.5}$$

Formula 39.2.4 (Maxwell's law[1]).

$$\nabla \times \vec{H} = \frac{\partial \vec{D}}{\partial t} + \vec{J} \tag{39.6}$$

---

[1]Also called the law of Maxwell-Ampère.

39.3 Potentials

39.3.1 Decomposition in potentials

Remembering the Helmholtz decomposition (equation 15.16) we can derive the following general form for $\vec{B}$ starting from Gauss' law 39.4:

$$\boxed{\vec{B} = \nabla \times \vec{A}}$$ (39.7)

where $\vec{A}$ is the magnetic potential.

Combining equation 39.7 with Faraday's law 39.5 and rewriting it a bit, gives the following general form for $\vec{E}$:

$$\boxed{\vec{E} = -\nabla V - \frac{\partial \vec{A}}{\partial t}}$$ (39.8)

where V is the electrostatic potential.

39.3.2 Conditions

Substituting the expressions 39.7 and 39.8 into Gauss' law 39.3 and Maxwell's law 39.6 gives the following two (coupled) conditions for the electromagnetic potentials:

$$\triangle \vec{A} - \varepsilon\mu\frac{\partial^2 \vec{A}}{\partial t^2} = \nabla\left(\nabla\cdot\vec{A} + \varepsilon\mu\frac{\partial V}{\partial t}\right) - \mu\vec{J}$$ (39.9)

$$\triangle V - \varepsilon\mu\frac{\partial^2 V}{\partial t^2} = -\frac{\partial}{\partial t}\left(\nabla\cdot\vec{A} + \varepsilon\mu\frac{\partial V}{\partial t}\right) - \frac{\rho}{\varepsilon}$$ (39.10)

39.3.3 Gauge transformations

Looking at equation 39.7, it is clear that a transformation $\vec{A} \to \vec{A} + \nabla\psi$ has no effect on $\vec{B}$ due to property 15.13. To compensate this in equation 39.8, we also have to perform the transformation $V \to V - \frac{\partial\psi}{\partial t}$.

The (scalar) function $\psi(\vec{r}, t)$ is called a **gauge function**. The transformations are called **gauge transformations**.

Definition 39.3.1 (Gauge fixing conditions). Conditions that fix a certain gauge (or class of gauge transformations) are called gauge fixing conditions. These select one of many physically equivalent configurations.

39.3.4   Lorenz gauge

A first example of a gauge fixing condition is the Lorenz gauge[2] :

$$\nabla \cdot \vec{A} + \varepsilon\mu \frac{\partial V}{\partial t} = 0 \tag{39.11}$$

When using this gauge fixing condition, equations 39.9 and 39.10 become uncoupled and can be rewritten as:

$$\Box\vec{A} = -\mu\vec{J} \tag{39.12}$$

$$\Box V = -\frac{\rho}{\varepsilon} \tag{39.13}$$

To see which gauge functions ψ are valid in this case we perform a transformation as explained above:

$$\vec{A}' = \vec{A} + \nabla\psi \qquad \text{and} \qquad V' = V - \frac{\partial\psi}{\partial t}$$

Substituting these transformations in equation 39.11 and using the fact that both sets of potentials $(\vec{A}, V)$ and $(\vec{A}', V)$ satisfiy the Lorenz gauge 39.11 gives the following condition for the gauge function ψ:

$$\Box\psi = 0 \tag{39.14}$$

Example 39.3.2 (Alternative gauges). Apart from the Lorenz gauge 39.11, there is also the Coulomb gauge:

$$\nabla \cdot \vec{A} = 0 \tag{39.15}$$

39.4   Energy and momentum

Definition 39.4.1 (Poynting vector).

$$\vec{S} = \vec{E} \times \vec{H} \tag{39.16}$$

Definition 39.4.2 (Energy density).

$$W = \frac{1}{2}\left(\vec{E} \cdot \vec{D} + \vec{B} \cdot \vec{H}\right) \tag{39.17}$$

---

[2]Named after Ludvig Lorenz. Not to be confused with Hendrik Lorentz.

Part IX

Relativity Theory

Chapter 40

Special relativity

40.1 Lortenz transformations

Formula 40.1.1 (Lorentz factor).

$$\gamma = \frac{1}{\sqrt{1 - \dfrac{v^2}{c^2}}} \tag{40.1}$$

Formula 40.1.2 (Lorentz transformations). Let $\mathbf{V}$ be a 4-vector. A Lorentz boost along the x^1-axis is given by the following transformation:

$$\boxed{\begin{array}{rcl} V'^0 &=& \gamma\left(V^0 - \frac{v}{c}V^1\right) \\ V'^1 &=& \gamma\left(V^1 - \frac{v}{c}V^0\right) \\ V'^2 &=& V^2 \\ V'^3 &=& V^3 \end{array}} \tag{40.2}$$

Remark 40.1.3. Putting $c = +\infty$ in the previous transformation formulas gives the known Galilean transformations from classical mechanics.

40.2 Energy and momentum

Formula 40.2.1 (4-velocity).

$$U^\mu = \left(\frac{dx^0}{d\tau}, \frac{dx^1}{d\tau}, \frac{dx^2}{d\tau}, \frac{dx^3}{d\tau}\right) \tag{40.3}$$

or by applying the formulas for proper time and time dilatation we obtain:

$$U^\mu = (\gamma c, \gamma \vec{\boldsymbol{u}}) \tag{40.4}$$

Formula 40.2.2 (4-momentum).

$$p^\mu = m_0 U^\mu \tag{40.5}$$

or by setting $E = cp^0$:

$$p^\mu = \left(\frac{E}{c}, \gamma m_0 \vec{u} \right) \tag{40.6}$$

Definition 40.2.3 (Relativistic mass). The factor γm_0 in the momentum 4-factor is called the relativistic mass. By introducing this quantity (and denoting it by m), the classic formula $\vec{p} = m\vec{u}$ for the 3-momentum is preserved.

Formula 40.2.4 (Relativistic energy relation).

$$\boxed{E^2 = m^2 c^4 + p^2 c^2} \tag{40.7}$$

This is also sometimes called the Einstein relation for energy.

Chapter 41

General Relativity

41.1 Einstein field equations

Formula 41.1.1 (Einstein field equations). The Einstein field equations without the cosmological constant Λ read:

$$G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \tag{41.1}$$

41.2 Schwarzschild metric

Formula 41.2.1 (Schwarzschild metric).

$$ds^2 = \left(1 - \frac{R_s}{r}\right) c^2 dt^2 - \left(1 - \frac{R_s}{r}\right)^{-1} dr^2 - r^2 d\Omega^2 \tag{41.2}$$

where R_s is the Schwarzschild radius given by $R_s = \frac{2GM}{c^2}$.

Theorem 41.2.2 (Birkhoff's theorem). *The Schwarzschild metric is the unique solution of the vacuum field equation with the additional constraints of asymptotic flatness and staticity.*

Part X

Quantum Mechanics

Chapter 42

Schrödinger equation

42.1   One dimension

42.1.1   Time independent Schrödinger equation (TISE)

Formula 42.1.1 (TISE).

$$\boxed{\hat{\boldsymbol{H}}\psi(x) = E\psi(x)} \tag{42.1}$$

where $\hat{\boldsymbol{H}}$ is the Hamiltonian of the system.

Property 42.1.2 (Orthonormality). Let (ψ_i) be the set of eigenfunctions of the TISE. These functions obey the following orthogonality relations:

$$\int \psi_i^*(x)\psi_j(x)dx = \delta_{ij} \tag{42.2}$$

42.1.2   Time dependent Schrödinger equation (TDSE)

Formula 42.1.3 (General TDSE).

$$\boxed{i\hbar\frac{\partial\psi}{\partial t} = \hat{\boldsymbol{H}}\psi} \tag{42.3}$$

where $\hat{\boldsymbol{H}}$ is the Hamiltonian of the system.

Formula 42.1.4 (Massive particle in a time-independent potential).

$$\boxed{i\hbar\frac{\partial}{\partial t}\psi(x,t) = \left(\frac{\hat{\boldsymbol{p}}\cdot\hat{\boldsymbol{p}}}{2m} + V(x)\right)\psi(x,t)} \tag{42.4}$$

Formula 42.1.5 (General solution).

$$\psi(x,t) = \sum_E c_E \psi_E(x) e^{-\frac{i}{\hbar}Et}$$

(42.5)

where the functions $\psi_E(x)$ are the eigenfunctions of the TISE 42.1. The coefficients c_E can be found using the orthogonality relations as follows:

$$c_E = \left(\int \psi_E^*(x')\psi(x',t)dx' \right) e^{\frac{i}{\hbar}Et_0}$$

(42.6)

Chapter 43

Mathematical formalism

43.1 Postulates

43.1.1 Postulate 6: eigenfunction expansion

Definition 43.1.1 (Observable). An operator $\hat{A}$ which possesses a complete set of eigenfunctions is called an observable.

Formula 43.1.2. Let $|\Psi\rangle$ be an arbitrary wavefunction representing the system. Let the set $\{|\psi_n\rangle\}$ be a complete set of eigenfunctions of an observable of the system. The wavefunction $|\Psi\rangle$ can then be expanded as a linear combination of those eigenfunctions:

$$\boxed{|\Psi\rangle = \sum_n c_n|\psi_n\rangle + \int c_a|\psi_a\rangle da} \tag{43.1}$$

where the summation ranges over the discrete spectrum and the integral over the continuous spectrum.

Formula 43.1.3 (Closure relation). For a complete set of discrete eigenfunctions the closure relation[1] reads:

$$\sum_n |\psi_n\rangle\langle\psi_n| = \mathbb{1} \tag{43.2}$$

For a complete set of continuous eigenfunctions we have the following counterpart:

$$\int |\psi_i\rangle\langle\psi_i|di = \mathbb{1} \tag{43.3}$$

For a mixed set of eigenfunctions a similar relation is obtained by summing over the discrete eigenfunctions and integrating over the continuous eigenfunctions.

Remark. To simplify the notation we will almost always use the notation of equation 43.2 but implicitly integrate over possible continuous eigenfunctions.

---

[1] This relation is also called the 'resolution of the identity'.

43.2   Uncertainty relations

Formula 43.2.1 (Commutator). Let $\hat{A}, \hat{B}$ be two operators. We define the commutator of $\hat{A}$ and $\hat{B}$ as follows:

$$\boxed{\left[\hat{A}, \hat{B}\right] = \hat{A}\hat{B} - \hat{B}\hat{A}} \tag{43.4}$$

Property 43.2.2.

$$\left[\hat{A}\hat{B}, \hat{C}\right] = \hat{A}\left[\hat{B}, \hat{C}\right] + \left[\hat{A}, \hat{C}\right]\hat{B} \tag{43.5}$$

Formula 43.2.3 (Anticommutator). Let $\hat{A}, \hat{B}$ be two operators. We define the anticommutator of $\hat{A}$ and $\hat{B}$ as follows:

$$\boxed{\left\{\hat{A}, \hat{B}\right\}_{+} = \hat{A}\hat{B} + \hat{B}\hat{A}} \tag{43.6}$$

Definition 43.2.4 (Compatible observables). Let $\hat{A}, \hat{B}$ be two observables. If there exists a complete set of functions $|\psi_n\rangle$ that are simultaneously eigenfunctions of $\hat{A}$ and $\hat{B}$, the two operators are called **compatible**.

Formula 43.2.5 (Uncertainty relation). Let $\hat{A}, \hat{B}$ be two observables. Let $\Delta A, \Delta B$ be the corresponding uncertainties. We have the following relation:

$$\boxed{\Delta A \Delta B = \frac{1}{4}\left|\left\langle\left[\hat{A}, \hat{B}\right]\right\rangle\right|^2} \tag{43.7}$$

43.3   Matrix representation

Formula 43.3.1. The following formula gives the (m, n)-th element of the matrix representation of $\hat{A}$ with respect to the basis orthonormal $\{\psi_n\}$.

$$\boxed{A_{mn} = \langle\psi_m|\hat{A}|\psi_n\rangle} \tag{43.8}$$

Remark 43.3.2. The basis $\{\psi_n\}$ need not consist out of eigenfunctions of $\hat{A}$.

43.4   Slater determinants

Theorem 43.4.1 (Symmetrization postulate). *A system of n identical particles is described by a wave function Ψ belonging to either $S^n(\mathcal{H})$ or $\Lambda^n(\mathcal{H})$, where $\mathcal{H}$ is the Hilbert space belonging to a single particle.*

Remark 43.4.2. In ordinary quantum mechanics this is a postulate, but in quantum field theory this is a consequence of the spin-statistics theorem of Pauli.

Formula 43.4.3. Let $\{\sigma\}$ be the set of all permutations of the sequence $(1, ..., n)$. Let $|\psi\rangle$ be the single-particle wave function. Fermionic systems are described by a wave function of the form

$$|\Psi_F\rangle = \sum_\sigma |\psi_{\sigma(1)}\rangle \cdots |\psi_{\sigma(n)}\rangle \tag{43.9}$$

Bosonic systems are described by a wave function

$$|\Psi_B\rangle = \sum_\sigma \text{sgn}(\sigma) |\psi_{\sigma(1)}\rangle \cdots |\psi_{\sigma(n)}\rangle \tag{43.10}$$

Definition 43.4.4 (Slater determinant). Let $\{\phi_i(\vec{q}_i)\}_{i \leq N}$ be the set of wave functions (spin orbitals) describing a system of N identical particles. The totally antisymmetric wave functions for the complete system is given by

$$\boxed{\psi(\vec{q}_1, ..., \vec{q}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\vec{q}_1) & \cdots & \phi_N(\vec{q}_1) \\ \vdots & & \vdots \\ \phi_1(\vec{q}_N) & \cdots & \phi_N(\vec{q}_N) \end{vmatrix}} \tag{43.11}$$

Chapter 44

Angular Momentum

44.1   General operator

Consider the general angular momentum operator $\hat{J} = \left(\hat{J}_x, \hat{J}_y, \hat{J}_z \right)$. This operator works on the Hilbert space spanned by the eigenbasis $\{|j, m\rangle\}$.

Property 44.1.1. The mutual eigenbasis of $\hat{J}^2$ and $\hat{J}_z$ is defined by the following two eigenvalue equations:

$$\hat{J}^2|j, m\rangle = j(j+1)\hbar^2|j, m\rangle \tag{44.1}$$

$$\hat{J}_z|j, m\rangle = m\hbar|j, m\rangle \tag{44.2}$$

Formula 44.1.2 (Commutation relation). The angular momentum operators satisfy following commutation relation:

$$\boxed{\left[\hat{J}_i, \hat{J}_j\right] = i\hbar\varepsilon_{ijk}\hat{J}_k} \tag{44.3}$$

Definition 44.1.3 (Ladder operators[1]). The raising and lowering operators[2] $\hat{J}_+$ and $\hat{J}_-$ are defined as:

$$\hat{J}_+ = \hat{J}_x + i\hat{J}_y \qquad \text{and} \qquad \hat{J}_- = \hat{J}_x - i\hat{J}_y \tag{44.4}$$

Corollary 44.1.4. From the definition of the ladder operators and the commutation relations of the angular momentum operators we can derive following commutation relations:

$$\left[\hat{J}_+, \hat{J}_-\right] = 2\hbar\hat{J}_z \tag{44.5}$$

---

[1]Also called the **creation** and **annihilation** operators.
[2]These operators will only affect the z-projection, not the total angular momentum.

Formula 44.1.5. The total angular momentum operator $\hat{J}^2$ can now be expressed in terms of these ladder operators and $\hat{J}_z$ using the commutation relation 44.3:

$$\hat{J}^2 = \hat{J}_+ \hat{J}_- + \hat{J}_z^2 - \hbar \hat{J}_z \tag{44.6}$$

Remark 44.1.6 (Casimir operator). From the definition of $\hat{J}^2$ it follows that this operator is a Casimir invariant[3] in the algebra generated by the operators $\hat{J}_i$.

44.2 Rotations

44.2.1 Infinitesimal rotation

Formula 44.2.1. An infinitesimal rotation $\hat{R}(\delta\vec{\varphi})$ is given by the following formula:

$$\boxed{\hat{R}(\delta\vec{\varphi}) = \mathbb{1} - \frac{i}{\hbar}\vec{J} \cdot \delta\vec{\varphi}} \tag{44.7}$$

A finite rotation can then be produced by applying this infinitesimal rotation repeatedly, which gives:

$$\hat{R}(\vec{\varphi}) = \left(\mathbb{1} - \frac{i}{\hbar}\vec{J} \cdot \frac{\vec{\varphi}}{n}\right)^n = \exp\left(-\frac{i}{\hbar}\vec{J} \cdot \vec{\varphi}\right) \tag{44.8}$$

Formula 44.2.2 (Matrix elements $\hat{R}_{ij}$). Applying a rotation around the z-axis to a state $|j, m\rangle$ gives:

$$\hat{R}(\varphi\vec{e}_z)|j, m\rangle = \exp\left(-\frac{i}{\hbar}\hat{J}_z\varphi\right)|j, m\rangle = \exp\left(-\frac{i}{\hbar}m\varphi\right)|j, m\rangle \tag{44.9}$$

Multiplying these states with a bra $\langle j', m'|$ and using the orthonormality of the eigenstates gives the matrix elements of the rotation operator:

$$\boxed{\hat{R}_{ij}(\varphi\vec{e}_z) = \exp\left(-\frac{i}{\hbar}m\varphi\right)\delta_{jj'}\delta_{mm'}} \tag{44.10}$$

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 a block diagonal matrix with respect to j.

We thus get the following expression for the rotation matrix:

$$\boxed{\langle j, m'|\hat{R}(\varphi\vec{n})|j, m\rangle = \mathcal{D}_{m,m'}^{(j)}(\hat{R})} \tag{44.11}$$

where the values $\mathcal{D}_{m,m'}^{(j)}(\hat{R})$ are the **Wigner D**-functions.

---

[3]See definition 22.2.25.

Remark (Wigner D-matrices). For every value of j there are $(2j+1)$ values for m. The matrix $\mathcal{D}^{(j)}(\hat{R})$ is thus a $(2j+1) \times (2j+1)$-matrix

Property 44.2.3. The rotations form a group and the Wigner D-matrices form a representation of this group. The latter implies that[4]:

$$\mathcal{D}^{(j)}(\hat{R}_1 \cdot \hat{R}_2) = \mathcal{D}^{(j)}(\hat{R}_1)\mathcal{D}^{(j)}(\hat{R}_2) \tag{44.12}$$

44.2.2 Spinor-representation

Definition 44.2.4 (Pauli matrices).

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{44.13}$$

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.

Formula 44.2.5. For a rotation around the z-axis the Wigner-D matrix reads as follows:

$$\mathcal{D}^{(1/2)}(\varphi \vec{e}_z) = \begin{pmatrix} e^{-i/2\varphi} & 0 \\ 0 & e^{i/2\varphi} \end{pmatrix} \tag{44.14}$$

44.3 Coupling of angular momenta

44.3.1 Total Hilbert space

Let $\mathcal{H}_i$ denote the Hilbert space of states belonging to the ith particle. The Hilbert space of the total system is given by the following tensor product:

$$\mathcal{H} = \mathcal{H}_1 \otimes ... \otimes \mathcal{H}_n$$

Due to the tensor product definition above, the angular momentum operator $\hat{J}_i$ should now be interpreted as $\mathbb{1} \otimes ... \otimes \hat{J}_i \otimes ... \otimes \mathbb{1}$. This implies that the angular momentum operators $\hat{J}_{l \neq i}$ do not act on the space $\mathcal{H}_i$, so one can pull these operators through the tensor product:

$$\hat{J}_i|j_1\rangle \otimes ... \otimes |j_n\rangle = |j_1\rangle \otimes ... \otimes \hat{J}_i|j_i\rangle \otimes ... \otimes |j_n\rangle$$

The basis as depicted above is called the **uncoupled basis**.

---

[4]As representations they are also group homomorphisms.

44.3.2 Clebsch-Gordan series

Let $\vec{J}$ denote the total angular momentum defined as:

$$\vec{J} = \hat{J}_1 + \hat{J}_2 \tag{44.15}$$

With this operator we can define a new **coupled** state $|\mathbf{J}, \mathbf{M}\rangle$ where $\mathbf{M}$ is the total magnetic quantum number which ranges from $-\mathbf{J}$ to $\mathbf{J}$.

Formula 44.3.1 (Clebsch-Gordan coefficients). As both bases (coupled and uncoupled) span the total Hilbert space $\mathcal{H}$ there has to exist a transformation between them. The transformation coefficients can be found by using the resolution of the identity:

$$\boxed{|\mathbf{J}, \mathbf{M}\rangle = \sum_{m_1=-j_1}^{j_1} \sum_{m_2=-j_2}^{j_2} |j_1, j_2, m_1, m_2\rangle\langle j_1, j_2, m_1, m_2|\mathbf{J}, \mathbf{M}\rangle} \tag{44.16}$$

These coefficients are called the Clebsch-Gordan coefficients.

Property 44.3.2. By acting with the operator $\vec{J}_z$ on both sides of equation 44.16 it possible to proof that the CG coefficient are non-zero if and only if $\mathbf{M} = m_1 + m_2$.

Chapter 45

Dirac equation

45.1   Dirac equation

Definition 45.1.1 (Dirac matrices). The time-like Dirac matrix γ^0 is defined as:

$$\gamma^0 = \begin{pmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{pmatrix} \tag{45.1}$$

where $\mathbb{1}$ is the 2-dimensional identity matrix. The space-like Dirac matrices γ^k, $k = 1, 2, 3$ are defined using the Pauli matrices σ^k:

$$\gamma^k = \begin{pmatrix} 0 & \sigma^k \\ -\sigma^k & 0 \end{pmatrix} \tag{45.2}$$

This form of the Dirac matrices fixes a basis called the **Dirac basis**. The **Weyl** or **chiral** basis is fixed by replacing the time-like matrix γ^0 by

$$\gamma^0 = \begin{pmatrix} 0 & \mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix} \tag{45.3}$$

Property 45.1.2. The Dirac matrices satisfy

$$\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = 2\eta^{\mu\nu} \mathbb{1} \tag{45.4}$$

This has the form of equation 19.4. The Dirac matrices can thus be used as the generating set of a Clifford algebra (see defiinition 19.1.1).

Notation 45.1.3 (Feynman slash notation). Let $a = a_\mu x^\mu \in V$ be a general 4-vector. The Feynman slash $\slashed{a}$ is defined as follows:

$$\slashed{a} = \gamma^\mu a_\mu \tag{45.5}$$

A more formal treatment of the Feynman slash notation shows us that it gives us a canonical map:

$$/ : V \to C\ell_V : a_\mu x^\mu \mapsto a_\mu \gamma^\mu \tag{45.6}$$

Formula 45.1.4 (Dirac equation). In covariant form the Dirac equation reads:

$$(i\hbar\partial\!\!\!/ - mc)\psi = 0 \tag{45.7}$$

Chapter 46

Perturbation theory

46.1 Rayleigh-Schrödinger perturbation theory

The basic of assumptions of the Rayleigh-Schrödinger perturbation theory are that the perturbation Hamiltonian is time-independent and that the eigenfunctions of the unperturbed Hamiltonian $\hat{H}_0$ also form a complete set for the perturbed Hamiltonian.

Notation 46.1.1. The perturbed eigenfunctions and energy-eigenvalues are denoted by $|\psi_n\rangle$ and E_n.

Formula 46.1.2. The perturbed eigenfunctions and eigenvalues can be expanded in the following way, where we assume that λ is a small perturbation parameter:

$$|\psi_n\rangle = \sum_{i=0}^{+\infty} \lambda^i |\psi_n^{(i)}\rangle \tag{46.1}$$

$$E_n = \sum_{i=0}^{+\infty} \lambda^i E_n^{(i)} \tag{46.2}$$

where 'i' denotes the order of the perturbation.

46.2 Time-dependent perturbation theory

In this section we consider perturbed Hamiltoninians of the following form:

$$\hat{H}(t) = \hat{H}_0 + \lambda \hat{V}(t) \tag{46.3}$$

46.2.1   Dyson series

Formula 46.2.1 (Tomonaga-Schwinger equation). The evolution operator $\hat{U}(t)$ satisfies the Schrödinger equation in the interaction image as follows:

$$i\hbar\frac{d}{dt}\hat{U}_I|\psi(0)\rangle_I = \hat{V}_I(t)\hat{U}_I|\psi(0)\rangle_I \tag{46.4}$$

Formula 46.2.2 (Dyson series). Together with the initial value condition $\hat{U}_I(0) = \mathbb{1}$ the Tomonaga-Schwinger equation becomes an initial value problem. A possible solution is given by:

$$\hat{U}_I(t) = \mathbb{1} - \frac{i}{\hbar}\int_0^t \hat{V}_I(t')\hat{U}_I(t')dt' \tag{46.5}$$

This solution can be iterated to become a series expansion of the evolution operator:

$$\hat{U}(t) = 1 - \frac{i}{\hbar}\int_0^t \hat{V}(t_1)dt_1 + \left(-\frac{i}{\hbar}\right)\int_0^t dt_1 \int_0^{t_1} dt_2 \hat{V}(t_1)\hat{V}(t_2) + ... \tag{46.6}$$

It is clear that the integrands obey a time-ordering. Introducing the **time-ordering operator** $\mathcal{T}$

$$\mathcal{T}\left(\hat{V}(t_1)\hat{V}(t_2)\right) = \begin{cases} \hat{V}(t_1)\hat{V}(t_2) & , & t_1 \geq t_2 \\ \hat{V}(t_2)\hat{V}(t_1) & , & t_2 > t_1 \end{cases} \tag{46.7}$$

the integrals can be rewritten in a more symmetric form:

$$\hat{U}(t) = 1 - \frac{i}{\hbar}\int_0^t \hat{V}(t_1)dt_1 + \frac{1}{2!}\left(-\frac{i}{\hbar}\right)\int_0^t dt_1 \int_0^t dt_2 \mathcal{T}\left(\hat{V}(t_1)\hat{V}(t_2)\right) + ... \tag{46.8}$$

or by comparing with the series expansion for exponential functions:

$$\boxed{\hat{U}(t) = \mathcal{T}\left(e^{-\frac{i}{\hbar}\int_0^t \hat{V}(t')dt'}\right)} \tag{46.9}$$

This last expansion is called the **Dyson series**.

46.3   Variational method

Formula 46.3.1 (Energy functional).

$$E(\psi) = \frac{\langle\psi|\hat{H}|\psi\rangle}{\langle\psi|\psi\rangle} \tag{46.10}$$

Property 46.3.2. The energy functional 46.10 satisfies following inequality:

$$E(\psi) \geq E_0 \tag{46.11}$$

where E_0 is the ground state energy.

Method 46.3.3. Assume that the trial function $|\psi\rangle$ depends on a set of parameters $\{c_i\}_{i\in I}$. The 'optimal' wave function is found by solving the following system of equations:

$$\frac{\partial\psi}{\partial c_i} = 0 \quad , \quad \forall i \in I \tag{46.12}$$

46.4 Adiabatic approximation

46.4.1 Berry phase

Consider a system for which the adiabatic approximation is valid. We then have a wavefunction of the form

$$\psi(t) = C_a(t)\psi_a(t)\exp\left[-\frac{i}{\hbar}\int_{t_0}^{t} E_a(t')dt'\right] \tag{46.13}$$

It follows from the orthonormality of the eigenstates $\psi_k(t)$ that the coefficient $C_a(t)$ is just a phase factor, so we can write it as

$$C_a(t) = e^{i\gamma_a(t)} \tag{46.14}$$

Substituting this ansatz in the wavefunction and the Schödinger equation gives a differential equation for the phase factor $\gamma_a(t)$. It can be readily integrated to obtain

$$\gamma_a(t) = i\int_{t_0}^{t}\left\langle\psi_a(t')\Big|\frac{\partial\psi_a(t')}{\partial t'}\right\rangle dt' \tag{46.15}$$

If the system is cyclic, then $\psi_a(t_0) = \psi_a(t_f)$. Due to time evolution the wavefunction accumulates a phase through the coefficient $C_a(t)$ over the period $t_0 - t_f$. This phase is called the **Berry phase**.

Lets try to apply a phase transformation to remove the Berry phase:

$$\psi_a'(t) = \psi_a(t)e^{i\eta(t)} \tag{46.16}$$

Entering this in equation 46.15 gives

$$\bar{\gamma}_a'(t) = \bar{\gamma}_a(t) - \eta(t_f) + \eta(t_0) \tag{46.17}$$

where the overhead bar denotes the integration between t_0 and t_f in equation 46.15. Combining 46.16 with the cyclic property of the wavefunction gives us

$$\eta(t_f) - \eta(t_0) = 2k\pi \qquad k \in \mathbb{N} \tag{46.18}$$

which means that the Berry phase cannot be eliminated and is thus an observable property of the system.

Definition 46.4.1 (Berry connection). The quantity

$$\mathbf{A}(\vec{c}) = i\langle\psi_a(\vec{c})|\nabla_{\vec{c}}\psi_a(\vec{c})\rangle \tag{46.19}$$

where $\nabla_{\vec{c}}$ denotes the gradient in phase space, is called the Berry connection (or Berry gauge potential). Applying Stokes' theorem to 46.15 gives us

$$\bar{\gamma}_a = \int \boldsymbol{\mathcal{B}} \cdot d\vec{S} \tag{46.20}$$

where $\boldsymbol{\mathcal{B}} = \nabla_{\vec{c}} \times \mathbf{A}(\vec{c})$ is called the **Berry curvature**. Although the Berry connection is gauge dependent, the Berry curvature is gauge invariant!

Chapter 47

Scattering theory

47.1 Cross section

Formula 47.1.1 (Differential cross section).

$$\frac{d\sigma}{d\Omega} = \frac{N(\theta, \varphi)}{F} \tag{47.1}$$

where is the incoming particle flux and N the detected flow rate[1].

47.1.1 Fermi's golden rule

Formula 47.1.2 (Fermi's golden rule). The transition propability from state i to state f is given by:

$$\boxed{\Gamma_{i \to f} = \frac{2\pi}{\hbar} |\langle f | \hat{H} | i \rangle|^2 \frac{dn}{dE_f}} \tag{47.2}$$

47.2 Lippman-Schwinger equations

In this section we consider Hamiltonians of the following form: $\hat{H} = \hat{H}_0 + \hat{V}$ where $\hat{H}_0$ is the free Hamiltonian and $\hat{V}$ the scattering potential. We will also assume that both the total Hamiltonian and the free Hamiltonian have the same eigenvalues.

Formula 47.2.1 (Lippman-Schwinger equation).

$$|\psi^{(\pm)}\rangle = |\varphi\rangle + \frac{1}{E - \hat{H}_0 \pm i\varepsilon} \hat{V} |\psi^{(\pm)}\rangle \tag{47.3}$$

where $|\varphi\rangle$ is an eigenstate of the free Hamiltonian with the same energy as $|\psi\rangle$.

---

[1] As N is not defined as a rate per unit area (flux), the differential cross section has the dimension of area.

Remark 47.2.2. The term $\pm i\varepsilon$ is added to the denominator as otherwise it would be singular. It has no real physical meaning.

Formula 47.2.3 (Born series). If we rewrite the Lippman-Schwinger equation as $|\psi\rangle = |\varphi\rangle + \hat{G}_0 \hat{V} |\psi\rangle$, were $\hat{G}_0$ is the Green's operator, then we can derive the following series expansion by iterating the equation:

$$|\psi\rangle = |\varphi\rangle + \hat{G}_0 \hat{V} |\varphi\rangle + \left(\hat{G}_0 \hat{V} \right)^2 |\varphi\rangle + \dots \tag{47.4}$$

Formula 47.2.4 (Born approximation). If we cut off the Born series at the first order term in $\hat{V}$ then we obtain the Born approximation:

$$|\psi\rangle = |\varphi\rangle + \hat{G}_0 \hat{V} |\varphi\rangle \tag{47.5}$$

Chapter 48

Entanglement & Quantum computing

48.1   Bipartite systems

48.1.1   Marginal density operators

Definition 48.1.1 (Marginal density operator). Let $|\Psi\rangle_{AB}$ be the state of a bipartite system. The marginal density operator $\hat{\boldsymbol{\rho}}_A$ of system A is defined as follows:

$$\hat{\boldsymbol{\rho}}_A = \mathrm{Tr}_B |\Psi\rangle_{AB\,AB}\langle\Psi| \tag{48.1}$$

Definition 48.1.2 (Purification). Let $\hat{\boldsymbol{\rho}}_A$ be the desnity operator of a system A. A purification of $\hat{\boldsymbol{\rho}}_A$ is a pure state $|\Psi\rangle_{AB}$ of a composite system AB such that:

$$\hat{\boldsymbol{\rho}}_A = \mathrm{Tr}_B |\Psi\rangle_{AB\,AB}\langle\Psi| \tag{48.2}$$

Property 48.1.3. Any two purifications of the same density operator $\hat{\rho}_A$ are related by a transformation $\mathbb{1}_A \otimes \hat{\boldsymbol{V}}$, where $\hat{\boldsymbol{V}}$ is a unitary operator on $\mathcal{H}_B$.

Chapter 49

Quantum Field Theory

In this chapter we adopt the standard Minkowskian signature $(+, -, -, -)$ unless otherwise stated. This follows the introductory literature and courses such as [3]. Furthermore we also work in natural units unless stated otherwise, i.e. $\hbar = c = 1$.

49.1 Noether's theorem

Theorem 49.1.1 (Noether's theorem[†]). *Consider a field transformation*

$$\phi(x) \to \phi(x) + \alpha\delta\phi(x) \tag{49.1}$$

where α is an infinitesimal quantity and $\delta\phi$ is a small deformation. In case of a symmetry we obtain the following conservation law:

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \delta\phi - \mathcal{J}^\mu \right) = 0 \tag{49.2}$$

The factor between parentheses can be interpreted as a conserved current $j^\mu(x)$. Noether's theorem states that every symmetry of the form 49.1 leads to such a current.

The conservation can also be expressed in terms of a charge[1]:

$$\frac{dQ}{dt} = \frac{d}{dt} \int j^0 d^3x = 0 \tag{49.3}$$

Definition 49.1.2 (Stress-energy tensor). Consider a field transformation

$$\phi(x) \to \phi(x + a) = \phi(x) + a^\mu \partial_\mu \phi(x)$$

Because the Lagrangian is a scalar it transforms similarly:

$$\mathcal{L} \to \mathcal{L} + a^\mu \partial_\mu \mathcal{L} = \mathcal{L} + a^\nu \partial_\mu (\delta^\mu{}_\nu \mathcal{L}) \tag{49.4}$$

---

[1]The conserved current and its associated charge are called the **Noether current** and **Noether charge**.

This leads to the existence of 4 conserved currents. These can be used to define the stress-energy tensor:

$$T^\mu_{\ \nu} = \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)}\partial_\nu \phi - \mathcal{L}\delta^\mu_{\ \nu} \tag{49.5}$$

49.2 Klein-Gordon Field

49.2.1 Lagrangian and Hamiltonian

The simplest Lagrangian (density) is given by:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)^2 - \frac{1}{2}m^2\phi^2 \tag{49.6}$$

Using the principle of least action we obtain the following Euler-Lagrange equations[2]:

$$\left(\partial^\mu \partial_\mu + m^2\right)\phi = 0 \tag{49.7}$$

or by introducing the **d'Alembertian** $\Box = \partial^\mu \partial_\mu$:

$$(\Box + m^2)\phi = 0 \tag{49.8}$$

This equation is called the Klein-Gordon equation. In the limit $m \to 0$ it reduces to the well-known wave equation.

From the Lagrangian 49.6 we can also derive a Hamiltonian function using relation 32.7:

$$H = \int d^3x \frac{1}{2}\left[\pi^2(x) + (\nabla\phi(x))^2 + m^2\phi^2(x)\right] \tag{49.9}$$

49.2.2 Raising and lowering operators

Fourier expanding the scalar field $\phi(\vec{x}, t)$ in momentum space and inserting it into the Klein-Gordon equation gives:

$$\left(\partial_t^2 + p^2 + m^2\right)\phi(\vec{p}, t) = 0 \tag{49.10}$$

This is the equation for a simple harmonic oscillator with frequency $\omega_{\vec{p}} = \sqrt{p^2 + m^2}$.

---

[2]See formula 32.4.

Analogous to ordinary quantum mechanics we define the raising and lowering operators $a_{\vec{p}}^{\dagger}$ and $a_{\vec{p}}$ such that:

$$\phi(\vec{x}) = \iiint \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2\omega_{\vec{p}}}} \left(a_{\vec{p}} e^{i\vec{p}\cdot\vec{x}} + a_{\vec{p}}^{\dagger} e^{-i\vec{p}\cdot\vec{x}} \right) \tag{49.11}$$

$$\pi(\vec{x}) = \iiint \frac{d^3p}{(2\pi)^3} (-i) \sqrt{\frac{\omega_{\vec{p}}}{2}} \left(a_{\vec{p}} e^{i\vec{p}\cdot\vec{x}} - a_{\vec{p}}^{\dagger} e^{-i\vec{p}\cdot\vec{x}} \right) \tag{49.12}$$

An equivalent definition is obtained by performing the transformation $\vec{p} \to -\vec{p}$ in the second term of $\phi(\vec{x})$ and $\pi(\vec{x})$:

$$\phi(\vec{x}) = \iiint \frac{d^3p}{(2\pi)^3} \frac{1}{\sqrt{2\omega_{\vec{p}}}} \left(a_{\vec{p}} + a_{-\vec{p}}^{\dagger} \right) e^{i\vec{p}\cdot\vec{x}} \tag{49.13}$$

$$\pi(\vec{x}) = \iiint \frac{d^3p}{(2\pi)^3} (-i) \sqrt{\frac{\omega_{\vec{p}}}{2}} \left(a_{\vec{p}} - a_{-\vec{p}}^{\dagger} \right) e^{i\vec{p}\cdot\vec{x}} \tag{49.14}$$

When we impose the commutation relation

$$[a_{\vec{p}}, a_{\vec{q}}^{\dagger}] = (2\pi)^3 \delta(\vec{p} - \vec{q}) \tag{49.15}$$

we obtain the following commutation relation for the scalar field and its conjugate momentum:

$$[\phi(\vec{x}), \pi(\vec{y})] = i\delta(\vec{x} - \vec{y}) \tag{49.16}$$

Combining the previous formulas gives us the following important commutation relations:

$$[H, a_{\vec{p}}^{\dagger}] = \omega_p a_{\vec{p}}^{\dagger} \tag{49.17}$$

$$[H, a_{\vec{p}}] = -\omega_p a_{\vec{p}} \tag{49.18}$$

The Hamiltonian can also be explicitly calculated:

$$H = \int \frac{d^3p}{(2\pi)^3} \omega_{\vec{p}} \left(a_{\vec{p}}^{\dagger} a_{\vec{p}} + \frac{1}{2} [a_{\vec{p}}, a_{\vec{p}}^{\dagger}] \right) \tag{49.19}$$

It is however immediately clear from 49.15 that the second term in this integral diverges. This is a consequence of both the fact that space is infinite, i.e. the d^3x integral diverges, and the "large p" limit in the the d^3p integral. The first divergence can be resolved by applying some kind of boundary and considering the energy density instead of the energy itself. The second divergence follows from the fact that by including very large values for p in the integral we enter a parameter range where our theory is likely to break down. So we should introduce a "high p" cut-off. A more practical solution is to note that only energy differences are physical and so we can drop the second term altogether as it is merely a "constant".

49.2.3 Complex scalar fields

Formula 49.2.1 (Pauli-Jordan function).

$$[\phi(x), \phi(y)] = i \underbrace{\int \frac{d^3p}{(2\pi)^3} \frac{1}{2\omega_p} \left(e^{-ip\cdot(x-y)} - e^{ip\cdot(x-y)}\right)}_{\Delta(x-y)} \tag{49.20}$$

In the case that $x^0 = y^0$ (ETCR) or $(x - y)^2 < 0$ the Pauli-Jordan function is identically 0.[3]

49.3 Lorentz invariant integrals

When applying a Lorentz boost Λ the delta function $\delta(\vec{p} - \vec{q})$ transforms[4] as $\delta(\Lambda\vec{p} - \Lambda\vec{q})\frac{\Lambda E}{E}$. This is clearly not a Lorentz invariant quantity and cannot be used for normalisation. It is however also clear that the quantity $2E\delta(\vec{p} - \vec{q})$ is Lorentz invariant. (The constant 2 is merely introduced for future convenience.) The correct normalisation for the momentum representation thus becomes:

$$\langle p|q \rangle = 2E_p(2\pi)^3\delta(\vec{p} - \vec{q}) \tag{49.21}$$

where the factors 2 and $(2\pi)^3$ are again a matter of convention.

The factor $2E_p$ does not only occur in the normalisation conditions. To define a Lorentz invariant measure for evaluating integrals in spacetime we define define the following integral:

$$\int \frac{d^3p}{2E_p} = \int d^4p \ \delta(p^2 - m^2)\Big|_{p^0>0} \tag{49.22}$$

This means that the integral of any Lorentz invariant function $f(p)$ using the measure $\frac{d^3p}{2E_p}$ will be Lorentz invariant.

Computing the quantity $\langle 0|\phi(\vec{x})|\vec{p}\rangle$ gives $e^{i\vec{x}\cdot\vec{p}}$. This coincides with the position representation from quantum mechanics of the state $|\vec{p}\rangle$ and so we will also interpret it in QFT as the position representation of the single particle state $|\vec{p}\rangle$.

49.4 Wick's theorem

49.4.1 Bosonic fields

Definition 49.4.1 (Contraction for neutral bosonic fields).

$$\overbrace{\phi(x)\phi(y)} = \begin{cases} [\phi(x)^{(+)}, \phi(y)^{(-)}] & x^0 > y^0 \\ [\phi(y)^{(+)}, \phi(x)^{(-)}] & y^0 > x^0 \end{cases} \tag{49.23}$$

---

[3]See also the axiom of microcausality 49.5.1
[4]This follows from property 10.9.

Formula 49.4.2 (Feynman propagator).

$$\overbrace{\phi(x)\phi(y)}^{} = i \lim_{\varepsilon \to 0^+} \underbrace{i \int \frac{d^4 k}{(2\pi)^4} \frac{e^{-ik \cdot (x-y)}}{k^2 - m^2 + i\varepsilon}}_{\Delta_F(x-y)} \tag{49.24}$$

Definition 49.4.3 (Contraction for charged bosonic fields).

$$\overbrace{\phi(x)\overline{\phi}(y)}^{} = \begin{cases} [\phi(x)^{(+)}, \overline{\phi}(y)^{(-)}] & x^0 > y^0 \\ [\phi(y)^{(+)}, \overline{\phi}(x)^{(-)}] & y^0 > x^0 \end{cases} \tag{49.25}$$

Remark 49.4.4. In the case of charged bosons, only contractions of the form $\overbrace{\phi(x)\overline{\phi}(y)}^{}$ will remain because $[a, b^+] = 0$ for charged bosons.

49.4.2   Fermionic fields

Definition 49.4.5 (Contraction).

$$\overbrace{\psi(x)\overline{\psi}(y)}^{} = \begin{cases} \{\psi(x)^{(+)}, \overline{\psi}(y)^{(-)}\}_+ & x^0 > y^0 \\ -\{\psi(y)^{(+)}, \overline{\psi}(x)^{(-)}\}_+ & y^0 > x^0 \end{cases} \tag{49.26}$$

Remark 49.4.6. Only contractions of the form $\overbrace{\psi(x)\overline{\psi}(y)}^{}$ will remain because $\{a, b^+\}_+ = 0$.

Formula 49.4.7 (Feynman propagator).

$$\overbrace{\psi(x)\overline{\psi}(y)}^{} = i \lim_{\varepsilon \to 0^+} \underbrace{\int \frac{d^4 p}{(2\pi)^4} \frac{\not{p} + m}{p^2 - m^2 + i\varepsilon} e^{-ip \cdot (x-y)}}_{S_F(x-y)} \tag{49.27}$$

49.5   Axiomatic approach

Theorem 49.5.1 (Axiom of microcausality). *Let $\hat{O}$ be an observable and let x, y be two spacetime points. If $x - y$ is a space-like vector then $[\hat{O}(x), \hat{O}(y)] = 0$.*

49.5.1   Wightman axioms

Part XI

Thermal Physics & Statistical Mechanics

Chapter 50

Thermodynamics

50.1   General definitions

Definition 50.1.1 (System). The part of space that we are examining.

Definition 50.1.2 (Surroundings). Everything outside the system.

Definition 50.1.3 (Immediate surrounding). The part of the surroundings that 'lies' immediately next to the system.

Definition 50.1.4 (Environment). Everything outside the immediate surroundings.

Definition 50.1.5 (Thermodynamic coordinates). Macroscopical quantities that describe the system.

Definition 50.1.6 (Intensive coordinate). Coordinate that does not depend on the total amount of material (or system size).

Definition 50.1.7 (Extensive coordinate). Coordinate that does depend on the amount of material.

Definition 50.1.8 (Thermodynamic equilibrium). A system in thermodynamic equilibrium is simultaneously in thermal, mechanical and chemical equilibrium. The system is also described by a certain set of constant coordinates.

Theorem 50.1.9. *During thermodynamic equilibrium, all intensive coordinates are uniform throughout the system.*

Definition 50.1.10 (Isolated system). An isolated system can't interact with its surroundings (due to the presence of impenetrable walls).

Definition 50.1.11 (Diathermic wall). A diathermic wall is a wall that allows heat transfer.

Definition 50.1.12 (Adiabatic wall). An adiabatic wall is a wall that does not allow heat transfer.

Definition 50.1.13 (Open system). An open system is a system that allows matter exchange.

Definition 50.1.14 (Closed system). A closed system is a system that does not allow matter exchange.

Definition 50.1.15 (Quasistatic process). A quasistatic process is a sequence of equilibrium states separated by infinitesimal changes.

Definition 50.1.16 (Path). The sequence of equilibrium states in a process is called the path.

50.2 Postulates

Theorem 50.2.1 (Zeroth law). *If two object are in thermal equilibrium with a third object then they are also in thermal equilibrium with eachother.*

Theorem 50.2.2 (First law).

$$U_f - U_i = W + Q \tag{50.1}$$

$$dU = \delta W + \delta Q \tag{50.2}$$

Remark. The δ in the heat and work differentials implies that these are 'inexact' differentials. This means that they cannot be expressed as functions of the thermodynamic coordinates. More formally a differential form dx is called inexact if the integral $\int dx$ is path dependent.

Theorem 50.2.3 (Kelvin-Planck formulation of the second law). *No machine can absorb an amount of heat and completely transform it into work.*

Theorem 50.2.4 (Clausius formulation of the second law). *Heat cannot be passed from a cooler object to a warmer object without performing work.*

Formula 50.2.5 (Clausius' inequality). In differential form, the inequality reads:

$$\frac{\delta Q}{T} \geq 0 \tag{50.3}$$

Theorem 50.2.6 (Third law). *No process can reach absolute zero in a finite sequence of operations.*

50.3 Gases

50.3.1 Ideal gases

Theorem 50.3.1 (Ideal gas law).

$$\boxed{PV = nRT} \tag{50.4}$$

Chapter 51

Statistical mechanics

51.1   Axioms

Theorem 51.1.1 (Ergodic principle). *All microstates corresponding to the same macroscopic state are equally propable.*

Theorem 51.1.2 (Boltzmann formula). *The central axiom of statistical mechanics gives following formula for the entropy:*

$$\boxed{S = k \ln \Omega(E, V, N, \alpha)} \tag{51.1}$$

where Ω denotes the number of microstates corresponding to the system at energy E, volume V, and so on.

51.2   Temperature

The temperature of a system in contact with a heat bath is defined as:

$$\boxed{T = \left(\frac{\partial E}{\partial S}\right)_V} \tag{51.2}$$

51.3   Maxwell-Boltzmann statistics

Consider a system of N indistinguishable non-interacting particles. Let ε_i be the energy associated with the i-th energy level with degeneracy g_i. The propability p_i of finding a particle in the i-th energy level is given by:

$$\boxed{p_i = \frac{g_i e^{-\beta \varepsilon_i}}{Z}} \tag{51.3}$$

where Z is the single particle **partition function** defined as:

$$Z = \sum_i g_i e^{-\beta \varepsilon_i}$$ (51.4)

51.4 Grand canonical system

Formula 51.4.1 (Grand canonical partition function).

$$\mathcal{Z}_i = \sum_{\varepsilon_i} e^{\beta n_i (\mu - \varepsilon_i)}$$ (51.5)

Corollary 51.4.2. In the case that $n_i \in \{0, 1\}$ this formula reduces to $\mathcal{Z}_i = e^{\beta \mu} Z_i$.

Definition 51.4.3 (Fugacity).

$$z = e^{\mu N}$$ (51.6)

51.5 Energy

Theorem 51.5.1 (Virial theorem).

$$\langle T \rangle = -\frac{1}{2} \sum_k \langle \vec{r}_k \cdot \vec{F}_k \rangle$$ (51.7)

Corollary 51.5.2. For potentials of the form $V = a r^{-n}$ this becomes:

$$2 \langle T \rangle = -n \langle V \rangle$$ (51.8)

Theorem 51.5.3 (Equipartition theorem). *Let x be any generalized coordinate (both position or momentum).*

$$\left\langle x^k \frac{\partial H}{\partial x^l} \right\rangle = \delta_{kl} k_b T$$ (51.9)

Corollary 51.5.4. For quadratic Hamiltonians this can be rewritten using Euler's theorem for homogeneous functions 5.10 as:

$$\langle T \rangle = \frac{1}{2} k_b T$$ (51.10)

Chapter 52

Photon gas

52.1 Black-body radiation

Formula 52.1.1 (Planck's law).

$$B_\nu(\nu, T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{\frac{h\nu}{kt}} - 1}$$
(52.1)

Formula 52.1.2 (Wien's displacement law).

$$\lambda_{max} T = b$$
(52.2)

where $b = 2.897\,772\,9(17) \times 10^{-3}$ Km is **Wien's displacement constant**.

Part XII

Solid State Physics

Chapter 53

Material physics

53.1   Crystals

Theorem 53.1.1 (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 53.1.2 (Zone). The collection of faces parallel to a given axis, is called a zone. The axis itself is called the zone axis.

53.1.1   Analytic representation

Definition 53.1.3 (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 53.1.4. Negative numbers are written as $\bar{a}$ instead of $-a$.

Formula 53.1.5 (Coordinates of 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 $(h_2\ k_2\ l_2)$, pointing in the direction of the point (au, bv, cw) is denoted by $[u\ v\ w]$. Where

$$u = \begin{vmatrix} k_1 & l_1 \\ k_2 & l_2 \end{vmatrix} \qquad v = \begin{vmatrix} l_1 & h_1 \\ l_2 & h_2 \end{vmatrix} \qquad w = \begin{vmatrix} h_1 & k_1 \\ h_2 & k_2 \end{vmatrix} \tag{53.1}$$

Theorem 53.1.6 (Hauy's law of rational indices). *The Miller indices of every natural face of a crystal will always have rational proportions.*

53.2   Symmetries

Definition 53.2.1 (Equivalent planes/axes). When applying certain symmetries to a plane or axis, it often occurs that we obtain a set of equivalent planes/axes. These equivalence classes are denoted respectively by $\{h\ k\ l\}$ and $< h\ k\ l >$.

Property 53.2.2 (Rotational symmetry). Only $1, 2, 3, 4$ and 6-fold rotational symmetries can occur.

53.3 Crystal lattice

Formula 53.3.1. For an orthogonal crystal lattice, the distance between planes of the family $(h\ k\ l)$ is given by:

$$d_{hkl} = \frac{1}{\sqrt{\left(\dfrac{h}{a}\right)^2 + \left(\dfrac{k}{b}\right)^2 + \left(\dfrac{l}{c}\right)^2}} \tag{53.2}$$

53.3.1 Bravais lattice

Definition 53.3.2 (Bravais lattice). A crystal lattice generated by a certain point group symmetry is called a Bravais lattice. 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 to a point group.

Definition 53.3.3 (Wigner-Seitz cell). The part of space consisting of all points closer to a given lattice point than to any other.

53.3.2 Reciprocal lattice

Formula 53.3.4 (Reciprocal basis vectors). The reciprocal lattice corresponding to a given Bravais lattice with primitive basis $\{\vec{a}, \vec{b}, \vec{c}\}$ is defined by the following reciprocal basis vectors

$$\vec{a}^* = 2\pi \frac{\vec{b} \times \vec{c}}{\vec{a} \cdot (\vec{b} \times \vec{c})} \tag{53.3}$$

The vectors $\vec{b}^*$ and $\vec{c}^*$ are obtained by permutation of (a, b, c). These vectors satisfy the relations

$$\vec{a} \cdot \vec{a}^* = 2\pi$$
$$\vec{b} \cdot \vec{b}^* = 2\pi \tag{53.4}$$
$$\vec{c} \cdot \vec{c}^* = 2\pi$$

$$\tag{53.5}$$

Notation 53.3.5 (Reciprocal lattice vector). The reciprocal lattice vector $\vec{r}^*_{hkl}$ is defined as follows:

$$\vec{r}^*_{hkl} = h\vec{a}^* + k\vec{b}^* + l\vec{c}^* \tag{53.6}$$

Property 53.3.6. 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.

- $||\vec{r}^{\,*}_{hkl}|| = \frac{2\pi n}{d_{hkl}}$

53.4 Diffraction

53.4.1 Constructive interference

Formula 53.4.1 (Laue conditions). Suppose that an incident beam makes angles α_0, β_0 and γ_0 with the lattice axes. The diffracted beam making angles α, β and γ with the axes will be observed if following conditions are satisfied:

$$a(\cos\alpha - \cos\alpha_0) = h\lambda$$
$$b(\cos\beta - \cos\beta_0) = k\lambda$$
$$c(\cos\gamma - \cos\gamma_0) = l\lambda$$

If these conditions have been met then we observe a diffracted beam of order hkl.

Remark 53.4.2. 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

- a rotating crystal and a monochromatic beam

Formula 53.4.3 (Vectorial Laue conditions). Let $\vec{k}_0, \vec{k}$ denote the wave vector of respectively the incident and diffracted beams. The Laue conditions can be reformulated in the following way:

$$\boxed{\vec{k} - \vec{k}_0 = \vec{r}^{\,*}_{hkl}} \tag{53.7}$$

Formula 53.4.4 (Bragg's law). Another equivalent formulation of the Laue conditions is given by following formula:

$$\boxed{2d_{hkl}\sin\theta = n\lambda} \tag{53.8}$$

where

λ : wavelength of the incoming beam

θ : the **Bragg angle**

d_{hkl} : distance between neighbouring planes

Remark 53.4.5. The angle between the incident and diffracted beams is given by 2θ.

Construction 53.4.6 (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 $\vec{k}_i$. Now construct a sphere with radius $\frac{2\pi}{\lambda}$ centered on the start of $\vec{k}_i$. All points on the sphere that coincide with a reciprocal lattice point satisfy the vectorial Laue condition 53.7. Therefore Bragg diffraction will occur in the direction of all the intersections of the Ewald sphere and the reciprocal lattice.

53.4.2 Intensity of diffracted beams

Definition 53.4.7 (Systematic extinctions). Every particle in the motive emits its own waves. These waves will interfere and some will cancel out which leads to the absence of certain diffraction spots. These absences are called systematic extinctions.

Definition 53.4.8 (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 53.4.9 (Structure factor). The waves coming from the individual atoms in the motive can also be combined, again taking into account the different phases, into a resulting wave. The amplitude of this wave is called the structure factor and it is given by:

$$F(hkl) = \sum_j f_j \exp\left[2\pi i (hx_j + ky_j + lz_j)\right] \tag{53.9}$$

where f_j is the atomic scattering factor of the jth atom in the motive.

Example 53.4.10. A useful example of systematic extinctions is the structure factor of an FCC or BCC lattice for the following specific situations:

If $h + k + l$ is odd, then $F(hkl) = 0$ for a BCC lattice. If h, k and l are not all even or all odd then $F(hkl) = 0$ for an FCC lattice.

Definition 53.4.11 (Laue indices). Higher order diffractions can be rewritten as a first order diffraction in the following way:

$$2d_{nhnknl} \sin\theta = \lambda \qquad \text{with} \qquad d_{nhnknl} = \frac{d_{hkl}}{n} \tag{53.10}$$

Following from the interpretation of the Bragg law as diffraction being a reflection at the lattice plane $(h\ k\ l)$ we 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.

53.5 Alloys

53.5.1 Hume-Rothery conditions

An element can be dissolved in a metal (forming a solid solution) if the following conditions are met:

- The difference between the atomic radii is $\leq 15\%$.

- The crystal structures are the same.

- The elements have a similar electronegativity.

- The valency is the same.

53.6   Lattice defects

53.6.1   Point defects

Definition 53.6.1 (Vacancy). A lattice point where an atom is missing. Also called a **Schottky defect**.

Formula 53.6.2 (Concentration of Schottky defects[†]). Let N denote the number of lattice points and n the number of vacancies. The following relation gives the temperature dependence of Schottky defects:

$$\frac{n}{n+N} = e^{-E_v/kT} \tag{53.11}$$

where T is the temperature and E_v the energy needed to create a vacancy.

Remark. A similar relation holds for interstitials.

Definition 53.6.3 (Interstitial). An atom placed at a position which is not a lattice point.

Definition 53.6.4 (Frenkel pair). An atom displaced from a lattice point to an interstitial location (hereby creating a vacancy-interstitial pair) is called a Frenkel defect.

Formula 53.6.5 (Concentration of Frenkel pairs). Let n_i denote the number of atoms displaced from the bulk of the lattice to any N_i possible interstitial positions and thus creating n_i vacancies. The following relation holds:

$$\frac{n_i}{\sqrt{NN_i}} = e^{-E_{fr}/2kT} \tag{53.12}$$

where E_{fr} denotes the energy needed to create a Frenkel pair.

Remark 53.6.6. In compounds the number of vacancies can be much higher than in mono-atomic lattices.

Remark 53.6.7. The existence of these defects creates the possibility of diffusion.

53.7   Electrical properties

53.7.1   Charge carriers

Formula 53.7.1 (Conductivity). Definition 37.1 can be modified to account for both positive and negative charge carriers:

$$\sigma = n_n q_n \mu_n + n_p q_p \mu_p \tag{53.13}$$

Remark. The difference between the concentration of positive and negative charge carriers can differ by orders of magnitude across different materials. It can differ by up to 20 orders of magnitude.

53.7.2 Band structure

Definition 53.7.2 (Valence band). The energy band corresponding to the outermost (partially) filled atomic orbital.

Definition 53.7.3 (Conduction band). The first unfilled energy band.

Definition 53.7.4 (Band gap). The energy difference between the valence and conduction bands (if they do not overlap). It is the energy zone[1] where no electron states can exist.

Definition 53.7.5 (Fermi level). The energy level having a 50% chance of being occupied at thermodynamic equilibrium.

Formula 53.7.6 (Fermi function). The following distribution gives the propability of a state with energy E_i being occupied by an electron:

$$\boxed{f(E_i) = \frac{1}{e^{(E_i - E_f)/kT} + 1}}$$
(53.14)

where E_f is the Fermi level as defined above.

53.7.3 Intrinsic semiconductors

Formula 53.7.7. Let n denote the charge carrier density as before. We find the following temperature dependence:

$$n \propto e^{-E_g/2kt}$$
(53.15)

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 $E_g \gg kT$.

53.7.4 Extrinsic semiconductors

Definition 53.7.8 (Doping). Intentionally introducing impurities to modify the (electrical) properties.

Definition 53.7.9 (Acceptor). Group III element added to create an excess of holes in the valence band. The resulting semiconductor is said to be a **p-type semiconductor**.

Definition 53.7.10 (Donor). Group IV element added to create an excess of electrons in the valence band. The resulting semiconductor is said to be an **n-type semiconductor**.

---

[1]For a basic derivation see [11].

53.7.5   Ferroelectricity

Some materials can exhibit certain phase transitions between a paraelectric and ferroelectric state.

Paraelectric materials have the property that the polarisation $\vec{P}$ and the electric field $\vec{E}$ 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 a symmetry breaking, i.e. the ions in the lattice have been shifted out of their 'central' positions and induce 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 53.7.11. Ferroelectricity can only occur in crystals with unit cells that do not have a center of symmetry. This would rule out the possiblity of having the asymmetry needed for the dipole moment.

Definition 53.7.12 (Saturation polarization). The maximum polarization obtained by a ferroelectric material. It it obtained when the domain formation also reaches a maximum.

Definition 53.7.13 (Remanent polarization). The residual polarization of the material when the external electric field is turned off.

Definition 53.7.14 (Coercive field). The electric field needed to cancel out the remanent polarization.

Definition 53.7.15 (Piezoelectricity). Materials that obtain a polarization when exposed to mechanical stress are called piezoelectric materials.

Remark 53.7.16. All ferroelectric materials are piezoelectric, but the converse is not true. All crystals without a center of symmetry are piezoelectric. This property is however only a necessary (and not a sufficient) condition for ferroelectricity, as mentioned above.

Example 53.7.17 (Transducer). A device that converts electrical to mechanical energy (and vice versa).

53.8   Magnetic properties

Definition 53.8.1 (Diamagnetism). In diamagnetic materials, the magnetization is oriented opposite to the applied field, so $B < H$. The susceptibility is small, negative and independent of the temperature.

Remark 53.8.2. All materials exhibit a diamagnetic character.

Definition 53.8.3 (Paramagnetism). The susceptibility is small, positive and inversely proportional to the temperature.

Definition 53.8.4 (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.

53.8.1 Paramagnetism

Formula 53.8.5 (Curie's law). If the interactions between the particles can be neglected, we obtain the following law:

$$\chi = \frac{C}{T} \tag{53.16}$$

Materials that satisfy this law are called **ideal paramagnetics**.

Formula 53.8.6 (Curie-Weiss law). If the interactions between particles cannot be neglected, we obtain the following law:

$$\chi = \frac{C}{T - \theta} \tag{53.17}$$

where $\theta = CN_W$ with N_W the **Weiss-constant**. This deviation of the Curie law is due to the intermolecular interactions that induce an internal magnetic field $H_m = N_W M$.

Formula 53.8.7 (Brillouin function B_J).

$$\boxed{B_J(y) = \frac{2J+1}{2J}\coth\left(\frac{2J+1}{2J}y\right) - \frac{1}{2J}\coth\left(\frac{y}{2J}\right)} \tag{53.18}$$

where $y = \frac{g\mu_B J B}{kT}$

Remark 53.8.8. Because $\coth(y \to \infty) \approx 1$ we have:

$$\text{if } T \to 0 \quad \text{then} \quad M = Ng\mu_B J B_J(y \to \infty) = Ng\mu_B J \tag{53.19}$$

This value is called the **absolute saturation magnetization**.

53.8.2 Ferromagnetism

Ferromagnetics are materials that have strong internal interactions which lead to large scale (with respect to the lattice constant) parallel ordering of the atomic magnetic (dipole) moments. This leads to the spontaneous magnetization of the material and consequently a nonzero total dipole moment.

Remark. In reality, ferromagnetic materials do not always spontaneously possess a magnetic moment in the absence of an external field. When stimulated by a small external field, they will however display a magnetic moment, much larger than paramagnetic materials would.

Definition 53.8.9 (Domain). The previous remark 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 then 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 53.8.10 (Bloch walls). A wall between two magnetic domains.

Definition 53.8.11 (Ferromagnetic Curie temperature). Above this this temperature the material loses its ferromagnetic properties and it becomes a paramagnetic material following the Curie-Weiss law.

Remark 53.8.12. For ferromagnetic (and ferrimagnetic) materials it is impossible to define a magnetic susceptibility as the magnetization is nonzero even in the absence of a magnetic field.[2] Above the critical temperature (Curie/Néel) it is however possible to define a susceptiblity as the materials become paramagnetic in this region.

53.8.3   Antiferromagnetism

When the domains in a magnetic material have an antiparallel ordering[3], the total dipole moment will be small. If the temperature rises, the thermal agitation however will disturb the orientation of the domains and the magnetic susceptibility will rise.

Definition 53.8.13 (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:

$$\chi = \frac{C}{T + \theta} \tag{53.20}$$

This resembles a generalization of the Curie-Weiss law with a negative and therefore virtual critical temperature.

53.8.4   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 53.8.14 (Néel hyperbola). Above the Néel temperature it is possible to define a susceptibility given by:

$$\frac{1}{\chi} = \frac{T}{C} - \frac{1}{\chi_0} - \frac{\sigma}{T - \theta'} \tag{53.21}$$

---

[2]This can be seen from equation 38.1: $M = \chi H$. The susceptibility should be infinite.

[3]This will occur if it is energetically more favourable.

53.9 Mathematical description

Theorem 53.9.1 (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.*

Part XIII

Appendices

Appendix A

Derivations: Mathematics

A.1 Linear algebra

A.1.1 Explanation for property 14.2.10

Pick a subspace $W \in \mathrm{Gr}(k, V)$. The stabilizer of W with respect to $GL(V)$ is the set

$$H_W = \{g \in GL(V) | g \cdot W = W\}$$

Due to the transitivity of the group action we have that

$$\forall X, Y \in Gr(k, V) : \exists h \in GL(V) : h \cdot X = Y$$

So for every $U \in \mathrm{Gr}(k, V)$ we can choose a group element g_U such that $g_U \cdot W = U$. For all elements in the coset $g_U H_W = \{g_U h \in GL(V) | h \in H_W\}$ the following equality is satisfied:

$$(g_U h_W) \cdot W = g_U \cdot (h_W W) = g_U \cdot W = U$$

This implies that the map $\Phi : GL(V)/H_W \to \mathrm{Gr}(k, V)$ is surjective.

Now we need to prove that Φ is also injective. We give a proof by contradiction. Choose two distinct cosets pH_W and qH_W. Then there exist two subspaces $P, Q \in \mathrm{Gr}(k, V)$ such that $p \cdot W = P$ and $q \cdot W = Q$. Now assume that $P = Q$. This means that

$$
\begin{aligned}
& p \cdot W = q \cdot W \\
\Longleftrightarrow\ & (q^{-1}p) \cdot W = W \\
\Longleftrightarrow\ & q^{-1}p \in H_W \\
\Longleftrightarrow\ & qH_W \ni q(q^{-1}p) = p
\end{aligned}
$$

This would imply that $pH_W = qH_W$ which is in contradiction to our assumption. It follows that $P \neq Q$ and that Φ is injective. $\qquad\square$

A.1.2 Proof for the equality of definitions 18.3.14 and 18.3.15

$$(u+v) \otimes (u+v) - u \otimes u - v \otimes v = u \otimes v + v \otimes u \tag{A.1}$$

The LHS is an element of the ideal I generated by $\{v \otimes v | v \in V\}$. Using the ideal generated by elements such as in the RHS gives the usual definition of the exterior algebra based on the wedge product as defined in 18.17 because it imposes the relation $u \wedge v = -v \wedge u$.

We do however have to pay attention to one little detail. As mentioned in 18.3.15 the general definition uses the ideal I to construct the quotient space. The other construction is only equivalent when working over a field with a characteristic different from 2. This follows from the fact that we have to divide by 2 when trying to obtain the ideal I from the RHS by setting $u = v$.

A.2 Vector fields & differential forms

A.2.1 Explanation for example 24.4.11

In this derivation we use the Landau little-o notation $o(t)$, i.e.:

$$\lim_{t \to 0} \frac{o(t)}{t} = 0 \tag{A.2}$$

Now assume that X is a smooth vector field and f is a smooth function. Because the Lie derivative is a local operation we can work in a local chart such that γ is (again locally) equivalent to a curve[1] $\beta_p : U \to \mathbb{R}^n$ and such that we can expand $\beta_p(t)$ around $p \in U$:

$$\begin{aligned}
\mathcal{L}_X f(p) &= \lim_{t \to 0} \left[\frac{f(\beta_p(0) + t\beta_p'(0) + o(t)) - f(p)}{t} \right] \\
&= \lim_{t \to 0} \left[\frac{f(p + tX(p) + o(t)) - f(p)}{t} \right] \\
&= \lim_{t \to 0} \left[\frac{f(p) + tDf(p) \cdot X(p) + o(t) - f(p)}{t} \right] \\
&= \sum_k \frac{\partial f}{\partial x^k}(p) X_k(p) + \lim_{t \to 0} \frac{o(t)}{t} \\
&= \sum_k \frac{\partial f}{\partial x^k}(p) X_k(p) \tag{A.3}
\end{aligned}$$

where we used the defining condition 24.26 for integral curves on line 2. If we now rewrite this equation as an operator equality, we obtain:

$$\boxed{\mathcal{L}_X = \sum_k X_k \frac{\partial}{\partial x^k}} \tag{A.4}$$

---

[1] 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.

A.2.2   Explanation for formula 24.4.12

For vector fields we cannot just take the difference at two different points because the tangent spaces generally do not coincide. We can solve this by using the flow 24.27:

$$\mathcal{L}_X Y = \lim_{t \to 0} \frac{(T\sigma_t)^{-1}[X(\gamma_p(t))] - X(p)}{t} \tag{A.5}$$

where the $T\sigma_t$ is the differential 24.3.4 of the flow which satisfies $(T\sigma)^{-1} = T\sigma_{-t}$. To see that this definition makes sense we have to show that $(T\sigma_t)^{-1}[X(\gamma_p(t))] \in T_p M$. This goes as follows:

$$
\begin{aligned}
(T\sigma_t)^{-1}[X(\gamma_p(t))](f) &= T\sigma_{-t}[X(\gamma_p(t))](f) \\
&= X(\sigma_{-t} \circ \gamma_p(t))(f \circ \sigma_{-t}) \\
&= X(\sigma_{-t} \circ \sigma_t(p))(f \circ \sigma_{-t}) \\
&= X(p)(f \circ \sigma_{-t}) \\
&\in T_p M
\end{aligned}
$$

for all $f \in C^k(M, \mathbb{R})$. On line 3 we used the definition of the flow 24.27.

We can also rewrite the second term in the numerator of A.5 using the flow:

$$X(p) = X(\sigma_0(p)) = T\sigma_0(X)$$

Using the definition of the pushforward of vector fields 24.23 the Lie derivative can be rewritten as:

$$
\begin{aligned}
\mathcal{L}_X Y &= \lim_{t \to 0} \frac{\sigma_{-t*} X(\gamma_p(t)) - \sigma_{0*} X(\gamma_p(0))}{t} \\
&= \frac{d}{dt}(\sigma_{-t*} X)(\gamma_p(t)) \Big|_{t=0}
\end{aligned}
$$

Or finally by using the relation between pushforward and pullback 24.24 this becomes:

$$\boxed{\mathcal{L}_X Y = \frac{d}{dt}(\sigma_t^* X)(\gamma_p(t)) \Big|_{t=0}} \tag{A.6}$$

A.2.3   Explanation for remark 24.5.10

Looking at formula 24.41 for the exterior derivative of a smooth function and remembering the definition of the gradient 15.2 we see that these two definitions appear very similar. The major difference lies in the fact that ∇f is a vector in $\mathbb{R}^3$ and df is a covector in $\mathbb{R}^{*3}$. However there exists an isomorphism between these spaces and so we can identify ∇f and df.

Similar relations hold for the rotor 15.8 and divergence 15.7, however here we have to use a different construction as we will be working with the spaces Λ^1 and Λ^2. However we can use the Hodge star 18.27 to obtain the correct dimensions.

Consider a vector $\vec{f} = (f_1, f_2, f_3)$ where f_i is smooth. Using these functions f_i we can construct a 1-form $\alpha = f_1 dx_1 + f_2 dx_2 + f_3 dx_3$ and a 2-form $\omega = f_1 dx_2 \wedge dx_3 + f_2 dx_3 \wedge dx_1 + f_3 dx_1 \wedge dx_2$. After applying the exterior derivative (in the corresponding spaces) we obtain:

$$d\alpha = \left(\frac{\partial f_3}{\partial x_2} - \frac{\partial f_2}{\partial x_3} \right) dx_2 \wedge dx_3 + \left(\frac{\partial f_1}{\partial x_3} - \frac{\partial f_3}{\partial x_1} \right) dx_3 \wedge dx_1 + \left(\frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2} \right) dx_1 \wedge dx_2$$

$$d\omega = \left(\frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} + \frac{\partial f_3}{\partial x_3} \right) dx_1 \wedge dx_2 \wedge dx_3$$

Using result 18.3.22 and the isomorphism $\sim : \mathbb{R}^{3*} \to \mathbb{R}^3$ we can rewrite this as:

$$\boxed{\sim df = \nabla f} \tag{A.7}$$

$$\boxed{\sim (*d\alpha) = \nabla \times \vec{f}} \tag{A.8}$$

$$\boxed{*d\omega = \nabla \cdot \vec{f}} \tag{A.9}$$

Appendix B

Derivations: Lagrangian formalism

B.1   d'Alembert's principle

In the following derivation we assume a constant mass.

$$\sum_k \left(\vec{F}_k - \dot{\vec{p}}_k \right) \dot{\vec{r}}_k = 0$$

$$\iff \sum_k \left(\vec{F}_k - \dot{\vec{p}}_k \right) \cdot \left(\sum_l \frac{\partial \vec{r}}{\partial q_l} \dot{q}_l \right) = 0$$

$$\iff \sum_l \left(\sum_k \vec{F}_k \cdot \frac{\partial \vec{r}}{\partial q_l} - \sum_k m \ddot{\vec{r}} \cdot \frac{\partial \vec{r}}{\partial q_l} \right) \dot{q}_l = 0$$

$$\iff \sum_l \left(Q_l - \sum_k m \ddot{\vec{r}} \cdot \frac{\partial \vec{r}}{\partial q_l} \right) \dot{q}_l = 0 \tag{B.1}$$

Now we look at the following derivative:

$$\frac{d}{dt} \left(\dot{\vec{r}} \cdot \frac{\partial \vec{r}}{\partial q_l} \right) = \ddot{\vec{r}} \cdot \frac{\partial \vec{r}}{\partial q_l} + \dot{\vec{r}} \cdot \frac{d}{dt} \left(\frac{\partial \vec{r}}{\partial q_l} \right)$$

$$\iff \ddot{\vec{r}} \cdot \frac{\partial \vec{r}}{\partial q_l} = \frac{d}{dt} \left(\dot{\vec{r}} \cdot \frac{\partial \vec{r}}{\partial q_l} \right) - \dot{\vec{r}} \cdot \frac{d}{dt} \left(\frac{\partial \vec{r}}{\partial q_l} \right)$$

$$\iff \ddot{\vec{r}} \cdot \frac{\partial \vec{r}}{\partial q_l} = \frac{d}{dt} \left(\dot{\vec{r}} \cdot \underbrace{\frac{\partial \vec{r}}{\partial q_l}}_{A} \right) - \dot{\vec{r}} \cdot \left(\frac{\partial \dot{\vec{r}}}{\partial q_l} \right) \tag{B.2}$$

326

To evaluate A we can take a look at another derivative:

$$\frac{\partial \dot{\vec{r}}}{\partial \dot{q}_l} = \frac{\partial}{\partial \dot{q}_l} \left(\sum_k \frac{\partial r}{\partial q_k} \dot{q}_k \right)$$

$$= \sum_k \frac{\partial r}{\partial q_k} \delta_{kl}$$

$$= \frac{\partial \vec{r}}{\partial q_l}$$

$$= A$$

Substituting this in formula B.2 gives:

$$\ddot{\vec{r}} \cdot \frac{\partial \vec{r}}{\partial q_l} = \frac{d}{dt} \left(\dot{\vec{r}} \cdot \frac{\partial \dot{\vec{r}}}{\partial \dot{q}_l} \right) - \dot{\vec{r}} \cdot \left(\frac{\partial \dot{\vec{r}}}{\partial q_l} \right)$$

$$= \frac{d}{dt} \left(\frac{1}{2} \frac{\partial \dot{\vec{r}}^2}{\partial \dot{q}_l} \right) - \frac{1}{2} \frac{\partial \dot{\vec{r}}^2}{\partial q_l} \tag{B.3}$$

If we multiply this by the mass m and sum over all particles we get :

$$\sum_k m_k \ddot{\vec{r}}_k \cdot \frac{\partial \vec{r}_k}{\partial q_l} = \frac{d}{dt} \frac{\partial}{\partial \dot{q}_l} \left(\sum_k \frac{1}{2} m \dot{\vec{r}}_k^2 \right) - \frac{\partial}{\partial q_l} \left(\sum_k \frac{1}{2} m \dot{\vec{r}}_k^2 \right)$$

$$= \frac{d}{dt} \frac{\partial T}{\partial \dot{q}_l} - \frac{\partial T}{\partial q_l} \tag{B.4}$$

Where we have denoted the total kinetic energy in the last line as T.
Plugging this result into formula B.1 gives us:

$$\sum_l \left(Q_l - \frac{d}{dt} \frac{\partial T}{\partial \dot{q}_l} - \frac{\partial T}{\partial q_l} \right) \dot{q}_l = 0 \tag{B.5}$$

As all the q_l are independent the following relation should hold for all l:

$$Q_l - \frac{d}{dt} \left(\frac{\partial T}{\partial \dot{q}_l} \right) - \frac{\partial T}{\partial q_l} = 0$$

$$\iff \boxed{\frac{d}{dt} \left(\frac{\partial T}{\partial \dot{q}_l} \right) - \frac{\partial T}{\partial q_l} = Q_l} \tag{B.6}$$

This last equation is known as a **Lagrange equation of the first kind**.
If we have a system with only conservative forces acting on it, we can write the force on the i-th particle as:

$$F_i = -\nabla_i V \tag{B.7}$$

With this in mind, lets take a look at the derivative of the potential V with respect to the l-th generalized coordinate:

$$\frac{\partial V}{\partial q_l} = \sum_i (\nabla_i V) \cdot \frac{\partial \vec{r}_i}{\partial q_l}$$

$$= -Q_l \tag{B.8}$$

The differentiation of V with respect to any generalized velocity $\dot{q}_l$ is trivially zero. This combined with the last formula B.8 and with formula B.6 gives:

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_l}\right) - \frac{\partial T}{\partial q_l} = Q_l$$

$$\Longleftrightarrow \frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_l}\right) - \frac{\partial T}{\partial q_l} = -\frac{\partial V}{\partial q_l} + \frac{\partial V}{\partial \dot{q}_l}$$

$$\Longleftrightarrow \frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_l} - \frac{\partial V}{\partial \dot{q}_l}\right) - \frac{\partial}{\partial q_l}(T - V) = 0 \tag{B.9}$$

If we introduce a new variable L, called the **Lagrangian**, we get the **Lagrangian equation of the second kind**:

$$\boxed{\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_l}\right) - \frac{\partial L}{\partial q_l} = 0} \tag{B.10}$$

B.2   Hamilton's principle

In this part we start from the principle of least action. First we the define the **action** as following:

$$\boxed{I = \int_{t_1}^{t_2} L\left(y(t), \dot{y}(t), t\right) dt} \tag{B.11}$$

Then we require that this action is minimal for the physically acceptable path. To do this we define a family of paths:

$$y(t, \alpha) = y(t) + \alpha \eta(t) \tag{B.12}$$

Where $\eta(t)$ is an arbitrary function with the following boundary conditions:

$$\begin{cases} \eta(t_1) = 0 \\ \eta(t_2) = 0 \end{cases} \tag{B.13}$$

If we define the action integral over this family of paths, the integral B.11 becomes a function of α:

$$I(\alpha) = \int_{t_1}^{t_2} L\left(y(t, \alpha), \dot{y}(t, \alpha), t\right) dt \tag{B.14}$$

Requiring that the action integral is stationary for $y(t)$ (thus $\alpha = 0$) is equivalent to:

$$\left(\frac{dI}{d\alpha}\right)_{\alpha=0} = 0 \tag{B.15}$$

This condition combined with formula B.14 gives us:

$$\frac{dI}{d\alpha} = \int_{t_1}^{t_2} \frac{d}{d\alpha} L\left(y(t, \alpha), \dot{y}(t, \alpha), t\right) dt \tag{B.16}$$

As we evaluate this derivative in $\alpha = 0$ we can replace $y(t, \alpha)$ by $y(t)$ due to definition B.12.

$$\frac{dI}{d\alpha} = \int_{t_1}^{t_2} \left[\frac{\partial L}{\partial y} \frac{\partial y}{\partial \alpha} + \frac{\partial L}{\partial \dot{y}} \frac{\partial \dot{y}}{\partial \alpha} \right] dt$$
$$= \int_{t_1}^{t_2} \left[\frac{\partial L}{\partial y} \eta(t) + \frac{\partial L}{\partial \dot{y}} \dot{\eta}(t) \right] dt \tag{B.17}$$

If we substitute $\frac{\partial L}{\partial \dot{y}} := h(t)$ and apply integration by parts to the second term in this integral, we get:

$$\frac{dI}{d\alpha} = \int_{t_1}^{t_2} \left[\frac{\partial L}{\partial y} \eta(x) + h(t)\dot{\eta}(t) \right] dt$$
$$= \int_{t_1}^{t_2} \left[\frac{\partial L}{\partial y} \eta(t) + h(t)\frac{d\eta}{dt} \right] dt$$
$$= \int_{t_1}^{t_2} \frac{\partial L}{\partial y} \eta(t) dt + \eta(t_2)h(t_2) - \eta(t_1)h(t_1) - \int_{t_1}^{t_2} \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{y}} \right) \eta(t) dt \tag{B.18}$$

Due to the initial conditions B.15 for the function $\eta(t)$, the two terms in the middle vanish and we obtain:

$$\frac{dI}{d\alpha} = \int_{t_1}^{t_2} \left[\frac{\partial L}{\partial y} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{y}} \right) \right] \eta(t) dt \tag{B.19}$$

Furthermore, as the function $\eta(t)$ was arbitrary, the only possible way that this derivative can become zero is when the integrand is identically zero:

$$\boxed{\frac{\partial L}{\partial y} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{y}} \right) = 0} \tag{B.20}$$

If we compare this result with formula B.10 we see that we can also obtain the **Lagrangian equations of the second kind** by starting from the principle of least action. (Where the variable y represents the generalized coordinates q_l and the variable $\dot{y}$ represents the generalized velocities $\dot{q}_l$)

Remark B.2.1. Differential equations of the form

$$\boxed{\frac{\partial f}{\partial y}(y, \dot{y}, x) = \frac{d}{dx}\left(\frac{\partial f}{\partial \dot{y}}(y, \dot{y}, x) \right)} \tag{B.21}$$

are known as **Euler-Lagrange equations**.

B.3   Explanation for Noether's theorem 49.1.1

The general transformation rule for the Lagrangian is:

$$\mathcal{L}(x) \to \mathcal{L}(x) + \alpha\delta\mathcal{L}(x) \tag{B.22}$$

To have a symmetry, i.e. keep the action invariant, the deformation factor has to be a 4-divergence:

$$\mathcal{L}(x) \to \mathcal{L}(x) + \alpha \partial_\mu \mathcal{J}^\mu(x) \tag{B.23}$$

To obtain formula 49.2 we vary the Lagrangian 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 B.20. Combining these formulas gives us:

$$\partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta\phi\right) - \partial_\mu\mathcal{J}^\mu(x) = 0 \tag{B.24}$$

From this equation we can conclude that the current

$$\boxed{j^\mu(x) = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta\phi - \mathcal{J}^\mu(x)} \tag{B.25}$$

is conserved.

Appendix C

Derivations: Optics and material physics

C.1 Optics

C.1.1 Law of Lambert-Beer 35.9

From formula 35.8 we now that the complex refractive index can be written as

$$\widetilde{n} = n + ik$$

Where k is called the **extinction coefficient**.
From classical optics we also know that in a material the speed of light obeys the following relation:

$$c = \widetilde{n}v$$

Where we have used the complex refractive index. It readily follows that the wavenumber (sadly also given the letter k) can be written as:

$$k = \frac{\omega}{v} = \widetilde{n}\frac{\omega}{c}$$

From classical electromagnetism we know that a plane wave can be written as:

$$E(x,t) = Re\left\{A \exp\left[i(kx - \omega t + \phi)\right]\right\}$$

So everything put together we get:

$$E(x,t) = Re\left\{A \exp\left[i\left((n + ik)\frac{\omega}{c}x - \omega t + \phi\right)\right]\right\}$$

or also:

$$E(x,t) = Re\left\{A \exp\left[in\frac{\omega}{c}x\right] \cdot \exp\left[-k\frac{\omega}{c}x\right] \cdot \exp\left[-i\omega t\right] \cdot \exp\left[i\phi\right]\right\}$$

We also know that the intensity is given by the following relation:

$$I(x) = |E(x)|^2 = E^*(x) \cdot E(x)$$

So only the second factor will remain. Dividing this by its value for $x = 0$ we get:

$$\frac{I(x)}{I(0)} = \frac{E(x) \cdot E^*(x)}{E(0) \cdot E^*(0)} = \exp\left[-\frac{2k\omega}{c}x\right] = \exp[-\alpha x]$$

Where α is the absorption coefficient as defined in formula 35.10. $\square$

Appendix D

Derivations: Classical and Statistical Mechanics

D.1 Moments of inertia

In this section we will always use formula 31.8 to calculate the moment of inertia.

D.1.1 Disk

The volume of a (solid) disk is given by:

$$V_{disk} = \pi R^2 d \tag{D.1}$$

where R is the radius and d is the thickness. The mass density is then given by:

$$\rho = \frac{M}{\pi R^2 d} \tag{D.2}$$

Using cylindrical coordinates the moment of inertia then becomes:

$$I = \frac{M}{\pi R^2 d} \int_0^{2\pi} d\varphi \int_0^d dz \int_0^R r^3 dr \tag{D.3}$$

$$= \frac{M}{\pi R^2 d} 2\pi d \frac{R^4}{4} \tag{D.4}$$

$$= \frac{1}{2} M R^2 \tag{D.5}$$

D.1.2 Solid sphere

The volume of a solid sphere is given by:

$$V_{sphere} = \frac{4}{3} \pi R^3 \tag{D.6}$$

where R is the radius. The mass density is then given by:

$$\rho = \frac{M}{\frac{4}{3}\pi R^3} \tag{D.7}$$

We will use spherical coordinates to derive the moment of inertia, but we have to be carefull. The r in formula 31.8 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 to be:

$$r = r' \sin \theta \tag{D.8}$$

where r' is the spherical coordinate. Now we can calculate the moment of inertia as follows:

$$I = \frac{M}{\frac{4}{3}\pi R^3} \int_0^{2\pi} d\varphi \int_0^R r'^4 dr' \int_0^\pi \sin^3 \theta d\theta \tag{D.9}$$

$$= \frac{M}{\frac{4}{3}\pi R^3} 2\pi \frac{R^5}{5} \frac{4}{3} \tag{D.10}$$

$$= \frac{2}{5} M R^2 \tag{D.11}$$

D.2 Schottky defects

Let E_v be the energy needed to remove a particle from its lattice point and move it to the surface. Furthermore we will neglect any surface effects and assume that the the energy E_v is indepednet 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

$$\Omega = \frac{(N+n)!}{n!N!} \tag{D.12}$$

where we used the fact that the removal of n particles creates n more lattice points at the surface. Using Boltzmann's entropy formula 51.1 and Stirling's formula we obtain

$$S(N,n) = k \ln \Omega = k\big[(N+n)\ln(N+n) - n\ln n - N\ln N\big] \tag{D.13}$$

Using 51.2 we can find the temperature:

$$\frac{1}{T} = \left(\frac{\partial S}{\partial E}\right)_{N,V} = \frac{dS}{dn}\frac{dn}{dE} = \frac{k}{E_v}\ln\frac{N+n}{n} \tag{D.14}$$

which can be rewritten as

$$\boxed{\frac{n}{N+n} = \exp\left(-\frac{E_v}{kT}\right)} \tag{D.15}$$

The density of Frenkel defects can be derived analogously.

Appendix E

Units and symbols

pk	745.7 W

Table E.1: Units

Bibliography

[1] Analysis, Manifolds and Physics, Part 1: Basics (Y. Choquet-Bruhat, C. DeWitt-Morette & M. Dillard-Bleick)

[2] Analysis, Manifolds and Physics, Part 2 (Y. Choquet-Bruhat, C. DeWitt-Morette)

[3] An Introduction to Quantum Field Theory (Michael E. Peskin & Daniel V. Schroeder, Westview Press)

[4] An Introduction to Tensors and Group Theory for Physicists (Nadir Jeevanjee, Birkhäuser)

[5] Branch cuts: http://www.math.odu.edu/ jhh/ch45.PDF

[6] Classical Mechanics (Goldstein, Safko & Poole, Pearson New International Edition)

[7] Elementary Symplectic Topology and Mechanics (Franco Cardin, Springer)

[8] Field Quantization (Reinhardt Greiner, Springer)

[9] Invariant Subspaces (H. Radjavi & P. Rosenthal)

[10] Principal Bundles: The Classical Case (Stephen Bruce Sontz, Springer)

[11] Quantum Mechanics (B.H. Bransden & C.J. Joachain, Prentice Hall)

[12] Representation Theory: A First Course (William Fulton & Joe Harris, Springer)

[13] Topology and Geometry for Physicists (Charles Nash & Siddharta Sen, Dover Publications)

Index