Concepts and Methods of Mathematical Physics

Catherine Meusburger [1]

Department Mathematik
FAU Erlangen-Nürnberg
Bismarckstr. 1 1/2
D - 91054 Erlangen
Germany

August 12, 2011

[1] catherine.meusburger@math.uni-erlangen.de

Disclaimer

This is a *preliminary draft version* of the lecture notes for the course "Concepts and Methods of Mathematical Physics", which was held as an intensive course for Master level students October 5-16 2009 and October 4-15 2010 at Hamburg University.

The lecture notes still contain many typos and mistakes, which we will try to correct gradually. However, we decline responsibility for any confusion that arises from them in the meantime. We are grateful for information about mistakes and also for suggestions and comments about these notes. These can be emailed to:

`catherine.meusburger@math.uni-erlangen.de`.

August 12, 2011, Catherine Meusburger

Contents

References and further reading

Chapter 1

- Frank Warner, Foundations of differentiable manifolds and Lie groups, Chapter 2
- R. Abraham, J. E. Marsden, R. Ratiu, Tensor Analysis, manifolds and applications, Chapter 5,6
- Theodor Bröcker, Lineare Algebra und analytische Geometrie: Ein Lehrbuch für Physiker und Mathematiker, Chapter 7
- Otto Forster, Analysis 3: Integralrechnung im $\mathbb{R}^n$ mit Anwendungen
- Henri Cartan, Differential forms
- Harley Flanders, Differential forms with applications to the physical sciences
- Jean A. Dieudonné, Treatise on Analysis 1, Appendix Linear Algebra

Chapter 2

- Andrew Baker, Matrix groups: an introduction to Lie group theory
- Brian C. Hall, Lie groups, Lie algebras, and representations: an elementary introduction
- Jürgen Fuchs, Christoph Schweigert, Symmetries, Lie Algebras and Representations: A Graduate Course for Physicists

Chapter 3

- Domenico Giulini, Algebraic and geometric structures of Special Relativity, http://arxiv.org/abs/math-ph/0602018
- Domenico Giulini, The Rich Structure of Minkowski Space, http://arxiv.org/abs/0802.4345

Chapter 4

- Otto Forster, Analysis 3: Integralrechnung im $\mathbb{R}^n$ mit Anwendungen
- Jean A. Dieudonné, Treatise on Analysis 1
- Jean A. Dieudonné, Treatise on Analysis 2

Chapter 5

- Jean A. Dieudonné, Treatise on Analysis 2
- William Averson: A short course on spectral theory
- John v. Neumann, Mathematical Foundations of Quantum Mechanics
- N. P. Landsman, Lecture notes on C*-algebras, Hilbert C*-modules, and quantum mechanics, http://arxiv.org/abs/math-ph/9807030
- Domenico Giulini, That strange procedure called quantisation, http://arxiv.org/abs/quant-ph/0304202

Chapter 1

Tensors and differential forms

1.1 Vector spaces

1.1.1 Notation and conventions

Vector spaces and related structures play an important role in physics because they arise whenever physical systems are linearised, i.e. approximated by linear structures. Linearity is a very strong tool. Non-linear systems such as general relativity or the fluid dynamics governed by Navier Stokes equation are very difficult to handle.

In this section, we consider finite dimensional vector spaces over the field $k = \mathbb{R}$ or $k = \mathbb{C}$ and linear maps between them. Vectors $\boldsymbol{x} \in V$ are characterised uniquely by their *coefficients with respect to a basis* $B = \{e_1, ...e_n\}$ of V

$$\boldsymbol{x} = \sum_{i=1}^{n} x^i e_i. \tag{1.1}$$

The set of linear maps $\phi : V \to W$ between vector spaces V and W over k, denoted $\mathrm{Hom}(V, W)$, also has the structure of a vector space with scalar multiplication and vector space addition given by

$$(\phi + \psi)(\boldsymbol{v}) = \phi(\boldsymbol{v}) + \psi(\boldsymbol{v}) \qquad (t\phi)(\boldsymbol{v}) = t\phi(\boldsymbol{v}) \qquad \forall \psi, \phi \in \mathrm{Hom}(V, W), t \in k. \tag{1.2}$$

A linear map $\phi \in \mathrm{Hom}(V, W)$ is called a *homomorphism* of vector spaces. It is called

- *monomorphism* if it is injective: $\phi(\boldsymbol{x}) = \phi(\boldsymbol{y})$ implies $\boldsymbol{x} = \boldsymbol{y}$
- *epimorphism* if it is surjective: $\forall \boldsymbol{x} \in W$ there exists a $\boldsymbol{y} \in V$ such that $\Phi(\boldsymbol{y}) = \boldsymbol{x}$
- *isomorphism* if it is bijective, i.e. injective and surjective. If an isomorphism between V and W exists, $\dim(V) = \dim(W)$ and we write $V \cong W$
- *endomorphism* if $V = W$. We denote the space of endomorphisms of V by $\mathrm{End}(V)$
- *automorphism* if it is a bijective endomorphism. We denote the space of automorphisms of V by $\mathrm{Aut}(V)$.

Linear maps $\phi \in \mathrm{Hom}(V, W)$ are characterised uniquely by their matrix elements with respect to a basis $B = \{e_1, ..., e_n\}$ of V and a basis $B' = \{f_1, ..., f_m\}$ of W

$$\phi(e_i) = \sum_{i=1}^{n} \phi_i^{\ j} f_j. \tag{1.3}$$

The matrix $A_\phi = (\phi^i_{\ j})$ is called the *representing matrix* of the linear map ϕ with respect to B and B'. The *transformation of the coefficients* of a vector $\boldsymbol{x} \in V$ under a linear map $\phi \in \mathrm{Hom}(V, W)$ is given by

$$\boldsymbol{x}' = \phi(\boldsymbol{x}) = \sum_{j=1}^{m} x'^j f_j \qquad x'^j = \sum_{i=1}^{n} \phi_i^{\ j} x^i. \tag{1.4}$$

1.1.2 Constructions with vector spaces

We will now consider the basic constructions that allow us to create new vector spaces from given ones. The first is the quotient of a vector space V by a linear subspace U. It allows one to turn a linear map into an *injective* linear map.

Definition 1.1.1: (Quotients of vector spaces)

Let U be a linear subspace of V. This is a subset $U \subset V$ that contains the null vector and is closed under the addition of vectors and under multiplication with k and therefore has the structure of a vector space.

We set $\boldsymbol{v} \sim \boldsymbol{w}$ is there exists an $\boldsymbol{u} \in U$ such that $\boldsymbol{w} = \boldsymbol{u} + \boldsymbol{v}$. Then $\sim$ defines an *equivalence relation* on V, i. e. it satisfies

1. *reflexivity*: $\boldsymbol{w} \sim \boldsymbol{w}$ for all $\boldsymbol{w} \in V$
2. *symmetry*: $\boldsymbol{v} \sim \boldsymbol{w} \Rightarrow \boldsymbol{w} \sim \boldsymbol{v}$ for all $\boldsymbol{v}, \boldsymbol{w} \in V$
3. *transitivity*: $\boldsymbol{u} \sim \boldsymbol{v}$, $\boldsymbol{v} \sim \boldsymbol{w} \Rightarrow \boldsymbol{u} \sim \boldsymbol{w}$ for all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$.

The *quotient* V/U is the set of *equivalence classes* $[\boldsymbol{v}] = \boldsymbol{v} + U = \{\boldsymbol{w} \in V \mid \boldsymbol{w} \sim \boldsymbol{v}\}$. It has the structure of a vector space with null vector $[0] = 0 + U = U$ and with addition and scalar multiplication

$$[\boldsymbol{v}] + [\boldsymbol{w}] = [\boldsymbol{v} + \boldsymbol{w}] \qquad t[\boldsymbol{v}] = [t\boldsymbol{v}] \qquad \forall \boldsymbol{v}, \boldsymbol{w} \in V, t \in \mathbb{R}.$$

Remark 1.1.2: It is important to show that the addition and scalar multiplication of equivalence classes are well-defined, i.e. that the result does not depend on the choice of the representative. In other words, we have to show that $[\boldsymbol{v}] = [\boldsymbol{v}']$ implies $t[\boldsymbol{v}] = t[\boldsymbol{v}']$ and $[\boldsymbol{v}] = [\boldsymbol{v}']$, $[\boldsymbol{w}] = [\boldsymbol{w}']$ implies $[\boldsymbol{v} + \boldsymbol{w}] = [\boldsymbol{v}' + \boldsymbol{w}']$.

If $[\boldsymbol{v}] = [\boldsymbol{v}']$, we have $\boldsymbol{v} - \boldsymbol{v}' \in U$. As U is a linear subspace of V this implies $t(\boldsymbol{v} - \boldsymbol{v}') = t\boldsymbol{v} - t\boldsymbol{v}' \in U$ for all $t \in \mathbb{R}$ and therefore $[t\boldsymbol{v}] = [t\boldsymbol{v}']$.

If $[\boldsymbol{v}] = [\boldsymbol{v}']$ and $[\boldsymbol{w}] = [\boldsymbol{w}']$, we have $\boldsymbol{v} - \boldsymbol{v}' \in U$, $\boldsymbol{w} - \boldsymbol{w}' \in U$. As U is a linear subspace of V, this implies $(\boldsymbol{v} - \boldsymbol{v}') + (\boldsymbol{w} - \boldsymbol{w}') = (\boldsymbol{v} + \boldsymbol{w}) - (\boldsymbol{v}' + \boldsymbol{w}') \in U$. Hence, we have $[\boldsymbol{v} + \boldsymbol{w}] = [\boldsymbol{v}' + \boldsymbol{w}']$. The vector space addition and scalar multiplication on V/U are therefore well-defined and V/U has the structure of a vector space with null vector $[0] = U$.

Example 1.1.3: We consider the vector space $V = \mathbb{R}^2$ and the linear subspace $U = \text{Span}(\boldsymbol{x}) = \{t\boldsymbol{x} \mid t \in \mathbb{R}\}$, where $\boldsymbol{x} \in \mathbb{R}^2 \setminus \{0\}$. Then, an equivalence class $[\boldsymbol{y}] = \{\boldsymbol{y}+t\boldsymbol{x} \mid t \in \mathbb{R}\}$ is a line through $\boldsymbol{y} \in \mathbb{R}^2$ with "velocity vector" $\boldsymbol{x}$. The quotient space $\mathbb{R}^2/U$ is the set of lines through points $\boldsymbol{y} \in \mathbb{R}^2$. The sum $[\boldsymbol{y}+\boldsymbol{z}]$ of two lines $[\boldsymbol{y}]$, $[\boldsymbol{z}]$ is a line through $\boldsymbol{y}+\boldsymbol{z}$ and the scalar multiplication of a line $[\boldsymbol{y}]$ with $t \in \mathbb{R}$ is the line $[t\boldsymbol{y}]$ through $t\boldsymbol{y}$.

Example 1.1.4: Consider a linear map $\phi \in \text{Hom}(V, W)$. Then $\ker(\phi) = \{\boldsymbol{v} \in V \mid \phi(\boldsymbol{v}) = 0\}$ is a linear subspace of V. We can define a linear map $\tilde{\phi} : V/\ker(\phi) \to W$ by setting

$$\tilde{\phi}([\boldsymbol{v}]) = \phi(\boldsymbol{v}) \qquad \forall \boldsymbol{v} \in V.$$

This map is well defined since $[\boldsymbol{v}] = [\boldsymbol{v}']$ if and only if there exists a $\boldsymbol{u} \in \ker(\phi)$ such that $\boldsymbol{v}' = \boldsymbol{v} + \boldsymbol{u}$. In this case, we have $\phi(\boldsymbol{v}') = \phi(\boldsymbol{v}) + \phi(\boldsymbol{u}) = \phi(\boldsymbol{v})$. It is linear since $\tilde{\phi}([\boldsymbol{v}]+[\boldsymbol{w}]) = \tilde{\phi}([\boldsymbol{v}+\boldsymbol{w}]) = \phi(\boldsymbol{v}+\boldsymbol{w}) = \phi(\boldsymbol{v})+\phi(\boldsymbol{w}) = \tilde{\phi}([\boldsymbol{v}])+\tilde{\phi}([\boldsymbol{w}])$ and $\tilde{\phi}(t[\boldsymbol{v}]) = \tilde{\phi}([t\boldsymbol{v}]) = \phi(t\boldsymbol{v}) = t\phi(\boldsymbol{v}) = t\tilde{\phi}([\boldsymbol{v}])$. The quotient construction therefore allows one to construct an *injective* map $\phi' \in \text{Hom}(V/\ker(\phi), W)$ from a linear map $\phi \in \text{Hom}(V, W)$.

Exercise 1: Show that for any vector space V and any linear map $\phi \in \text{End}(V)$, we have $V \cong \ker(\phi) \oplus \text{Im}(\phi)$. Show that $V/\ker(\phi) \cong \text{Im}(\phi)$.

Hint: Choose a basis $B_1 = \{e_1, .., e_k\}$, $e_i = \phi(g_i)$ of $\text{Im}(\phi)$ and complete it to a basis $B = B_1 \cup \{f_1, ..., f_{n-k}\}$ of V. By subtracting suitable linear combinations from the basis vectors $f_1, ..., f_n$, you can construct vectors that lie in $\ker(\phi)$.

We now consider the *dual of a vector space.* Dual vector spaces play an important role in physics. Among others, they encode the relations between particles and anti-particles. In particle physics, elementary particles are given by representations of certain Lie algebras on vector spaces. The duals of those vector spaces correspond to the associated anti-particles.

Definition 1.1.5: (Dual of a vector space, dual basis)

1. The *dual of a vector space* V over k, denoted V^*, is the space $\text{Hom}(V, k)$ of *linear forms on* V, i.e. of linear maps

$$\alpha : V \to k \qquad \alpha(t\boldsymbol{x} + s\boldsymbol{y}) = t\alpha(\boldsymbol{x}) + s\alpha(\boldsymbol{y}) \qquad \forall t, s \in k, \boldsymbol{x}, \boldsymbol{y} \in V. \tag{1.5}$$

   We have $\dim(V) = \dim(V^*)$ and $(V^*)^* \cong V$ for any finite dimensional vector space V.

2. For any basis $B = \{e_1, ..., e_n\}$ of V there is a *dual basis* $B^* = \{e^1, ..., e^n\}$ of V^* characterised by $e^i(e_j) = \delta^i_j$. Any element of V^* can be expressed as linear combination of the basis elements as

$$\alpha = \sum_{i=1}^{n} \alpha_i e^i \qquad \alpha_i \in k. \tag{1.6}$$

The choice of a basis and its dual gives rise to an isomorphism $\phi : V \to V^*$ defined by

$$\boldsymbol{v} = \sum_{i=1}^{n} v^i e_i \mapsto \phi(\boldsymbol{v}) = \sum_{i,j=1}^{n} \delta^{ij} v^i e_j = \sum_{i=1}^{n} v^i e^i.$$

Remark 1.1.6: Note that the identification between a vector space and its dual is *not canonical*, as it makes use of the choice of a basis and depends on this choice. With a different choice of a basis one obtains a different isomorphism $\phi : V \to V^*$.

Note also that $(V^*)^* \cong V$ does in general not hold for infinite dimensional vector spaces. This is due to the fact that there is no good isomorphism $V \to V^*$.

Definition 1.1.7: (Dual of a linear map)

For any linear map $\phi \in \mathrm{Hom}(V, W)$ there is a *dual linear map* $\phi^* \in \mathrm{Hom}(W^*, V^*)$ defined by

$$\phi^*(\alpha) = \alpha \circ \phi \qquad \forall \alpha \in W^*. \tag{1.7}$$

If the transformation of a basis B and of the coefficients under a linear map $\phi \in \mathrm{Hom}(V, W)$ are given by, respectively, (1.3) and (1.4), the transformation of the dual basis B^* and of the coefficients under the dual ϕ^* takes the form

$$\phi^*(f^j) = \sum_{i=1}^n {\phi_i}^j f^i \qquad \phi^*(\alpha)_i = \sum_{j=1}^n {\phi_i}^j \alpha_j. \tag{1.8}$$

The representing matrix A_{ϕ^*} of the dual map ϕ^* is the *transpose* of the representing matrix of ϕ: $A_{\phi^*} = A_\phi^T$.

Lemma 1.1.8: The duals ϕ^* of linear maps $\phi \in \mathrm{Hom}(V, W)$ satisfy the relations:

$$\begin{aligned} &(\mathrm{id}_V)^* = \mathrm{id}_{V^*} && \\ &(t\phi + s\psi)^* = t\phi^* + s\psi^* && \forall \phi, \psi \in \mathrm{Hom}(V, W), t, s \in \mathbb{R} \\ &(\phi \circ \psi)^* = \psi^* \circ \phi^* && \forall \psi \in \mathrm{Hom}(V, W), \phi \in \mathrm{Hom}(W, U) \end{aligned} \tag{1.9}$$

Remark 1.1.9: (Covariant and contravariant quantities, Einstein summation convention)

1. Under a linear map $\phi \in \mathrm{End}(V)$, quantities with *lower indices* such as the the basis vectors $e_i \in B$ and the coefficients with respect to the dual basis B^* transform as in (1.3), (1.8)
$$e_i \mapsto \sum_{j=1}^n {\phi_i}^j e_j \qquad \alpha_i \mapsto \sum_{j=1}^n {\phi_i}^j \alpha_j.$$
They are said to *transform contravariantly* or to be *contravariant quantities.* Quantities with *upper indices* such as the coefficients with respect to a basis B or the basis vectors $e^i \in B^*$ of the dual basis, transform as in (1.4), (1.8).
$$e^i \mapsto \sum_{j=1}^n {\phi_j}^i e^j \qquad x^i \mapsto \sum_{j=1}^n {\phi_j}^i x^j.$$
They are said to *transform covariantly* or to be *covariant quantities.*

2. In physics the sum symbols are often omitted and quantities are expressed in **Einstein summation convention**: all indices that occur twice in an expression, once as an upper and once as a lower index, are summed over. Equations (1.1), (1.3), (1.6) and (1.8) then take the form
$$x = x^i e_i \qquad \phi(e_i) = {\phi_i}^j e_j \qquad \alpha = \alpha_i e^i \qquad \phi^*(e^i) = {\phi_j}^i e^j. \tag{1.10}$$

Exercise 2: (Harmonic oscillator)

1. We consider the classical harmonic oscillator with equations of motion

$$\ddot{x} + \omega^2 x = 0 \qquad x : \mathbb{R} \to \mathbb{R},\ t \mapsto x(t). \tag{1.11}$$

Show or recall that the space of solutions of this equation is a two-dimensional vector space V over $\mathbb{R}$ with basis

$$e_1(t) = \sin(\omega t) \qquad e_2(t) = \cos(\omega t). \tag{1.12}$$

Show that the maps $\delta, -\dot{\delta} : V \to \mathbb{R}$

$$\delta : f \mapsto f(0) \qquad -\dot{\delta} : f \mapsto \dot{f}(0). \tag{1.13}$$

form a basis of the dual vector space V^* and that this is a dual basis of $B = \{e_1, e_2\}$. We will see later that they are the restriction of the delta distribution and its derivative to the space of solutions of (1.11).

2. Show that the maps

$$\Phi_\theta : f \to f_\theta \qquad f_\theta(t) = f(t + \theta) \tag{1.14}$$

are elements of $\mathrm{Aut}(V)$. Calculate the representing matrix of $\Phi_{\pi/(2\omega)}$ and $\Phi_{\pi/(4\omega)}$ with respect to the basis (1.12). Express the basis dual to $B = \{\Phi_{\pi/4\omega}(e_1), \Phi_{\pi/4\omega}(e_2)\}$ in terms of $e^1 = -\dot{\delta}$ and $e^2 = \delta$. Describe the dual of the maps Φ_θ with respect to the dual basis (1.13).

Hint:

$$\cos(\alpha + \beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta \qquad \sin(\alpha + \beta) = \cos\alpha\sin\beta + \cos\beta\sin\alpha. \tag{1.15}$$

Exercise 3: (Spin 1/2 particle)

We consider the quantum mechanical spin 1/2 particle, which is described by a two-dimensional vector space over $\mathbb{C}$ with basis

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{1.16}$$

We consider the set of hermitian, traceless matrices, i.e. matrices $A \in M(2, \mathbb{C})$ with $A^\dagger = \bar{A}^T = A$, $\mathrm{Tr}\,(A) = 0$. Show that these matrices form a vector space W and that a basis of W is given by

$$J_x = \tfrac{1}{2}\begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} \qquad J_y = \tfrac{1}{2}\begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \qquad J_z = \tfrac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.17}$$

The *matrix element* $M_{\boldsymbol{x},\boldsymbol{y}}(A)$ of a matrix $A \in M(2, \mathbb{C})$ with respect to vectors $\boldsymbol{x} = \alpha e_1 + \beta e_2$, $\boldsymbol{y} = \gamma e_1 + \delta e_2$, $\alpha, \beta, \gamma, \delta \in \mathbb{C}$ is defined as

$$M_{\boldsymbol{x},\boldsymbol{y}}(A) = \boldsymbol{y}^T \cdot A \cdot \boldsymbol{x} = \begin{pmatrix} \bar{\gamma} & \bar{\delta} \end{pmatrix} \cdot A \cdot \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \tag{1.18}$$

The expectation value $E_{\boldsymbol{x}}(A)$ of a matrix $A \in M(2, \mathbb{C})$ with respect to the vector $\boldsymbol{x} = \alpha e_1 + \beta e_2 \in \mathbb{C}^2$, $\alpha, \beta \in \mathbb{C}$, is the matrix element $M_{\boldsymbol{x},\boldsymbol{x}}$

$$E_{\boldsymbol{x}}(A) = M_{\boldsymbol{x},\boldsymbol{x}}(A) = \boldsymbol{x}^T \cdot A \cdot \boldsymbol{x} = \begin{pmatrix} \bar{\alpha} & \bar{\beta} \end{pmatrix} \cdot A \cdot \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \tag{1.19}$$

Show that for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^2$, the expectation value $E_{\boldsymbol{x}} : A \mapsto E_{\boldsymbol{x}}(A)$ and the real and imaginary part of the matrix elements $M_{\boldsymbol{x},\boldsymbol{y}} : A \mapsto E_{\boldsymbol{x},\boldsymbol{y}}(A)$ are elements of W^*. Determine vectors $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ such that the expectation values $E_{\boldsymbol{x}}$, $E_{\boldsymbol{y}}$, $E_{\boldsymbol{z}}$ form a dual basis for $\{J_x, J_y, J_z\}$.

Hint: Show first that for $\boldsymbol{x} = (\alpha, \beta)^T$, $\alpha, \beta \in \mathbb{C}$, we have

$$E_{\boldsymbol{x}}(J_x) = -\mathrm{Im}(\bar{\alpha}\beta) \qquad E_{\boldsymbol{x}}(J_y) = -\mathrm{Re}(\bar{\alpha}\beta) \qquad E_{\boldsymbol{x}}(J_z) = \tfrac{1}{2}(|\alpha|^2 - |\beta|^2).$$

We will now investigate constructions that allow us to build a new vector space out of two given ones. The first is the direct sum of two vector spaces. It allows one to combine families of linear maps $\phi_i : V_i \to W_i$, $i = 1, ..., n$ into a single linear map $\phi = \phi_1 \oplus ... \oplus \phi_i : V_1 \oplus ... \oplus V_n \to W_1 \oplus ... \oplus W_n$.

Definition 1.1.10: (Direct sum of vector spaces)

The *direct sum* $V \oplus W$ of two vector spaces V and W is the set of tuples $\{(\boldsymbol{x}, \boldsymbol{y}) \mid \boldsymbol{x} \in V, \boldsymbol{y} \in W\}$, equipped with the vector space addition and scalar multiplication

$$(\boldsymbol{x}_1, \boldsymbol{y}_1) + (\boldsymbol{x}_2, \boldsymbol{y}_2) = (\boldsymbol{x}_1 + \boldsymbol{x}_2, \boldsymbol{y}_1 + \boldsymbol{y}_2) \qquad t \cdot (\boldsymbol{x}, \boldsymbol{y}) = (t\boldsymbol{x}, t\boldsymbol{y}) \tag{1.20}$$

for all $\boldsymbol{x}, \boldsymbol{x}_1, \boldsymbol{x}_2 \in V$, $\boldsymbol{y}, \boldsymbol{y}_1, \boldsymbol{y}_2 \in W$ and $t \in k$. If $B_V = \{e_1, .., e_n\}$ is a basis of V and $B_W = \{f_1, ..., f_m\}$ a basis of W, a basis of $V \oplus W$ is given by

$$B_{V \oplus W} = \{(e_i, 0) \mid e_i \in B_V\} \cup \{(0, f_j) \mid f_j \in B_W\}. \tag{1.21}$$

The maps $\imath_V : V \to V \oplus W$, $\imath_W : W \to V \oplus W$

$$\imath_V(\boldsymbol{v}) = (\boldsymbol{v}, 0) \qquad \imath_W(\boldsymbol{w}) = (0, \boldsymbol{w}) \qquad \forall \boldsymbol{v} \in V, \boldsymbol{w} \in W$$

are called the *canonical injections* of V and W into $V \oplus W$. In the following we often omit the tuple notation and write $\boldsymbol{x} + \boldsymbol{y}$ for $\imath_V(\boldsymbol{v}) + \imath_W(\boldsymbol{w}) = (\boldsymbol{x}, \boldsymbol{y})$.

The direct sum of two linear maps $\phi \in \mathrm{Hom}(V, V')$, $\psi \in \mathrm{Hom}(W, W')$ is the linear map $\phi \oplus \psi \in \mathrm{Hom}(V \oplus W, V' \oplus W')$ defined by

$$\phi \oplus \psi(\boldsymbol{v}, \boldsymbol{w}) = (\phi(\boldsymbol{v}), \psi(\boldsymbol{w})) \qquad \forall \boldsymbol{v} \in V, \boldsymbol{w} \in W.$$

An alternative but equivalent definition of the direct sum is via a *universality property*.

Definition 1.1.11: The direct sum of two vector spaces V, W over k is a vector space, denoted $V \oplus W$ together with two linear maps $\imath_V : V \to V \oplus W$, $i_W : W \to V \oplus W$ such that for any pair of linear maps $\beta : V \to X$, $\gamma : W \to X$ there exists a unique bilinear map $\phi_{\beta,\gamma} : V \oplus W \to X$ such that $\phi_{\beta,\gamma} \circ \imath_V = \beta$, $\phi_{\beta,\gamma} \circ i_W = \gamma$.

1.1.3 (Multi)linear forms

Definition 1.1.12: (Multilinear forms)

1. A *n-linear form* on a vector space V is a map $\alpha : V \times ... \times V \to k$ that is linear in all arguments

$$\alpha(.., t\boldsymbol{y} + s\boldsymbol{z}, ...) = t\,\alpha(..., \boldsymbol{y}, ...) + s\,\alpha(..., \boldsymbol{z}, ...) \qquad \forall t, s \in k, \boldsymbol{x}_i, \boldsymbol{y}, \boldsymbol{z} \in V.$$

2. For $n = 2$, the form is called bilinear. The *representing matrix* of a bilinear form $\alpha : V \times V \to k$ with respect to a basis $B = \{e_1, ..., e_n\}$ of V is the matrix $A_\alpha = (\alpha_{ij})_{i,j=1,...,n}$ with $\alpha_{ij} = \alpha(e_i, e_j)$. A bilinear form $\alpha : V \times V \to k$ is called

- *non-degenerate* if $\alpha(\boldsymbol{x}, \boldsymbol{y}) = 0 \ \forall \boldsymbol{y} \in V$ implies $\boldsymbol{x} = 0 \Leftrightarrow \det(A_\alpha) \neq 0$.
- *symmetric* if $\alpha(\boldsymbol{x}, \boldsymbol{y}) = \alpha(\boldsymbol{y}, \boldsymbol{x}) \ \forall \boldsymbol{x}, \boldsymbol{y} \in V \Leftrightarrow A_\alpha = A_\alpha^T$.
- *anti-symmetric* if $\alpha(\boldsymbol{x}, \boldsymbol{y}) = -\alpha(\boldsymbol{y}, \boldsymbol{x}) \ \forall \boldsymbol{x}, \boldsymbol{y} \in V \Leftrightarrow A_\alpha = -A_\alpha^T$
- *positive definite* if $\alpha(\boldsymbol{x}, \boldsymbol{x}) > 0 \ \forall \boldsymbol{x} \in V \setminus \{0\} \Leftrightarrow \sum_{i,j=1}^n x^i x^j \alpha_{ij} > 0 \ \forall \boldsymbol{x}, \in V \setminus \{0\}$.
- *positive semi-definite* if $\alpha(\boldsymbol{x}, \boldsymbol{x}) \geq 0 \ \forall \boldsymbol{x} \in V \setminus \{0\} \Leftrightarrow \sum_{i,j=1}^n x^i x^j \alpha_{ij} \geq 0 \ \forall \boldsymbol{x}, \in V \setminus \{0\}$.
- *scalar product over* $\mathbb{R}$ if $k = \mathbb{R}$ and α is non-degenerate, positive definite and symmetric.

Exercise 4: Prove that the identities for the representing matrix in the second part of Def. 1.1.12 are equivalent to the corresponding conditions on the bilinear form.

There is a related concept which for vector spaces over $\mathbb{C}$ which plays an important role in quantum mechanics.

Definition 1.1.13: (Hermitian forms)

Let V be a vector space over $\mathbb{C}$. A map $\alpha : V \times V \to \mathbb{C}$ is called *hermitian* if it is linear in the second argument, anti-linear in the first argument and switching its arguments results in complex conjugation:

$$\begin{aligned} &\alpha(\boldsymbol{x}, \boldsymbol{y}) = \overline{\alpha(\boldsymbol{y}, \boldsymbol{x})} \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in V \\ &\alpha(t\boldsymbol{y} + s\boldsymbol{z}, \boldsymbol{x}) = \bar{t}\,\alpha(\boldsymbol{y}, \boldsymbol{x}) + \bar{s}\,\alpha(\boldsymbol{z}, \boldsymbol{x}) \\ &\alpha(\boldsymbol{x}, t\boldsymbol{y} + s\boldsymbol{z}) = t\,\alpha(\boldsymbol{x}, \boldsymbol{y}) + s\,\alpha(\boldsymbol{x}, \boldsymbol{z}) \qquad \forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V, t, s \in \mathbb{C}. \end{aligned}$$

If, additionally, α is positive definite: $\alpha(\boldsymbol{x}, \boldsymbol{x}) > 0 \ \forall \boldsymbol{x} \in V \setminus \{0\}$, it is called a *hermitian product* or *scalar product over* $\mathbb{C}$.

Example 1.1.14: (Non-degenerate 2-forms)

1. The Euclidean or standard scalar product g_E on $\mathbb{R}^n$ given by $g_E(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^n x^i y^i$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, is a symmetric, non-degenerate and positive definite bilinear form on $\mathbb{R}^n$.
2. The Minkowski metric g_M on $\mathbb{R}^n$ which is given by $g_M(\boldsymbol{x}, \boldsymbol{y}) = -x^0 y^0 + \sum_{i=1}^{n-1} x^i y^i$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ is a symmetric and non-degenerate form on $\mathbb{R}^n$, but not positive definite.
3. The standard hermitian product on $\mathbb{C}^n$ given by $\alpha(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^n \bar{x}^i y^i$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$ is a hermitian form on $\mathbb{C}^n$.
4. The components of the cross or wedge product of vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^3$ define an antisymmetric bilinear form $\alpha : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ on $\mathbb{R}^3$

$$\alpha_i(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x} \wedge \boldsymbol{y})^i \qquad \text{where} \quad (\boldsymbol{x} \wedge \boldsymbol{y} = (x^2 y^3 - y^2 x^3, x^3 y^1 - x^1 y^3, x^1 y^2 - x^2 y^1).$$

Example 1.1.15: (Hermitian form on the space of matrices)

We consider the space of matrices $M(n, \mathbb{C})$ as an n^2-dimensional vector space over $\mathbb{C}$. The *trace* of a matrix $A = (a_{ij})_{i,j=1,\ldots,n} \in M(n, \mathbb{C})$ is given by $\mathrm{Tr}\,(A) = \sum_{i=1}^n a_{ii}$. The hermitian conjugate $A^\dagger$ of A is the matrix $A^\dagger = \bar{A}^T$, where $\bar{}$ denotes complex conjugation. We define a map $\alpha : M(n, \mathbb{C}) \times M(n, \mathbb{C}) \to \mathbb{C}$ by setting

$$\alpha(A, B) = \mathrm{Tr}\left(A^\dagger \cdot B\right) \qquad \forall A, B \in M(2, \mathbb{C})$$

Then α defines a hermitian product on the space of matrices $M(2, \mathbb{C})$.

Proof: Exercise

Exercise 5: Show that the map $\alpha : M(2,\mathbb{C}) \times (2,\mathbb{C}) \to \mathbb{C}$ defined by

$$\alpha(A,B) = \mathrm{Tr}\left(A^\dagger B\right) \qquad \forall A, B \in M(2,\mathbb{C}).$$

is anti-linear in the first and linear in the second argument as well as positive definite and therefore defines a hermitian product on $M(2,\mathbb{C})$. Prove the *Cauchy Schwartz inequality*

$$|\alpha(A,B)| \leq \sqrt{\alpha(A,A)}\sqrt{\alpha(B,B)}.$$

Lemma 1.1.16: A non-degenerate bilinear form α induces two isomorphisms $\Phi^1, \Phi^2 : V \to V^*$ between V and V^*

$$\Phi^1(\boldsymbol{x}) = \alpha(\boldsymbol{x}, \cdot) : V \to k \qquad \Phi^2(\boldsymbol{x}) = \alpha(\cdot, \boldsymbol{x}) : V \to k. \tag{1.22}$$

The maps coincide (up to a minus sign) if and only if α is symmetric (antisymmetric).

Exercise 6: (Orthonormal basis)

1. Show that for any symmetric or hermitian bilinear form α there exists a basis $B = \{e_1, ..., e_n\}$ of V such that $\alpha(e_i, e_j) = \epsilon_i \delta_{ij}$ with $\epsilon_i \in \{0, 1, -1\}$. Hint: Express α in terms of coefficients with respect to a general basis $\tilde{B} = \{\tilde{e}_1, ..., \tilde{e}_n\}$ of V. What does the fact that α is symmetric or hermitian imply for the matrix formed by its coefficients? Use results from linear algebra to show that the matrix can be diagonalised and that this corresponds to a basis transformation.

2. Show that if α is also non-degenerate, we have $\epsilon_i \in \{\pm 1\}$. The associated basis is called *orthonormal basis for* α, and the coefficients ϵ_i are called eigenvalues. The *signature* of α is the number of *negative eigenvalues.*

Proof: Exercise.

Exercise 7: (Orthogonal complement)

The *orthogonal complement* $U^{\perp_\alpha}$ of a subspace $U \subset V$ with respect to a non-degenerate bilinear or hermitian form α on V is the set

$$U^{\perp_\alpha} = \{\boldsymbol{y} \in V : \ \alpha(\boldsymbol{y}, \boldsymbol{x}) = 0 \ \forall \boldsymbol{x} \in U\}. \tag{1.23}$$

For $U = \mathrm{Span}(\boldsymbol{x}) = \{t\boldsymbol{x} \mid t \in \mathbb{R}\}$ we write $U^{\perp_\alpha} = \boldsymbol{x}^{\perp_\alpha}$.

1. Show that $U^{\perp_\alpha}$ is a linear subspace of V.
2. Show that for linear subspaces $W \subset U \subset V$, we have $V^{\perp_\alpha} \subset W^{\perp_\alpha}$.
3. Show that for a scalar product or hermitian product α, $U \cap U^{\perp_\alpha} = \{0\}$ and $V = U \oplus U^{\perp_\alpha}$.
4. Show that for a scalar product α and any vector $\boldsymbol{x} \in V \setminus \{0\}$, we have $\dim(\boldsymbol{x}^{\perp_\alpha}) = \dim(V) - 1$ and $\boldsymbol{x}$ is not contained in $\boldsymbol{x}^{\perp_\alpha}$.
5. Give an example in which α is symmetric and non-degenerate and $\boldsymbol{x} \in \boldsymbol{x}^{\perp_\alpha}$, $\boldsymbol{x} \neq 0$. Hint: think of lightlike vectors.
6. Show: For two subspaces $U, W \subset V$, $U \cap W = \{0\}$, we have $(U \oplus W)^{\perp_\alpha} = U^{\perp_\alpha} \cap W^{\perp_\alpha}$.

1.2 Tensors

As we have seen in the previous section, the direct sum of vector spaces allowed us to combine families of linear maps between different vector spaces into a single linear map between their direct sum. The notion of a dual vector space allowed us to understand the vector space structure of linear forms. We will now consider *tensor products* of vector spaces. Just as the direct sum of vector spaces allows one to combine families of linear maps into a single linear map, tensor products allow one to view multilinear maps between certain vector spaces as linear maps between their tensor products.

Tensor products play an important role in physics as they arise whenever several physical systems are coupled to a single system. Examples are multi-particle systems, coupled harmonic oscillators and, finally, the Fock space arising in second quantisation. Tensors and tensor fields also play an important role in general relativity.

Definition 1.2.1: (Tensor product of vector spaces)

The *tensor product* of two vector spaces V, W over k is a vector space $V \otimes W$ over k together with a linear map $\kappa : V \times W \to V \otimes W$, $(v, w) \mapsto v \otimes w$ that has the *universality property*: For any bilinear map $\beta : V \times W \to X$ there exists exactly one linear map $\phi_\beta : V \otimes W \to X$ such that $\beta = \phi_\beta \circ \kappa$. The elements of the vector space $V \otimes W$ are called tensors.

Lemma 1.2.2: The universality property defines the tensor product consistently and up to a unique isomorphism.

Proof:

Existence:

Any (multi)linear map between vector spaces is given uniquely by its values on a basis. We select a vector space of dimension $\dim(V) \cdot \dim(W)$, which we denote $V \otimes W$, and a basis denoted $B = \{e_i \otimes f_j \mid i = 1, ..., \dim(V), j = 1, ..., \dim(W)\}$. Furthermore, we choose a basis $B_V = \{e_1, ..., e_n\}$ of V and a basis $B_W = \{f_1, ..., f_m\}$ of W and define a bilinear map $\kappa : V \times W \to V \otimes W$ by its values on a basis. We set $\kappa(e_i, f_j) = e_i \otimes f_j$. For general vectors $\boldsymbol{x} = \sum_{i=1}^n x^i e_i$, $\boldsymbol{z} = \sum_{j=1}^m y^j f_j$, we then have

$$\kappa(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^n \sum_{j=1}^m x^i y^j e_i \otimes f_j.$$

Uniqueness:

Suppose there were two such maps $\kappa : V \times W \to V \otimes W$, $\tilde{\kappa} : V \times W \to V \tilde{\otimes} W$, both having the universality property. Then for any linear map $\alpha : V \times W \to X$, we have two maps $\phi_\alpha : V \otimes W \to X$, $\tilde{\phi}_\alpha : V \tilde{\otimes} W \to X$ such that $\alpha = \phi_\alpha \circ \kappa = \tilde{\phi}_\alpha \circ \tilde{\kappa}'$. If we take $\alpha = \kappa$, we obtain a map $\tilde{\phi}_\kappa : V \tilde{\otimes} W \to V \otimes W$ satisfying $\tilde{\phi}_\kappa \circ \tilde{\kappa} = \kappa$. If we set $\alpha = \tilde{\kappa}$, we obtain a map $\phi_{\tilde{\kappa}} : V \otimes W \to V \tilde{\otimes} W$ satisfying $\phi_{\tilde{\kappa}} \circ \kappa = \tilde{\kappa}$. This implies for the composition $\tilde{\phi}_\kappa \circ \phi_{\tilde{\kappa}} \circ \kappa = \kappa : V \times W \to V \otimes W$ and $\phi_{\tilde{\kappa}} \circ \tilde{\phi}_\kappa \circ \tilde{\kappa} = \tilde{\kappa} : V \times W \to V \tilde{\otimes} W$. Hence, we have $\tilde{\phi}_\kappa \circ \phi_{\tilde{\kappa}} = \mathrm{id}_{V \otimes W}$, $\phi_{\tilde{\kappa}} \circ \tilde{\phi}_\kappa = \mathrm{id}_{V \tilde{\otimes} W}$, which implies that $\tilde{\phi}_\kappa$ and ϕ_κ are isomorphisms. □

Definition 1.2.1 gives a precise definition of the tensor product. However, it is not very useful in concrete calculations. For those, it is advantageous to use the following identities instead.

Lemma 1.2.3: (Properties of the tensor product)

1. The tensor product is *bilinear*. With $\boldsymbol{v} \otimes \boldsymbol{w} = \kappa(\boldsymbol{v}, \boldsymbol{w})$, we have

$$(\boldsymbol{v}_1 + \boldsymbol{v}_2) \otimes \boldsymbol{w} = \boldsymbol{v}_1 \otimes \boldsymbol{w} + \boldsymbol{v}_2 \otimes \boldsymbol{w} \qquad \forall \boldsymbol{v}_1, \boldsymbol{v}_2 \in V, \boldsymbol{w} \in W \tag{1.24}$$
$$\boldsymbol{v} \otimes (\boldsymbol{w}_1 + \boldsymbol{w}_2) = \boldsymbol{v} \otimes \boldsymbol{w}_1 + \boldsymbol{v} \otimes \boldsymbol{w}_2 \qquad \forall \boldsymbol{v} \in V, \boldsymbol{w}_1, \boldsymbol{w}_2 \in W \tag{1.25}$$
$$t(\boldsymbol{v} \otimes \boldsymbol{w}) = (t\boldsymbol{v}) \otimes \boldsymbol{w} = \boldsymbol{v} \otimes (t\boldsymbol{w}) \qquad \forall \boldsymbol{v} \in V, \boldsymbol{w} \in W, t \in k. \tag{1.26}$$

2. There are *canonical* isomorphisms

$$\begin{aligned}
&R: \; V \otimes W \to W \otimes V, \; \boldsymbol{v} \otimes \boldsymbol{w} \mapsto \boldsymbol{w} \otimes \boldsymbol{v} \\
&A: \; U \otimes (V \otimes W) \to (U \otimes V) \otimes W, \; \boldsymbol{u} \otimes (\boldsymbol{v} \otimes \boldsymbol{w}) \to (\boldsymbol{u} \otimes \boldsymbol{v}) \otimes \boldsymbol{w} \\
&K_L: \; k \otimes V \to V, t \otimes \boldsymbol{v} \mapsto t\boldsymbol{v} \\
&K_R: \; V \otimes k \to V, \boldsymbol{v} \otimes t \mapsto t\boldsymbol{v}.
\end{aligned}$$

3. A basis of $V \otimes W$ is given by the set $B_{V \otimes W} = \{e_i \otimes f_j \mid i = 1, ..., n, j = 1, ..., m\}$, where $B = \{e_1, ..., e_n\}$ is a basis of V and $B' = \{f_1, ..., f_m\}$ a basis of W. This basis is called the *tensor product basis*. With respect to this basis, a general element $\boldsymbol{x} \in V \otimes W$ can be expressed as a sum

$$\boldsymbol{x} = \sum_{i=1}^{n} \sum_{j=1}^{m} x^{ij} e_i \otimes f_j. \tag{1.27}$$

Remark 1.2.4: Note that although elements of the form $\boldsymbol{v} \otimes \boldsymbol{w}$ with $\boldsymbol{v} \in V, \boldsymbol{w} \in W$ span the tensor product $V \otimes W$, *not every* element of $V \otimes W$ is of this form.

Example 1.2.5:

1. $\mathbb{R}^n \otimes \mathbb{R}^m \cong \mathbb{R}^{m \cdot n}$. An isomorphism is given by $e_i \otimes e_j \mapsto e_{(m-1)i+j}$, where $B = \{e_1, ..., e_k\}$ is the canonical basis of $\mathbb{R}^k$.

2. We can view $\mathbb{C}$ as a two-dimensional vector space over $\mathbb{R}$ with basis $B = \{1, i\}$. Then, the tensor product $V_{\mathbb{C}} := V \otimes \mathbb{C}$ for a vector space V over $\mathbb{R}$ is called *complexification of V*. With the definition $a \cdot (v \otimes z) := v \otimes (az)$ for $a, z \in \mathbb{C}$, $v \in V$, $V_{\mathbb{C}}$ becomes a vector space over $\mathbb{C}$.

3. Let X, Y be finite sets and $\mathcal{F}(X)$, $\mathcal{F}(Y)$ the set of functions on X and Y with values in $k = \mathbb{R}$ or $k = \mathbb{C}$. Clearly, $\mathcal{F}(X)$ and $\mathcal{F}(Y)$ have the structure of vector spaces with respect to the pointwise addition of functions and multiplication by k. The tensor product $\mathcal{F}(X) \otimes \mathcal{F}(Y)$ is canonically isomorphic to the space of functions on $X \times Y$. The isomorphism $\phi : \mathcal{F}(X) \otimes \mathcal{F}(Y) \to \mathcal{F}(X \times Y)$ is given by

$$\phi(f \otimes g)(x, y) = f(x)g(y) \qquad \forall x \in X, y \in Y, f \in \mathcal{F}(X), g \in \mathcal{F}(Y).$$

We consider the $B_X = \{\delta_x \mid x \in X\}$ of $\mathcal{F}(X)$, $B_Y = \{\delta_y \mid y \in Y\}$ of $\mathcal{F}(Y)$ where δ_x, δ_y are the Kronecker delta functions

$$\delta_x(z) = \begin{cases} 1 & z = x \\ 0 & z \neq x \end{cases}.$$

Clearly, we have for any $f \in \mathcal{F}(X)$, $f = \sum_{x\in X} f(x)\delta_x$ and similarly for Y. In terms of these bases, the isomorphism ϕ takes the form

$$\phi(\delta_x \otimes \delta_y) = \delta_{(x,y)} \qquad \delta_{(x,y)}(v,w) = \begin{cases} 1 & x = v \text{ and } y = w \\ 0 & \text{otherwise.} \end{cases}$$

Its inverse is given by

$$\phi^{-1} : \mathcal{F}(X \times Y) \to \mathcal{F}(X) \otimes \mathcal{F}(Y),\ \delta_{(x,y)} \mapsto \delta_x \otimes \delta_y.$$

As the linear maps between vector spaces also form a vector space, we can consider the tensor product of vector spaces of linear maps. We obtain the following lemma.

Lemma 1.2.6: (Tensor product of linear maps, functoriality)

For vector spaces U, V, W, Z over k the tensor product $\mathrm{Hom}(U,V) \otimes \mathrm{Hom}(W,Z)$ is canonically isomorphic to $\mathrm{Hom}(U \otimes W, V \otimes Z)$. The isomorphism $\tau : \mathrm{Hom}(U,V) \otimes \mathrm{Hom}(W,Z) \to \mathrm{Hom}(U \otimes W, V \otimes Z)$ is given by

$$\tau(\psi \otimes \phi)(\boldsymbol{u} \otimes \boldsymbol{v}) = \psi(\boldsymbol{u}) \otimes \phi(\boldsymbol{v}) \qquad \forall \psi \in \mathrm{Hom}(U,V), \phi \in \mathrm{Hom}(W,Z), \boldsymbol{u} \in U, \boldsymbol{v} \in V.$$

In terms of the maps $\kappa_{V\otimes Z} : V \times Z \to V \otimes Z$, $\kappa_{U\otimes W} : U \times W \to U \otimes W$ introduced in Def. 1.2.1, we have

$$\kappa_{V\otimes Z} \circ (\psi, \phi) = (\psi \otimes \phi) \circ \kappa_{U\otimes W} \qquad \forall \psi \in \mathrm{Hom}(U,V), \phi \in \mathrm{Hom}(W,Z). \tag{1.28}$$

We say the tensor product is *functorial.*

Lemma 1.2.7: (Transformation behaviour of coefficients)

If we characterise $\psi \in \mathrm{Hom}(U,V)$, $\phi \in \mathrm{Hom}(W,Z)$ by their matrix coefficients with respect to bases $B_U = \{e_1, ..., e_n\}$, $B_V = \{f_1, ..., f_m\}$,$B_W = \{g_1, ..., g_p\}$, $B_Z = \{h_1, ..., h_s\}$

$$\psi(e_i) = \sum_{j=1}^{m} {\psi_i}^j f_j \qquad \phi(g_k) = \sum_{l=1}^{s} {\phi_k}^l h_l, \tag{1.29}$$

the matrix coefficients of the tensor product $\psi \otimes \phi$ with respect to the bases $B_{U\otimes W} = \{e_i \otimes g_k \mid i = 1, ..., n, k = 1, ..., p\}$ and $B_{V\otimes Z} = \{f_j \otimes h_l \mid j = 1, ..., m, l = 1, ..., s\}$ are given by

$$(\psi \otimes \phi)(e_i \otimes g_k) = \sum_{j=1}^{m} \sum_{l=1}^{s} {\psi_i}^j {\phi_k}^l f_j \otimes h_l \tag{1.30}$$

Alternatively, we can characterise the linear map $\psi \otimes \phi$ by its action on the coefficients. For

$$\boldsymbol{x} = \sum_{i=1}^{n} \sum_{k=1}^{p} x^{ik} e_i \otimes g_k \qquad (\phi \otimes \psi)(\boldsymbol{x}) = \sum_{j=1}^{m} \sum_{l=1}^{s} x'^{jl} f_j \otimes h_l, \tag{1.31}$$

we have

$$x'^{jl} = \sum_{i=1}^{n} \sum_{k=1}^{p} {\psi_i}^j {\phi_k}^l x^{ik}.$$

Exercise 8: (Two spin 1/2 particles)

We consider a system of two spin 1/2 particles as in Example 3. The vector space associated with the combined system is the tensor product $\mathbb{C}^2 \otimes \mathbb{C}^2 \cong \mathbb{C}^4$. A basis of $\mathbb{C}^2 \otimes \mathbb{C}^2$, the *tensor product basis*, is given by

$$B = \{e_1 \otimes e_1, e_1 \otimes e_2, e_2 \otimes e_1, e_2 \otimes e_2\} \tag{1.32}$$

with e_1, e_2 as in (1.16). We consider the *total spin* or *total angular momentum*

$$J^2_{tot} := \sum_{a=1}^{3} (1 \otimes J_a + J_a \otimes 1)(1 \otimes J_a + J_a \otimes 1) = \sum_{a=1}^{3} J_a^2 \otimes 1 + 1 \otimes J_a^2 + 2J_a \otimes J_a \in \mathrm{End}(\mathbb{C}^2 \otimes \mathbb{C}^2)$$

and the *total spin* or *total angular momentum in the z-direction*

$$J_z^{tot} = 1 \otimes J_z + J_z \otimes 1 \in \mathrm{End}(\mathbb{C}^2 \otimes \mathbb{C}^2).$$

Determine the matrix coefficients of J^2_{tot} and J_z^{tot} with respect to the tensor product basis. Show that another basis of $\mathbb{C}^2 \otimes \mathbb{C}^2$ is given by

$$B' = \{e_1 \otimes e_1, e_2 \otimes e_2, \tfrac{1}{\sqrt{2}}(e_1 \otimes e_2 - e_2 \otimes e_1), \tfrac{1}{\sqrt{2}}(e_1 \otimes e_2 + e_2 \otimes e_1)\}. \tag{1.33}$$

Determine the matrix coefficients of J^2_{tot} and J_z^{tot} with respect to this basis.

After defining the tensor product, we are now ready to investigate the relation between tensors and multilinear forms. We start by considering the relation between tensor products and dual vector spaces.

Theorem 1.2.8: (Tensor products and duals)

1. For vector spaces V, W over k, the tensor product of their duals is canonically isomorphic to the dual of their tensor product: $V^* \otimes W^* \cong (V \otimes W)^*$. The isomorphism is given by

$$\tau : \alpha \otimes \beta \mapsto \tau(\alpha \otimes \beta) \qquad \tau(\alpha \otimes \beta)(\boldsymbol{v} \otimes \boldsymbol{w}) = \alpha(\boldsymbol{v})\beta(\boldsymbol{w}) \quad \forall \boldsymbol{v} \in V, \boldsymbol{w} \in W. \tag{1.34}$$

2. For vector spaces V, W over k, we have $V^* \otimes W \cong \mathrm{Hom}(V, W)$. The isomorphism is

$$\mu : \alpha \otimes \boldsymbol{w} \mapsto \mu(\alpha \otimes \boldsymbol{w}) \qquad \mu(\alpha \otimes \boldsymbol{w})(\boldsymbol{v}) = \alpha(\boldsymbol{v})\boldsymbol{w} \quad \forall \boldsymbol{v} \in V, \boldsymbol{w} \in W. \tag{1.35}$$

Corollary 1.2.9: The n-fold tensor product $V^* \otimes ... \otimes V^*$ can be identified with the space of n-forms $\alpha : V \times ... \times V \to k$. The identification is given by

$$\mu(\alpha_1 \otimes ... \otimes \alpha_n)(\boldsymbol{v}_1, ..., \boldsymbol{v}_n) = \alpha_1(\boldsymbol{v}_1) \cdots \alpha_n(\boldsymbol{v}_n) \qquad \forall \boldsymbol{v}_1, ..., \boldsymbol{v}_n \in V, \alpha_1, ..., \alpha_n \in V^*. \tag{1.36}$$

In the following, we will often omit the isomorphisms μ, τ in theorem 1.2.8 and write $\alpha \otimes \beta(\boldsymbol{v} \otimes \boldsymbol{w}) = \alpha(\boldsymbol{v})\beta(\boldsymbol{w})$ for (1.34) and $(\alpha \otimes \boldsymbol{w})(\boldsymbol{v}) = \alpha(\boldsymbol{v})\boldsymbol{w}$ for (1.35).

These relations between the tensor product of vector spaces, their duals and multilinear forms extend to multiple tensor products of vector spaces and their duals. To define this, we introduce the tensor algebra.

Definition 1.2.10: (Tensor algebra)

1. For a vector space V and $r, s \geq 0$, we define the *tensor space of type (r,s)*

$$\bigotimes_{r,s} V := \underbrace{V \otimes ... \otimes V}_{r\times} \otimes \underbrace{V^* \otimes ... \otimes V^*}_{s\times} \qquad \bigotimes_{0,0} V = k. \tag{1.37}$$

elements of $\bigotimes_{r,s} V$ are called *homogeneous tensors of type (r,s)*.

2. The *tensor algebra* $T(V)$ is the direct sum

$$T(V) = \bigoplus_{r,s\geq 0} \bigotimes_{r,s} V. \tag{1.38}$$

We define a multiplication $\otimes : T(V) \otimes T(V) \to T(V)$ via its action on the homogeneous tensors and extend it bilinearly to $T(V)$. For $\boldsymbol{u}_1 \otimes ... \otimes \boldsymbol{u}_r \otimes \alpha_1 \otimes ... \otimes \alpha_s \in \bigotimes_{r,s} V$ and $\boldsymbol{x}_1 \otimes ... \otimes \boldsymbol{x}_p \otimes \beta_1 \otimes ... \otimes \beta_q \in \bigotimes_{p,q} V$ we set

$$(\boldsymbol{u}_1 \otimes ... \otimes \boldsymbol{u}_r \otimes \alpha_1 \otimes ... \otimes \alpha_s) \otimes (\boldsymbol{x}_1 \otimes ... \otimes \boldsymbol{x}_p \otimes \beta_1 \otimes ... \otimes \beta_q) \tag{1.39}$$

$$:= \boldsymbol{u}_1 \otimes ... \otimes \boldsymbol{u}_r \otimes \boldsymbol{x}_1 \otimes ... \otimes \boldsymbol{x}_p \otimes \alpha_1 \otimes ... \otimes \alpha_s \otimes \beta_1 \otimes ... \otimes \beta_p \in \bigotimes_{r+p,s+q} V. \tag{1.40}$$

Remark 1.2.11: With the multiplication defined in (1.2.10), the tensor algebra becomes an *associative, graded algebra.*

- *associative* means $(a \otimes b) \otimes c = a \otimes (b \otimes c)$ for all $a, b, c \in T(V)$.
- *graded* means that $T(V)$ can be expressed as a direct sum of homogeneous spaces as in (1.38) and the multiplication adds their degrees $\otimes : \bigotimes_{r,s} V \times \bigotimes_{p,q} V \to \bigotimes_{r+p,s+q} V$.

Remark 1.2.12: The tensor algebra plays an important role in quantum field theory. The Fock spaces constructed in the formalism of second quantisation are tensor algebras of certain vector spaces.

Remark 1.2.13: (Interpretation of (r, s)-tensors)

Together, Theorem 1.2.8 and Definition 1.2.10 imply that we can interpret an (r, s) tensor $\boldsymbol{v}_1 \otimes ... \otimes \boldsymbol{v}_r \otimes \alpha_1 \otimes ... \otimes \alpha_s \in \bigotimes_{r,s} V$ as

- a multilinear map

$$\underbrace{V \times ... \times V}_{s\times} \to \bigotimes_{r,0} V$$
$$(\boldsymbol{x}_1, ..., \boldsymbol{x}_s) \mapsto \alpha_1(\boldsymbol{x}_1) \cdots \alpha_s(\boldsymbol{x}_s) \boldsymbol{v}_1 \otimes ... \otimes \boldsymbol{v}_r$$

- a multilinear map

$$\underbrace{V^* \times ... \times V^*}_{r\times} \to \bigotimes_{0,s} V$$
$$(\beta_1, ..., \beta_r) \mapsto \beta_1(\boldsymbol{v}_1) \cdots \beta_r(\boldsymbol{v}_r) \alpha_1 \otimes ... \otimes \alpha_s$$

- a linear map

$$\bigotimes_{s,0} V \to \bigotimes_{r,0} V$$
$$\boldsymbol{x}_1 \otimes ... \otimes \boldsymbol{x}_s \mapsto \alpha_1(\boldsymbol{x}_1)\cdots\alpha_s(\boldsymbol{x}_s)\boldsymbol{v}_1 \otimes ... \otimes \boldsymbol{v}_r$$

- a linear map

$$\bigotimes_{0,r} \to \bigotimes_{0,s} V$$
$$\beta_1 \otimes ... \otimes \beta_r \mapsto \beta_1(\boldsymbol{v}_1)\cdots\beta_r(\boldsymbol{v}_r)\alpha_1 \otimes ... \otimes \alpha_s$$

- a linear form on $\bigotimes_{s,r} V$

$$\bigotimes_{s,r} V \to k$$
$$\boldsymbol{x}_1 \otimes ... \otimes \boldsymbol{x}_s \otimes \beta_1 \otimes ... \otimes \beta_r \mapsto \alpha_1(\boldsymbol{x}_1)\cdots\alpha_s(\boldsymbol{x}_s)\beta_1(\boldsymbol{v}_1)\cdots\beta_r(\boldsymbol{v}_r).$$

Example 1.2.14: Vectors are $(1,0)$-tensors. Linear forms are $(0,1)$-tensors. Linear maps are $(1,1)$-tensors. Bilinear forms are $(0,2)$-tensors.

Lemma 1.2.15: (Coefficients and transformation behaviour)

1. With respect to a basis $B=\{e_1,...,e_n\}$ of V and dual basis $B^*=\{e^1,...,e^n\}$, a general element in $\bigotimes_{r,s} V$ can be expressed as

$$\boldsymbol{x} = \sum_{i_1,...,i_r,j_1,...,j_s=1}^{n} x^{i_1...i_s}_{j_1...j_s}\ e_{i_1}\otimes ... \otimes e_{i_r} \otimes e^{j_1} \otimes ... \otimes e^{j_s} \tag{1.41}$$

The upper indices are called *covariant* indices and lower indices *contravariant* indices.

2. We consider linear maps $\phi_1,..,\phi_r,\psi_1,...\psi_s \in \mathrm{End}(V)$ with matrix elements

$$\phi_k(e_i) = \sum_{j=1}^{n} (\phi_k)_i{}^j e_j \qquad \psi_l(e_i) = \sum_{j=1}^{n} (\psi_l)_i{}^j e_j \tag{1.42}$$

Then, the transformation of a (r,s) tensor under $\phi_1 \otimes ... \otimes \phi_r \otimes \psi_1 \otimes ... \otimes \psi_s$ is given by

$$\phi_1 \otimes ... \otimes \phi_r \otimes \psi_1 \otimes ... \otimes \psi_s(e_{i_1} \otimes ... \otimes e_{i_r} \otimes e^{j_1} \otimes ... \otimes e^{j_s}) = \tag{1.43}$$
$$\sum_{k_1,...,k_r,l_1,...l_s=1}^{n} (\phi_1)_{i_1}{}^{k_1}\cdots(\phi_r)_{i_r}{}^{k_r}(\psi_1)_{l_1}{}^{j_1}\cdots(\psi_s)_{l_s}{}^{j_s}\ e_{k_1}\otimes ... \otimes e_{k_r} \otimes e^{l_1} \otimes ... \otimes e^{l_s}.$$

and the coefficients transform according to

$$x^{i_1...i_s}_{j_1...j_s} \mapsto \sum_{k_1,...,k_r,l_1,...l_s=1}^{n} (\phi_1)_{i_1}{}^{k_1}\cdots(\phi_r)_{i_r}{}^{k_r}(\psi_1)_{l_1}{}^{j_1}\cdots(\psi_s)_{l_s}{}^{j_s}\ x^{l_1...l_s}_{k_1...k_r} \tag{1.44}$$

Definition 1.2.16: (Trace and contraction of tensors)

The *trace* is the canonical linear map $\mathrm{Tr} : \mathrm{End}(V) \cong V \otimes V^* \to k$

$$\mathrm{Tr}: \ \boldsymbol{v} \otimes \alpha \mapsto \alpha(\boldsymbol{v}) \qquad \forall \boldsymbol{v} \in V, \alpha \in V^*. \tag{1.45}$$

When expressed in terms of the coefficients with respect to a basis $B = \{e_1, ..., e_n\}$ of V and its dual basis $B^* = \{e^1, ..., e^n\}$, the trace coincides with the matrix trace.

$$\mathrm{Tr}\left(\sum_{i,j=1}^{n} x^i_j e_i \otimes e^j\right) = \sum_{i=1}^{n} x^i_i \tag{1.46}$$

Exercise 9:

Suppose that the following quantities denote the coefficients of a (r, s)-tensor, i.e. an element of $\bigotimes_{r,s} V$, with respect to a basis $B = \{e_1, ..., e_n\}$ of V and its dual $B^* = \{e^1, ..., e^n\}$

$$g_{ij} \qquad x^{ij}_{mn} \qquad y^i_{jik} \qquad z^{ijk}. \tag{1.47}$$

Determine (r, s) for each of them and give expressions for the five maps in remark 1.2.13 in terms of B and B^* and in terms of the coefficients. Use Einstein summation convention.

Example:

w^k_{ij} is the coefficient of a (1,2)-tensor. The associated map $\varphi : V \times V \to V$ is given by

$$\varphi(e_i, e_j) = w^k_{ij} e_k, \qquad \varphi(\boldsymbol{x}, \boldsymbol{y})^k = w^k_{ij} x^i y^j.$$

The associated map $\phi : V^* \to V^* \otimes V^*$ is given by

$$\phi(e^k) = w^k_{ij} e^i \otimes e^j \qquad \phi(\alpha)_{ij} = w^k_{ij} \alpha_k.$$

The associated map $\psi : V \otimes V \to V$ is given by

$$\psi(e_i \otimes e_j) = w^k_{ij} e_k \qquad \psi(\boldsymbol{x} \otimes \boldsymbol{y})^k = w^k_{ij} x^i y^j.$$

1.3 Alternating forms and exterior algebra

We will now consider a special type of multilinear forms, namely the ones that are *alternating* or *anti-symmetric*, and we will construct the associated dual object.

Definition 1.3.1: (Alternating or exterior k-form)

An *alternating* (or *exterior*) k-form on V is a k-form on V such that

$$\sigma\alpha(\boldsymbol{v}_1, ..., \boldsymbol{v}_k) = \alpha(\boldsymbol{v}_{\sigma(1)}, ..., \boldsymbol{v}_{\sigma(k)}) = \mathrm{sig}(\sigma)\alpha(\boldsymbol{v}_1, ..., \boldsymbol{v}_k) \qquad \forall \boldsymbol{v}_1, ..., \boldsymbol{v}_k \in V, \sigma \in S(k),$$

where the *sign sig*(σ) *of the permutation* $\sigma \in S(k)$ is $\mathrm{sig}(\sigma) = (-1)^n$, where n is the number of elementary transpositions, i.e. exchanges of neighbouring elements needed to obtain σ or, equivalently, the number of pairs (i, j), $i, j \in \{1, ..., k\}$ with $i < j$ and $\sigma(i) > \sigma(j)$. Explicitly,

$$\mathrm{sig}(\sigma) = \Pi_{i<j} \frac{i - j}{\sigma(i) - \sigma(j)}. \tag{1.48}$$

The alternating k-forms on V form a vector space which we denote by $\mathrm{Alt}^k(V)$. The *vector space Alt*(V) *of alternating forms on* V is the direct sum $\mathrm{Alt}(V) = \bigoplus_{k=0}^{\infty} \mathrm{Alt}^k(V)$.

Example 1.3.2: For $\sigma \in S(3)$, $\sigma : (1,2,3) \mapsto (3,1,2)$ $\mathrm{sig}(\sigma) = 1$, since there are two pairs, $(1,3)$ and $(2,3)$, whose order is inverted by sigma and two elementary transpositions are needed to obtain sigma

$$(1,2,3) \mapsto (1,3,2) \mapsto (3,1,2).$$

For $\sigma \in S(4)$, $\sigma : (1,2,3,4) \mapsto (2,4,1,3)$, we have $\mathrm{sig}(\sigma) = -1$ since there are three pairs, $(1,2)$, $(1,4)$ and $(3,4)$, with $i < j$ and $\sigma(i) > \sigma(j)$ and three elementary transpositions are needed to obtain σ

$$(1,2,3,4) \mapsto (1,2,4,3) \mapsto (2,1,4,3) \mapsto (2,4,1,3).$$

Example 1.3.3: We have $\mathrm{Alt}^1(V) = V^*$. Alternating one-forms are one-forms. Alternating two-forms are antisymmetric bilinear forms on V.

Lemma 1.3.4: (Properties of $\mathrm{Alt}^k(V)$)

1. An alternating k-form vanishes on $(\boldsymbol{v}_1, ..., \boldsymbol{v}_k)$ if are $\boldsymbol{v}_1, ..., \boldsymbol{v}_k$ linearly dependent.
2. This implies in particular that it is antisymmetric

$$\alpha(\boldsymbol{v}_1, ..., \boldsymbol{v}_j, ..., \boldsymbol{v}_i, ...\boldsymbol{v}_n) = -\alpha(\boldsymbol{v}_1, ..., \boldsymbol{v}_i, ..., \boldsymbol{v}_j, ..., \boldsymbol{v}_n). \tag{1.49}$$

3. $\mathrm{Alt}^k(V) = \{0\}$ for all $k > \dim(V)$.

As we have seen in the previous section, there is a duality between n-forms on V and tensor products $\otimes^n V$, that allows us to interpret n-forms as elements of $\bigotimes^n V^*$ and led to the definition of the tensor algebra. We would now like to specialise this definition to *alternating* k-forms and construct an object that takes the role of tensor algebra for alternating forms. This object is the exterior algebra of a vector space V. To define it, we have to introduce the *alternator*.

Definition 1.3.5: (Alternator, k-fold exterior product of V)

The *alternator* is the bilinear map

$$\boldsymbol{a}_k : \bigotimes_{k,0} V \to \bigotimes_{k,0} V \tag{1.50}$$

$$\boldsymbol{v}_1 \otimes .... \otimes \boldsymbol{v}_k \mapsto \boldsymbol{a}_k(\boldsymbol{v}_1 \otimes ... \otimes \boldsymbol{v}_k) = \tfrac{1}{k!} \sum_{\sigma \in S(k)} \mathrm{sig}(\sigma)\ \boldsymbol{v}_{\sigma(1)} \otimes ... \otimes \boldsymbol{v}_{\sigma(k)}. \tag{1.51}$$

Its image $\mathrm{Im}(\boldsymbol{a}_k) \subset \otimes_{k,0} V$ is called the *k-fold exterior product* and denoted $\Lambda^k V$.

Lemma 1.3.6: The alternator is a projector from $\bigotimes_{k,0} V$ onto its image $\Lambda^k V$: $\boldsymbol{a}_k \circ \boldsymbol{a}_k = \boldsymbol{a}_k$.

Proof: Exercise.

Lemma 1.3.7: $\mathrm{Alt}^k(V) = \Lambda^k(V^*)$. $\Lambda^k V$ has properties analogous to those in Lemma 1.3.4:

1. $\boldsymbol{x} \in \Lambda^k(V) \Rightarrow \boldsymbol{x}(\alpha_1, ..., \alpha_k) = 0$ if $\alpha_1, ..., \alpha_k$ linearly dependent.
2. $\Lambda^k V = \{0\}$ for $k > \dim(V)$.

Definition 1.3.8: (Exterior algebra, wedge product)

1. The *exterior algebra* or *Grassmann algebra* is the vector space given as the direct sum
$$\Lambda V = \bigoplus_{k=0}^{\dim(V)} \Lambda^k V. \tag{1.52}$$

2. The *wedge product* $\wedge : \Lambda^k V \times \Lambda^l V \to \Lambda^{k+l} V$ is the unique bilinear map that satisfies
$$\boldsymbol{a}_k(\boldsymbol{x}) \wedge \boldsymbol{a}_l(\boldsymbol{y}) = \tfrac{(k+l)!}{k!l!} \boldsymbol{a}_{k+l}(\boldsymbol{x} \otimes \boldsymbol{y}) \qquad \forall \boldsymbol{x} \in \Lambda^k V, \boldsymbol{y} \in \Lambda^l V. \tag{1.53}$$

3. By extending the wedge product bilinearly to ΛV, we obtain a bilinear map $\wedge : \Lambda V \times \Lambda V \to \Lambda V$. This gives ΛV the structure of an *associative, graded algebra.*

Remark 1.3.9: Equation (1.53) defines the wedge product consistently and uniquely.

Lemma 1.3.10: (Properties of the wedge product)

The wedge product is

1. bilinear:
$\alpha \wedge (t\beta + s\gamma) = t\alpha \wedge \beta + s\alpha \wedge \gamma$, $(t\beta + s\gamma) \wedge \alpha = t\beta \wedge \alpha + s\gamma \wedge \alpha$ $\forall t, s \in k$, $\alpha, \beta, \gamma \in \Lambda V$
2. associative: $(\alpha \wedge \beta) \wedge \gamma = \alpha \wedge (\beta \wedge \gamma)$ for all $\alpha, \beta, \gamma \in \Lambda V$
3. graded anti-commutative (skew): For $\alpha \in \Lambda^k V, \beta \in \Lambda^l V$: $\alpha \wedge \beta = (-1)^{kl} \beta \wedge \alpha$.
4. natural: $\phi^*(\alpha \wedge \beta) = \phi^*(\alpha) \wedge \phi^*(\beta)$ for all $\phi \in \mathrm{End}(V)$, $\alpha, \beta \in \Lambda(V^*) = \mathrm{Alt}(V)$.
5. related to the determinant:
$$\alpha_1 \wedge ... \wedge \alpha_k(v_1, ..., v_k) = \det\left(\alpha_i(v_j)\right)_{i,j \in \{1,...,k\}} \qquad \forall \alpha_1, ..., \alpha_k \in \Lambda^1 V^* \tag{1.54}$$
6. given by the identity
$$\boldsymbol{x}_1 \wedge ... \wedge \boldsymbol{x}_l = \frac{(k_1 + ... + k_l)!}{k_1! k_2! \cdots k_l!} \boldsymbol{a}_{k_1+...+k_l}(\boldsymbol{x}_1 \otimes ... \otimes \boldsymbol{x}_l) \qquad \boldsymbol{x}_i \in \Lambda^{k_i} V. \tag{1.55}$$

Theorem 1.3.11: (Basis of $\Lambda^k V$)

For any basis $B = \{e_1, ..., e_n\}$ of V, a basis of $\Lambda^k V$ is given by the ordered k-fold wedge products of elements in B
$$B_{\Lambda^k V} = \{e_{i_1} \wedge e_{i_2} ... \wedge e_{i_k}, 0 < i_1 < i_2 < ... < i_k \leq n\}. \tag{1.56}$$

This implies $\dim(\Lambda^k V) = \begin{pmatrix} n \\ k \end{pmatrix} = \frac{n!}{k!(n-k)!}$. In particular, $\Lambda^n V$ is one-dimensional and spanned by $e_1 \wedge e_2 \wedge ... \wedge e_n$.

Exercise 10: (Transformation under linear maps)

Show that for any linear map $\phi \in \mathrm{End}(V)$ with $\det(\phi) = 1$ and associated matrix $A_\phi = (\phi^i{}_j)_{i,j \in \{1,...,n\}}$, we have
$$\phi^*(e^1) \wedge ... \wedge \phi^*(e^n)(\boldsymbol{v}_1, ..., \boldsymbol{v}_n) = e^1 \wedge ... \wedge e^n(\phi(\boldsymbol{v}_1), ..., \phi(\boldsymbol{v}_n)) = \det(A_\phi). \tag{1.57}$$

The fact that the dimension of $\Lambda^k V$ is $\dim(\Lambda^k V) = \frac{n!}{k!(n-k)!} = \dim(\Lambda^{n-k} V)$ suggests that there should be an identification between the vector spaces $\Lambda^k V$ and $\Lambda^{n-k} V$. Note, however, that this identification is not canonical, since it requires additional structure, namely a non-degenerate symmetric (or hermitian) bilinear form on V.

Lemma 1.3.12: (Hodge operator)

1. Consider a vector space V over k and let α be a non-degenerate symmetric or hermitian bilinear form on V. We obtain a symmetric, non-degenerate bilinear form on $\Lambda^k V$ by choosing a basis $B = \{e_1, ..., e_n\}$ of V and setting

$$\tilde\alpha : \Lambda^k V \times \Lambda^k V \to \mathbb{R} \qquad \tilde\alpha(e_{i_1} \wedge ... \wedge e_{i_k}, e_{j_1} \wedge ... \wedge e_{j_k}) = \alpha(e_{i_1}, e_{j_1}) \cdots \alpha(e_{i_k}, e_{j_k})$$
$$\text{for } 0 < i_1 < i_2 < ... < i_k \leq n, 0 < j_1 < j_2 < ... < j_n \leq n.$$

   By Exercise 6 there exists a basis in which $\alpha(e_i, e_j) = \epsilon_i \delta_{ij}$ with $\epsilon_i \in \{\pm 1\}$. For this basis, we have

$$\tilde\alpha(e_{i_1} \wedge ... \wedge e_{i_k}, e_{j_1} \wedge ... \wedge e_{j_k}) = \epsilon_{i_1} \cdots \epsilon_{i_k} \delta_{i_1, j_1} \cdots \delta_{i_k, j_k} \tag{1.58}$$
$$\text{for } 0 < i_1 < i_2 < ... < i_k \leq n, 0 < j_1 < j_2 < ... < j_n \leq n.$$

2. There is a unique bilinear map $*_\alpha : \Lambda^k V \to \Lambda^{n-k} V$ the *Hodge operator* or *Hodge star* that satisfies $\boldsymbol{x} \wedge (*_\alpha \boldsymbol{y}) = \tilde\alpha(\boldsymbol{x}, \boldsymbol{y}) e_1 \wedge ... \wedge e_n$ for all $\boldsymbol{x}, \boldsymbol{y} \in \Lambda^k V$. It is given by its values on a basis of $\Lambda^k V$

$$*_\alpha(e_{i_1} \wedge ... \wedge e_{i_k}) = (-1)^{\mathrm{sgn}(\sigma)} \epsilon_{i_1} \cdots \epsilon_{i_k} e_{j_1} \wedge ... \wedge e_{j_{n-k}},$$

   where $0 < i_1 < i_2 < ... < i_k \leq n$, $0 < j_1 < ... < j_{n-k} \leq n$, $\{i_1, ..., i_k\} \cup \{j_1, ..., j_{n-k}\} = \{1, ..., n\}$ and $\sigma \in S_n$ is the permutation

$$\sigma > \ (i_1, ..., i_k, j_1, ..., j_{n-k}) \to (1, ..., n).$$

   The Hodge operator satisfies

   (a) $*_\alpha(e_1 \wedge ... \wedge e_n) = (-1)^{\epsilon_1 \cdots \epsilon_n}$, $*_\alpha 1 = e_1 \wedge ... \wedge e_n$
   (b) For all $\boldsymbol{x} \in \Lambda^k V$, $\boldsymbol{y} \in \Lambda^{n-k} V$: $\tilde\alpha(\boldsymbol{x}, *_\alpha \boldsymbol{y}) = (-1)^{k(n-k)} \tilde\alpha(*_\alpha \boldsymbol{x}, \boldsymbol{y})$
   (c) For all $\boldsymbol{x} \in \Lambda^k V$: $*_\alpha(*_\alpha \boldsymbol{x}) = (-1)^{k(n-k) + \epsilon_1 \cdots \epsilon_n} \boldsymbol{x}$.

Example 1.3.13: (Angular momentum, wedge product)

We consider $\mathbb{R}^3$ with the standard scalar product $\alpha = g_E$. For vectors $\boldsymbol{x} = \sum_{i=1}^3 x^i e_i$, $\boldsymbol{p} = \sum_{i=1}^n p^i e_i$, we set

$$\boldsymbol{l} = \sum_{i=1}^3 l^i e_i = *_\alpha(\boldsymbol{x} \wedge \boldsymbol{p}). \tag{1.59}$$

Then, the components of $\boldsymbol{l} = \sum_{i=1}^3 l^i e_i$ are given by

$$l^1 = x^2 p^3 - p^2 x^3 \qquad l^2 = x^3 p^1 - x^1 p^3 \qquad l^3 = x^1 p^2 - x^2 p^1. \tag{1.60}$$

We recover the usual expression for the angular momentum. This also explains the use of the symbol $\wedge$ for the wedge or cross product of two vectors in $\mathbb{R}^3$.

We now consider the transformation of $\boldsymbol{l}$ under the linear map $\phi : \mathbb{R}^3 \to \mathbb{R}^3$, $\boldsymbol{y} \to -\boldsymbol{y}$. While the vectors $\boldsymbol{x}$ and $\boldsymbol{p}$ transform as $\boldsymbol{x} \to -\boldsymbol{x}$, $\boldsymbol{p} \to -\boldsymbol{p}$, we have $\boldsymbol{x} \wedge \boldsymbol{p} \to \boldsymbol{x} \wedge \boldsymbol{p}$. with the definition $\boldsymbol{l}' = *_\alpha(\phi(\boldsymbol{x}) \wedge \phi(\boldsymbol{p}))$, we find $\boldsymbol{l}' = \boldsymbol{l}$. This is due to the fact that the bilinear form α is *not invariant* under the linear map ϕ, as can be shown using Exercise 10.

The angular momentum therefore does not transform as a usual vector under the map ϕ. One says that $\boldsymbol{l}$ *transforms as a pseudovector or axial vector.* Similarly, one can define a *pseudoscalar* s by setting $s = *_\alpha \boldsymbol{x} \wedge \boldsymbol{p} \wedge \boldsymbol{k}$, where $\boldsymbol{x}, \boldsymbol{p}, \boldsymbol{k}$ are three linearly independent vectors in $\mathbb{R}^3$. Under the map ϕ, s transforms as $s \to s' = *_\alpha(\boldsymbol{x}' \wedge \boldsymbol{p}' \wedge \boldsymbol{k}') = -s$.

We generalise this observation to the following definition and lemma

Definition 1.3.14: (Pseudovector, pseudoscalar)

Let V be a vector space of odd dimension over $\mathbb{R}$, α a non-degenerate bilinear form as in Lemma 1.3.12 with associated orthonormal basis $B = \{e_1, ..., e_n\}$ and exterior algebra ΛV. Then:

- elements of $\Lambda^0 V = k$ are called *scalars*
- elements of $\Lambda^1 V = V$ are called *vectors*
- elements of $\Lambda^{n-1} V \cong V$ are called *pseudovectors*
- elements of $\Lambda^n V \cong k$ are called *pseudoscalars.*

Lemma 1.3.15: An *oriented* vector space V is a vector space V over $\mathbb{R}$ together with the choice of an *ordered* basis B. Two ordered bases are said to *induce the same orientation* if they are related by a linear map $\phi \in \mathrm{Aut}(V)$ with positive determinant. A linear map $\phi \in \mathrm{Aut}(V)$ is called *orientation preserving* if $\det \phi > 0$ and *orientation reversing* if $\det \phi < 0$.

Let V be a vector space over $\mathbb{R}$ of odd dimension. Under an orientation reversing linear map $\phi \in \mathrm{Aut}(\phi)$, $\det \phi = -1$ with matrix coefficients $\phi(e_i) = \sum_{j=1}^n {\phi_i}^j e_j$ scalars, pseudoscalars, vectors and pseudovectors transform according to

- *scalars* $s \in \Lambda^0 V$: $s \to s$
- *vectors* $\boldsymbol{x} = \sum_{i=1}^n x^i e_i \in \Lambda^1 V$: $x^i \to {\phi_j}^i x^j$
- *pseudovectors* $\boldsymbol{y} = \sum_{i=1}^n y^i e_1 \wedge ...\widehat{e_i}... \wedge e_n$: $y^i \to -{\phi_j}^i y^j$
- *pseudoscalars* $r \cdot e_1 \wedge ... \wedge e_n \in \Lambda^n V$: $r \to -r$.

Definition 1.3.14 and Lemma 1.3.15 solve the puzzle about the transformation behaviour of pseudovectors and pseudoscalars. They are not really vectors or scalars, but, respectively, the tuple of coefficients of $n-1$-forms and n-forms. In odd-dimensional spaces, an orientation reversing map changes the sign of the form which spans the space of n-forms and leaves invariant the sign of the basis vectors which span the space of $n-1$-forms. This explains the resulting transformation behaviour of the coefficients.

1.4 Vector fields and differential forms

Definition 1.4.1: (Submanifolds of $\mathbb{R}^n$)

A subset $M \subset \mathbb{R}^n$ is called a *submanifold of dimension k* of $\mathbb{R}^n$ if for all points $\forall p \in M$ there exist open neighbourhoods $U \subset \mathbb{R}^n$ and functions $f_1, ..., f_{n-k} : U \to \mathbb{R} \in \mathcal{C}^\infty(\mathbb{R}^n)$ such that

1. $M \cap U = \{x \in U : f_1(x) = ... = f_{n-k}(x) = 0\}$.
2. The gradients $\text{grad} f_1(p)$,... ,$\text{grad} f_n(p)$ are linearly independent, where $\text{grad} f = \sum_{i=1}^n \partial_i f e_i$.

Example 1.4.2: (Submanifolds of $\mathbb{R}^n$)

1. Any open subset $U \subset \mathbb{R}^n$ is a n-dimensional submanifold of $\mathbb{R}^n$.
2. Linear subspaces $V \subset \mathbb{R}^n$ are submanifolds of $\mathbb{R}^n$.
3. The $(n-1)$-sphere $S^{n-1} = \{\boldsymbol{x} = \sum_{i=1}^n x^i e_i \in \mathbb{R}^n \mid \sum_{i=1}^n (x^i)^2 = 1\}$ is a $n-1$-dimensional submanifold of $\mathbb{R}^n$.
4. The $(n-1)$-dimensional hyperbolic space $\mathbb{H}^{n-1} = \{\boldsymbol{x} = \sum_{i=0}^{n-1} x^i e_i \in \mathbb{R}^n \mid -(x^0)^2 + \sum_{i=1}^{n-1} (x^i)^2 = -1\}$ is a $n-1$-dimensional submanifold of $\mathbb{R}^n$.

Definition 1.4.3: (Tangent vector, tangent space)

A vector $\boldsymbol{v} \in \mathbb{R}^n$ is called a *tangent vector* in p at a submanifold $M \subset \mathbb{R}^n$ if there exists a smooth curve $c :]-\epsilon, \epsilon[\to M$, $c(0) = p$, $\dot{c}(0) = \boldsymbol{v}$.

The *tangent space on M at p* is the set $T_pM = \{\boldsymbol{v} \in \mathbb{R}^n \mid \boldsymbol{v}\,\text{tangent vector at}\ M\ \text{in}\ p\}$. It is a k-dimensional linear subspace of $\mathbb{R}^n$. The dual vector space T_p^*M is called the *cotangent space* on M at p.

The set $TM = \bigcup_{p\in M} T_pM$ is called *tangent bundle*, the set $T^*M = \bigcup_{p\in M} (T_pM)^*$ the *cotangent bundle* of M. The tangent bundle TM and the cotangent bundle T^*M are equipped with canonical projections $\pi_M : TM \to M$, $\boldsymbol{v} \in T_pM \mapsto p \in M$ and $\tilde{\pi}_M : T^*M \to M$, $\alpha \in (T_p^M)^* \mapsto p \in M$.

Example 1.4.4:

1. For open subsets $U \subset \mathbb{R}^n$, we have $T_pU = \mathbb{R}^n$ for all $p \in U$.
2. If $V \subset \mathbb{R}^n$ is linear subspace, we have $T_pV = V$ for all $p \in V$.
3. For all $p \in S^{n-1}$ the tangent space in p is given by $T_pS^{n-1} = p^\perp = \{\boldsymbol{x} \in \mathbb{R}^n : \sum_{i=1}^n x^i p^i = 0\}$.
4. For all $p \in H^{n-1}$, we have $T_pH^{n-1} = p^\perp = \{\boldsymbol{x} \in \mathbb{R}^n : x^0p^0 - \sum_{i=1}^{n-1} x^i p^i = 0\}$.

Proof: The first two examples are trivial. In the case of S^{n-1}, curves $c :]-\epsilon, \epsilon[\to M$ with $c(0) = p$ must satisfy $\sum_{i=1}^n c_i(t)^2 = 1$. This implies

$$\frac{d}{dt}|_{t=0}\left(\sum_{i=1}^n c^i(t)^2\right) = 2\sum_{i=1}^n \dot{c}^i(0)c^i(0) = 2\sum_{i=1}^n \dot{c}^i(0)p^i = 0$$

and therefore $\dot{c}(0) \in p^\perp$. The proof for H^{n-1} is analogous. □

Definition 1.4.5: (Vector fields, tensor fields and differential forms)

1. We consider a submanifold $M \subset \mathbb{R}^n$. A *vector field* on M is a map

$$\boldsymbol{Y} : M \to TM \qquad p \mapsto \boldsymbol{Y}(p) \in T_pM.$$

   The set of all vector fields on M, denoted $\mathrm{Vec}(M)$, is a vector space over $\mathbb{R}$ with respect to pointwise addition and multiplication by $\mathbb{R}$.

2. A (r,s)-*tensor field* on M is a map

$$g : M \to \bigotimes_{r,s} TM \qquad p \mapsto g(p) \in \bigotimes_{r,s} T_pM.$$

   The set of all tensor fields on M is a vector space over $\mathbb{R}$ with respect to pointwise addition and multiplication by $\mathbb{R}$.

3. A *differential form of order* k on M (or k*-form on* M) is a map

$$\omega : M \to \bigcup_{p\in M} \Lambda^k(T_p^*M) \qquad \boldsymbol{p} \mapsto \omega(\boldsymbol{p}) \in \Lambda^k T_p^*(M).$$

   The set of all k-forms on M, denoted $\Omega^k(M)$, is a vector space on M with respect to pointwise addition and multiplication by $\mathbb{R}$. 0-forms are functions on M.

In the following we will mostly restrict attention to submanifolds $U \subset \mathbb{R}^n$ that are open subsets. However, most of the statements that follow generalise to the situation where U is a general submanifold of $\mathbb{R}^n$ or, generally, a manifold.

Lemma 1.4.6: (Expression in coordinates)

Let $U \subset \mathbb{R}^n$ be an open subset and $x^1, ..., x^n$ be the canonical coordinates on $\mathbb{R}^n$ with respect to its standard basis $B = \{e_1, ..., e_n\}$. We denote by $B^* = \{e^1, ..., e^n\}$ the dual basis. The *basis vector fields* associated with these coordinates are

$$\partial_i : U \to TU \qquad p \mapsto \partial_i(p) = e_i \qquad \forall p \in U, i \in \{1, ..., n\}. \tag{1.61}$$

The associated *basis one-forms* are

$$dx^i : U \to T^*U \qquad p \mapsto dx^i(p) = e^i \qquad \forall p \in U, i \in \{1, ..., n\}. \tag{1.62}$$

1. A vector field on U can be expressed as a linear combination of the basis vector fields with coefficients that are functions on U

$$\boldsymbol{Y} = \sum_{i=1}^n y^i \partial_i \qquad y^i : U \to \mathbb{R}. \tag{1.63}$$

   A vector field is called *continuous*, *differentiable*, *smooth* etc. if all coefficient functions are continuous, differentiable, smooth etc. We denote by $\mathrm{Vec}(U)$ the vector space of smooth vector fields on U.

2. A k-form $\omega : U \to \Lambda^k T^*U$ on U can be expressed as:

$$\omega = \sum_{1\leq i_1<\ldots<i_k\leq n} f_{i_1\ldots i_k} dx^{i_1}\wedge\ldots\wedge dx^{i_k} \qquad f_{i_1\ldots i_k} : U \to \mathbb{R}. \tag{1.64}$$

   A k-form ω is called *continuous*, *differentiable*, *smooth* etc. if all coefficient functions $f_{i_1\ldots i_k} : U \to \mathbb{R}$ are continuous, differentiable, smooth etc. We denote by $\Omega^k(U)$ the vector space of smooth k-forms on U.

3. A (r,s)-tensor field $g : U \to \bigotimes_{r,s} TM$ can be expressed in terms of the basis vector fields and the basis of one-forms as

$$g = \sum_{i_1,\ldots,i_r,j_1,\ldots,j_s=1}^{n} g^{i_1\cdots i_r}_{j_1\cdots j_s} \partial_{i_1}\otimes\ldots\otimes\partial_{i_r}\otimes dx^{j_1}\otimes\ldots\otimes dx^{j_s} \qquad g^{i_1\cdots i_r}_{j_1\cdots j_s} : U \to \mathbb{R}.$$

   A (r,s)-tensor field is called *continuous*, *differentiable*, *smooth* etc. if all coefficient functions $g^{i_1\cdots i_r}_{j_1\cdots j_s} : U \to \mathbb{R}$ are continuous, differentiable, smooth etc.

Definition 1.4.7: (Action of a vector field on a function)

The action of a smooth vector field $\boldsymbol{X} \in \mathrm{Vec}(U)$ on functions $f \in \mathcal{C}^\infty(U)$ defines a map $\mathcal{C}^\infty(U) \to \mathcal{C}^\infty(U)$, $f \to \boldsymbol{X}.f$ given by

$$\boldsymbol{X}.f(p) := \frac{d}{dt}|_{t=0} f(p + t\boldsymbol{X}(p)). \tag{1.65}$$

For $\boldsymbol{X} = \sum_{i=1}^n x^i\partial_i$, $x^i \in \mathcal{C}^\infty(U)$, we have

$$\boldsymbol{X}f(p) = \sum_{i=1}^n x^i(p)\partial_i f(p). \tag{1.66}$$

In particular, the basis vector fields ∂_i act by partial derivatives $\partial_i.f(p) = \frac{d}{dt}|_{t=0} f(p + te_i)$.

We can thus interpret vector fields as *differential operators* on the vector space $\mathcal{C}^\infty(M)$ of smooth functions on a manifold M. Moreover, the preceding definitions imply that we can think of a vector field as something that attaches an element of a vector space V to every point of a submanifold $M \subset \mathbb{R}^n$ and of a differential form as something that attaches an element of the dual vector space V^* to each point of M. All structures that we have encountered in the previous sections - duality, exterior algebra - will give rise to corresponding structures for vector fields and differential forms. These structures are obtained by defining them pointwise.

Definition 1.4.8: (Operations on k-forms)

1. *Sum of k-forms and products with functions:* For k-forms $\omega, \alpha \in \Omega^k(U)$ and functions $f \in \mathcal{C}^\infty(U)$, the sum $\omega+\alpha \in \Omega^k(U)$ and the product $f\cdot\omega \in \Omega^k(U)$ are defined pointwise

$$(\omega+\alpha)(p) = \omega(p) + \alpha(p) \qquad (f\omega)(p) = f(p)\omega(p) \qquad \forall p \in U. \tag{1.67}$$

2. *Wedge product:* The wedge product of differential forms $\omega \in \Omega^k(U)$, $\alpha \in \Omega^l(U)$, is the $(k+l)$-form $\omega\wedge\alpha \in \Omega^{k+l}$ defined pointwise as

$$(\omega\wedge\alpha)(p) = \omega(p)\wedge\alpha(p) \qquad \forall p \in U. \tag{1.68}$$

Remark 1.4.9: All properties and identities we derived previously for the exterior algebra hold pointwise. In particular, the wedge product satisfies identities analogous to the ones in Lemma 1.3.10.

To generalise the Hodge star to k-forms, we need additional structure, namely an assignment of a non-degenerate symmetric bilinear form on T_pU to every point $p \in U$. This turns U into a *pseudo-Riemannian manifold* or, if the form is positive definite, into a *Riemannian manifold.*

Definition 1.4.10: (Riemannian and pseudo-Riemannian manifold)

A submanifold $M \subset \mathbb{R}^n$ is called a *pseudo-Riemannian manifold* if it is equipped with a smooth symmetric non-degenerate $(0,2)$-tensor field, i.e. a smooth map

$$g : M \to T_p^*M \otimes T_p^*M \qquad p \mapsto g(p) \in T_p^*M \otimes T_p^*M \tag{1.69}$$

that assigns to every point $p \in M$ a symmetric, non-degenerate bilinear form $g(p)$ on T_pM. The $(0,2)$-tensor field g is called a *pseudo-Riemannian metric on* M. If $g(p)$ is positive definite for all $p \in M$, g is called a *Riemannian metric on* M and M is called a *Riemannian manifold.*

In local coordinates the metric g takes the form

$$g = \sum_{i,j=1}^n g_{ij} dx^i \otimes dx^j \qquad g_{ij} = g_{ji} : M \to \mathbb{R} \text{ smooth.} \tag{1.70}$$

Definition 1.4.11: (Hodge operator)

Let $U \subset \mathbb{R}^n$ be an open subset of $\mathbb{R}^n$, g a pseudo-Riemannian metric on U. Then there exist vector fields $E_1, ..., E_n \in \text{Vec}(U)$, the *orthonormal basis* for g, such that $E_1(p), ..., E_n(p)$ are linearly independent for all $p \in U$ and

$$g_p(E_i(p), E_j(p)) = \epsilon_i \delta_{ij} \qquad \epsilon_i \in \{\pm 1\} \qquad \forall p \in U.$$

In terms of the basis vector fields $\partial_1, ..., \partial_n \in \text{Vec}(U)$, these vector fields are given by coefficient functions $U_i^{\;j} \in \mathcal{C}^\infty(U)$

$$E_i = \sum_{j=1}^n U_i^{\;j} \partial_j$$

such that the matrix $U_i^{\;j}(p)$ is invertible for all $p \in U$. We denote by $(U^{-1})_i^{\;j} \in \mathcal{C}^\infty(U)$ the components of its inverse, i.e. the functions characterised by the condition $\sum_{j=1}^n U_i^{\;j} (U^{-1})_j^{\;k} = \delta_i^k$. Then the one-forms dX^i defined by

$$dX^i = \sum_{j=1}^n (U^{-1})_j^{\;i} dx^j$$

are dual to the vector fields E_i: $dX^j(E_i) = \delta_i^j$. We define a non-degenerate symmetric bilinear form $\tilde{g}$ on $\Omega^k(U)$ by setting

$$\begin{aligned} &\bar{g}(dX^i, dX^j) = \epsilon_i \delta^{ij} \\ &\tilde{g}(dX^{i_1} \wedge ... \wedge dX^{i_k}, dX^{j_1} \wedge ... \wedge dX^{j_k}) = \bar{g}(dX^{i_1}, dX^{j_1}) \cdots \bar{g}(dX^{i_k}, dX^{j_k}) \end{aligned} \tag{1.71}$$

for $i_1 < i_2 < ... < i_k$, $j_1 < ... < j_k$, $i_k, j_k \in \{1, ..., n\}$. The *Hodge operator* is the unique map $*_g : \Omega^k(U) \to \Omega^{n-k}(U)$ defined by

$$\omega \wedge *_g\eta = \tilde{g}(\omega, \eta) \cdot dx^1 \wedge ... \wedge dx^n. \tag{1.72}$$

In terms of the one-forms dX^i, we have

$$*_g(dX^{i_1} \wedge \ldots \wedge dX^{i_k}) = (-1)^{\text{sgn}(\sigma)}\epsilon_{i_1} \cdots \epsilon_{i_k} dX^{j_1} \wedge \ldots \wedge dX^{j_{n-k}},$$

where $0 < i_1 < \cdots < i_k \leq n$, $0 < j_1 < \cdots < j_{n-k} \leq n$, $\{i_1, ..., i_k\} \cup \{j_1, ..., j_{n-k}\} = \{1, ..., n\}$ and $\text{sgn}(\sigma)$ is the sign of the permutation

$$\sigma : (i_1, ..., i_k, j_1, ..., j_{n-k}) \mapsto (1, ..., n).$$

Remark 1.4.12: The properties and identities in Lemma 1.3.12 generalise to this situation. What is different is that the orthonormal basis of $g(p)$ changes with $p \in U$. The vector fields E_i that diagonalise g and the associated one-forms dX^i are therefore not a constant linear combination of the basis vector fields ∂_i and one-forms dx^i, but a linear combination with coefficients that are *functions* on U.

It remains to investigate how differential forms behave under functions that map the manifold M or the open subset $U \subset \mathbb{R}^n$ to other manifolds. The concept which characterises their behaviour under such transformations is the pull-back.

Definition 1.4.13: (Pull-back of differential forms)

Let $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^m$ be open subsets and let $\omega \in \Omega^k(U)$ be a k-form that is given by

$$\omega = \sum_{1 \leq i_1 < ... < i_k \leq n} f_{i_1 ... i_k}\, dx^{i_1} \wedge ... \wedge dx^{i_n}. \tag{1.73}$$

Suppose there exists a continuously differentiable map $\phi = (\phi_1, ..., \phi_n) : V \subset \mathbb{R}^m \to U \subset \mathbb{R}^n$. Then, the *pull-back* of ω with ϕ is the differential form $\phi^*\omega \in \Omega^k(V)$ defined by

$$\phi^*\omega(p)(\boldsymbol{v}_1, ..., \boldsymbol{v}_n) = \omega(\phi(p))\,(d_p\phi\, \boldsymbol{v}_1, ..., d_p\phi\, \boldsymbol{v}_n) \qquad \forall \boldsymbol{v}_1, ..., \boldsymbol{v}_n \in T_pU,$$

where $d_p\phi$ is the matrix $d_p\phi = (\partial_i\phi_j(p))_{ij}$. In local coordinates, we have

$$\phi^*\omega = \sum_{1 \leq i_1 < ... < i_k \leq n} f_{i_1 ... i_k} \circ \phi\;\; d\phi^{i_1} \wedge ... \wedge d\phi^{i_k} \qquad d\phi^i = \sum_{j=1}^n \partial_j\phi^i\, dx^j \tag{1.74}$$

Lemma 1.4.14: (Properties of the pull-back)

The pull-back of differential forms

1. is *linear*: $\phi^*(t\omega_1 + s\omega_2) = t\phi^*\omega_1 + s\phi^*\omega_2$ for all $\omega_1, \omega_2 \in \Omega^k(U)$, $t, s \in \mathbb{R}$
2. *commutes with the wedge product*: $\phi^*(\omega \wedge \sigma) = (\phi^*\omega) \wedge (\phi^*\sigma)$ for all $\omega \in \Omega^k(M)$, $\sigma \in \Omega^l(M)$
3. satisfies $(\phi \circ \psi)^*\omega = \psi^*(\phi^*\omega)$.

After generalising the structures encountered for alternating forms and the exterior algebra to differential forms, we will now investigate the new structures which arise from the dependence on the points $p \in M$. The first is the *exterior differential of a differential form.*

Definition 1.4.15: (Exterior Differential)

Let $U \subset \mathbb{R}^n$ be an open subset of $\mathbb{R}^n$. The *exterior differential* is a map $d : \Omega(U) \to \Omega(U)$, $d(\Omega^k(U)) \subset \Omega^{k+1}(U)$, defined by

$$d\omega = \sum_{1 \leq i_1 < \ldots < i_k \leq n} df_{i_1,\ldots,i_k} \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_k} = \sum_{1 \leq i_1 < \ldots < i_k \leq n, j \neq i_1,\ldots,i_k} \partial_j f_{i_1 \ldots i_k} \, dx^j \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_k}$$

$$\text{for } \omega = \sum_{1 \leq i_1 < \ldots < i_k \leq n} f_{i_1,\ldots,i_k} \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_k} \in \Omega^k(U) \qquad f_{i_1 \cdots i_k} \in \mathcal{C}^\infty(U). \tag{1.75}$$

Remark 1.4.16: Note that this definition is independent of the choice of coordinates. The exterior differential depends only on the k-form ω itself and is therefore well-defined.

Lemma 1.4.17: (Properties of exterior differential)

The exterior differential

1. is *linear*: $d(t\omega_1 + s\omega_2) = td\omega_1 + sd\omega_2 \ \forall \omega_1, \omega_2 \in \Omega^k(U)$, $t, s \in \mathbb{R}$
2. is *graded anti-commutative (skew)*
$$d(\omega \wedge \sigma) = (d\omega) \wedge \sigma + (-1)^k \omega \wedge d\sigma \qquad \forall \omega \in \Omega^k(U), \sigma \in \Omega^l(U)$$
3. satisfies $d^2\omega = d(d\omega) = 0$
4. Commutes with the *pull-back*: For any continuously differentiable map $\phi = (\phi_1, \ldots, \phi_n) : V \subset \mathbb{R}^m \to U \subset \mathbb{R}^n$, we have $d(\phi^*\omega) = \phi^*(d\omega)$ for all $\omega \in \Omega^k(U)$.

Proof: Exercise.

Remark 1.4.18: The compatibility between exterior differential and pull-back $d(\phi^*)\omega = \Phi^*(d\omega)$ and the identity $d^2\omega = 0$ arise from the fact that the exterior derivative and differential forms are *antisymmetric*. Roughly speaking, the antisymmetry of differential forms and of the exterior derivative kills all second order derivatives which arise when one applies the exterior differential to the pull-back or to the differential form $d\omega$.

Exercise 11:

1. Show that the exterior differential is well-defined and independent of the choice of coordinates. Consider first its transformation under a linear map $\phi : \mathbb{R}^n \to \mathbb{R}^n$ and show that $d(\phi^*)\omega = \Phi^*(d\omega)$ in that case. Prove now the identity $d(\phi^*)\omega = \Phi^*(d\omega)$ for a general (continuously differentiable) map $\phi = (\phi_1, \ldots, \phi_n) : V \subset \mathbb{R}^m \to U \subset \mathbb{R}^n$.

2. We consider an open subset $U \subset \mathbb{R}^n$ and attempt to define a symmetric exterior differential by setting

$$df = \sum_{i=1}^n \partial_i f dx^i \quad \forall f \in \mathcal{C}^\infty(U) \tag{1.76}$$

$$d\omega = \sum_{j=1}^n \tfrac{1}{2} \partial_j f^i (dx^i \otimes dx^j + dx^j \otimes dx^i) \quad \text{for} \quad \omega = \sum_{i=1}^n f_i dx^i \in \Omega^1(U).$$

Show that this differential does not satisfy $d^2 f = 0$ for all $f \in \mathcal{C}^\infty(U)$. Investigate how it changes under the pull-back. Does the identity $d(\phi^*\omega) = \phi^*(d\omega)$ hold for ω?

Example 1.4.19: (Gradient, divergence, curl)

1. The exterior differential of a function $f \in \Omega^0(U)$ is the *gradient*

$$df = \sum_{i=1}^n \partial_i f\, dx^i = \sum_{i=1}^n (\text{grad} f)_i\, dx^i \qquad \forall f \in \Omega^0(U) \tag{1.77}$$

   Note that the gradient defined in this way is a one-form. In order to obtain a gradient which is a vector field on U, we need an identification $(T_pU)^* \cong T_pU$ for all $p \in U$. Such an identification is obtained via a Riemannian metric or via the choice of a basis of T_pU and a dual basis of $(T_pU)^*$ for all $p \in U$.

2. The exterior differential of a one-form $\omega \in \Omega^1(U)$ is the *curl*:

$$d\omega = \tfrac{1}{2} \sum_{i,j=1}^n (\partial_i f_j - \partial_j f_i) dx^i \wedge dx^j \qquad \forall \omega = \sum_{i=1}^n f_i\, dx^i \in \Omega^1(U) \tag{1.78}$$

   For $U \subset \mathbb{R}^3$, we can set $d\omega = g_1\, dx^2 \wedge dx^3 + g_2\, dx^3 \wedge dx^1 + g_3\, dx^1 \wedge dx^2$ and recover the familiar expression for the curl:

$$g_i = (\vec{\nabla} \times \vec{f})_i = \epsilon_i^{\;jk}(\partial_j f_k - \partial_k f_j),$$

   where $\epsilon_i^{\;jk}$ is the totally antisymmetric tensor with $\epsilon_1^{\;23} = 1$.

3. The exterior differential of a $n-1$-form is related to the *divergence* $\text{div}\vec{f} = \sum_{i=1}^n \partial_i f_i$. We consider a $(n-1)$-form

$$\omega = \sum_{i=1}^n (-1)^{i-1} f_i\, dx^1 \wedge dx^2 \wedge \ldots \wedge \widehat{dx^i} \wedge \ldots \wedge dx^n = *_{g_E}\left(\sum_{i=1}^n f_i\, dx^i\right)$$

   where $*_{g_E}$ is the Hodge operator associated with the Euclidean metric g_E on $\mathbb{R}^n$. The differential of ω is given by

$$d\omega = \left(\sum_{i=1}^n \partial_i f_i\right) \cdot dx^1 \wedge \ldots \wedge dx^n = \text{div}\vec{f} \cdot dx^1 \wedge \cdots \wedge dx^n \in \Omega^n(U) \tag{1.79}$$

Exercise 12: Consider an open subset $U \subset \mathbb{R}^3$.

1. Show that for a function $f \in \mathcal{C}^\infty(U)$, the identity $d^2 f = 0$ can be reformulated as

$$\vec{\nabla} \times (\text{grad} f) = 0.$$

2. Show that for a one-form $\omega = \sum_{i=1}^3 f_i\, dx^i$, the identity $d^2\omega = 0$ can be reformulated as

$$\text{div}(\vec{\nabla} \times \vec{f}) = 0.$$

Lemma 1.4.20: (Co-differential)

Let $U \subset \mathbb{R}^n$ be an open subset of $\mathbb{R}^n$, g a pseudo-Riemannian metric of signature $s \leq n$ with associated Hodge star $*_g$. Then, the *co-differential* $\delta = - *_g \circ d \circ *_g : \Omega^{k+1}(U) \to \Omega^k(U)$ has properties analogous to the ones in lemma 1.4.17. In particular: $\delta^2 = *_g \circ d \circ *_g \circ *_g \circ d \circ *_g = 0$

Definition 1.4.21: (closed, exact)

A k-form $\omega \in \Omega^k(U)$ is called *closed* if $d\omega = 0$. It is called *exact* if there exists a $(k-1)$-form $\theta \in \Omega^{k-1}(U)$ such that $\omega = d\theta$.

Clearly, due to the identity $d^2 = 0$, all exact differential forms on an open subset $U \subset \mathbb{R}^n$ are closed. The answer to the question if the converse is also true is in general negative. However, it can be shown that for subsets $U \subset \mathbb{R}^n$ which are star-shaped, every closed form is exact. This is the content of Poincaré's lemma.

Theorem 1.4.22: (Poincaré lemma)

Let $U \subset \mathbb{R}^n$ be open and *star-shaped*, i.e. such that there exists a point $p \in U$ such that for any $q \in U$, the segment $[p, q] = \{(1-t)p + tq \mid t \in [0, 1]\}$ lies in U: $[p, q] \subset U$. Let $\omega \in \Omega^k(U)$ be closed. Then ω is exact.

Remark 1.4.23: Do not confuse *star-shaped* with *convex*. Star-shaped means that there *exists a point* $\boldsymbol{p} \in U$ such that the segments connecting it to any other point in U lie in U. Convex means that this is the case for *all points* $\boldsymbol{p} \in U$. Convex therefore implies star-shaped, but not the other way around.

Exercise 13: (Uniqueness of the exterior differential)

The exterior differential d is a linear map $d : \Omega^k(U) \to \Omega^{k+1}(U)$ that satisfies $d^2 = 0$. In this exercise, we investigate how unique such a structure is. Given the structures we have introduced so far, we attempt to construct other linear maps $\tilde{d} : \Omega^k(U) \to \Omega^{k+1}(U)$ that differ from d non-trivially and satisfy $\tilde{d}^2 = 0$.

1. As a first guess, we could consider modifying d by multiplying it with a function $f \in \Omega^0(M)$. Show that the map $d_f : \Omega^k(M) \to \Omega^{k+1}(M)$, $\omega \mapsto f \cdot d\omega$ does *not* satisfy $d_f^2 = 0$ unless f is constant. Multiplying d by a function therefore does not yield a new differential.

2. As a second guess, we could consider modifying d by using a one-form $\theta \in \Omega^1(U)$. As we need a map $d : \Omega^k(U) \to \Omega^{k+1}(U)$, the canonical way of implementing this would be to define

$$d_\theta(\omega) := d\omega + \theta \wedge \omega \qquad \forall \omega \in \Omega^k(U). \tag{1.80}$$

Show that we have $d_\theta^2 = 0$ if and only if θ is closed. Given a closed $\theta \in \Omega^1(U)$, we call the associated map $d_\theta : \Omega^k(U) \to \Omega^{k+1}(U)$ the *covariant derivative* with respect to θ.

3. We now consider a vector space V with basis $B = \{e_1, .., e_n\}$ and one-forms $\omega \in \Omega^k(U, V)$ which take values in the vector space V. With respect to the basis $B = \{e_1, .., e_n\}$, such vector space valued one-forms can be expressed as

$$\omega = \sum_{a=1}^{n} \omega_a \, e^a \qquad \text{with } \omega^a \in \Omega^k(U) \;\; \forall a \in \{1, ..., n\}. \tag{1.81}$$

To define a *covariant derivative*, we use one-forms $\theta \in \Omega^1(U, \mathrm{End}(V))$ which take values in the vector space $\mathrm{End}(V)$ of endomorphisms of V. With respect to the basis B, such one-forms can be expressed as

$$\theta(e_a) = \sum_{b=1}^{n} \theta_a{}^b\, e_b \qquad \text{with } \theta_a{}^b \in \Omega^1(U) \;\; \forall a, b \in \{1, ..., n\}. \tag{1.82}$$

We define the *covariant derivative with respect to* θ as

$$d_\theta \omega_a = d\omega_a + \sum_{b=1}^{n} \theta_a{}^b \wedge \omega_b. \tag{1.83}$$

Show that $d_\theta^2 \omega_a = 0$ for all $\omega \in \Omega^k(U, V)$, $a = 1, ..., n$, if and only if θ is *flat*:

$$d\theta_a{}^b + \sum_{c=1}^{n} \theta_a{}^c \wedge \theta_c{}^b = 0. \tag{1.84}$$

Remark 1.4.24: Such generalised derivatives play an important role in gauge and field theories. The one-forms θ used to construct the generalised derivatives d_θ corresponds to gauge fields. One-forms $\theta \in \omega^1(U)$ correspond to abelian gauge theories such as electromagnetism. One-forms $\theta \in \Omega^1(U, \text{End}(V))$ with values in the set of endomorphisms $\text{End}(V)$ correspond to non-abelian gauge theories.

We will now investigate the interaction of vector fields and differential forms. As we have seen in the context of (multi)linear forms, we can interpret a bilinear form or $(0,2)$-tensor $g \in V^* \otimes V^*$ on a vector space V as a bilinear map $g : V \times V \to \mathbb{R}$ or as a linear map $g : V \to V^*$. In the context of vector fields and differential forms, we consider $(0,2)$-tensor fields on an open subset $U \subset \mathbb{R}^n$, which assign to every point $p \in U$ a bilinear form $\omega(p) : T_pU \times T_pU \to \mathbb{R}$. An example are the (pseudo-) Riemannian metrics in Definition 1.4.10, which correspond to *symmetric* non-degenerate (0,2)-tensor fields.

We will now investigate what structures are obtained from *anti-symmetric* non-degenerate (0,2)-tensor fields or 2-*forms* on U. One finds that every non-degenerate 2-form on U gives rise to a map $\text{Vec}(U) \times \text{Vec}(U) \to \mathcal{C}^\infty(U)$ and to an identification of $\text{Vec}(U)$ and $\Omega^1(U)$. If one combines these structures with the exterior differential of functions $f \in \mathcal{C}^\infty(U)$, one obtains a map $\mathcal{C}^\infty(U) \times \mathcal{C}^\infty(U) \to \mathcal{C}^\infty(U)$. If the 2-form is also closed, this map is a Poisson bracket.

Lemma 1.4.25: (Identification of vector fields and forms: the Poisson bracket)

Let $U \subset \mathbb{R}^n$ be open and $\omega \in \Omega^2(U)$ a closed, non-degenerate 2-form on U. Such a 2-form is also called a *symplectic form* on U. The non-degenerate 2-form ω induces an isomorphism between $\text{Vec}(U)$ and $\Omega^1(U)$

$$\Phi_\omega : \text{Vec}(U) \to \Omega^1(U) \quad \boldsymbol{X} \mapsto \omega(\boldsymbol{X}, \cdot). \tag{1.85}$$

It therefore gives rise to a map

$$X_\cdot : \mathcal{C}^\infty(U) \to \text{Vec}(U) \qquad\qquad f \mapsto \boldsymbol{X}_f = \Phi_\omega^{-1}(df) \tag{1.86}$$

and a map

$$\begin{aligned} \{\,,\,\} :\; & \mathcal{C}^\infty(U) \times \mathcal{C}^\infty(U) \to \mathcal{C}^\infty(U) \\ & (f, g) \mapsto \{f, g\} = df(\boldsymbol{X}_g) = -dg(\boldsymbol{X}_f) = \omega(\boldsymbol{X}_f, \boldsymbol{X}_g) \end{aligned} \tag{1.87}$$

The latter has the properties of a Poisson bracket.

1. It is *bilinear*: $\{tf + sg, h\} = t\{f, h\} + s\{g, h\}$ for all $f, g, h \in \mathcal{C}^\infty(U)$, $t, s \in \mathbb{R}$.
2. It is *antisymmetric*: $\{f, g\} = -\{g, f\}$ for all $f, g \in \mathcal{C}^\infty U$.
3. It satisfies the *Leibnitz identity* or, in other words, it is a *derivation*:
$$\{f \cdot g, h\} = f \cdot \{g, h\} + g \cdot \{f, h\} \qquad \forall f, g, h \in \mathcal{C}^\infty(U).$$
4. It satisfies the *Jacobi identity*:
$$\{\{f, g\}, h\} + \{\{h, f\}, g\} + \{\{g, h\}, f\} = 0 \qquad \forall f, g, h \in \mathcal{C}^\infty(U).$$

If U is star-shaped, there exists a one form θ such that $\omega = d\theta$. Such a one-form is called a *symplectic potential*.

Proof: The bilinearity and antisymmetry follow from the fact that all maps involved in the definition are linear and from the antisymmetry of the two-form ω. The Leibnitz identity follows from the identity $d(f \cdot g) = f \cdot dg + g \cdot df$. The Jacobi identity follows from the fact that ω is closed and the existence of the symplectic potential from Poincaré's lemma. □

Remark 1.4.26: The Poisson bracket is well known from classical mechanics. There it often arises from a *canonical symplectic form* associated with the action that characterises a physical system. The functions in $\mathcal{C}^\infty(U)$ correspond to the functions on phase space.

1.5 Electrodynamics and the theory of differential forms

An electrodynamical system is described by two fields:

- the *electrical field* $\vec{E} : \mathbb{R}^4 \to \mathbb{R}^3$ which assigns to each point $\vec{x}$ in space and each time t the value of the electrical field $\vec{E}(t, x) \in \mathbb{R}^3$ at $\vec{x}$ at time t .
- the *magnetic field* $\vec{B} : \mathbb{R}^4 \to \mathbb{R}^3$ which assigns to each point $\vec{x}$ in space and each time t the value $\vec{B}(t, x) \in \mathbb{R}^3$ of the magnetic field at $\vec{x}$ at time t.

together with two densities that describe the behaviour of the electrical charges:

- the *charge density* $\rho : \mathbb{R}^4 \to \mathbb{R}$ which assigns to each point $\vec{x}$ in space and each time t the density $\rho(t, \vec{x})$ of the electrical charge at $\vec{x}$ at time t.
- the *current density* $\vec{j} : \mathbb{R}^4 \to \mathbb{R}^3$ which assigns to each point $\vec{x}$ in space and each time t the density $\vec{j}(t, \vec{x})$ of the electrical current at $\vec{x}$ at time t.

The behaviour of the electromagnetic system is described by the *Maxwell equations.*

Definition 1.5.1: (Maxwell equations)

We consider an electrical field $\vec{E} : \mathbb{R}^4 \to \mathbb{R}^3$, a magnetic field $\vec{B} : \mathbb{R}^4 \to \mathbb{R}$ a charge density $\rho : \mathbb{R}^4 \to \mathbb{R}$ and a current density $\vec{j} : \mathbb{R}^4 \to \mathbb{R}^3$. In units in which the speed of light, the dielectric and the magnetic constant are one: $\epsilon_0 = \mu_0 = c = 1$, the Maxwell equations read

$$\mathrm{div}\vec{E} = \rho \qquad \mathrm{div}\vec{B} = 0 \tag{1.88}$$
$$\vec{\nabla}\times\vec{E} = -\partial_t\vec{B} \qquad \vec{\nabla}\times\vec{B} = \vec{j} + \partial_t\vec{E}. \tag{1.89}$$

The Maxwell equation $\mathrm{div}\,\vec{B} = 0$ is often paraphrased as *the absence of magnetic sources*, while $\vec{\nabla}\times\vec{E} = -\partial_t\vec{B}$ is known as *Faraday's law*. The equation $\mathrm{div}\vec{E} = \rho$ is often referred to as *Gauss' law* and $\vec{\nabla}\times\vec{B} = \vec{\jmath} + \partial_t\vec{E}$ as *Ampère's law*.

In addition to the Maxwell equations, we have the *charge conservation law*

$$\partial_t\rho + \mathrm{div}\,\vec{\jmath} = 0. \tag{1.90}$$

Remark 1.5.2: The Maxwell equations (1.88) can be viewed as *consistency conditions* on the electrical and magnetic field $\vec{E}(t,\cdot)$, $\vec{B}(t,\cdot)$ at a given time t. The charge conservation law describes the time evolution of the charge density, and the Maxwell equations (1.89) the time evolution of the electrical and magnetic field.

If the electrical field $\vec{E}(t,\cdot)$, the magnetic field $\vec{B}(t,\cdot)$ and the charge and current densities $\rho(t,\cdot), \vec{\jmath}(t,\cdot)$ are known for a given time t, the Maxwell equations uniquely determine their values for all $t' > t$.

We will now show that the Maxwell equations can be reformulated in terms of differential forms and that this formulation is useful for understanding their properties. We consider the vector space $\mathbb{R}^4$ with Basis $B = \{e_0, e_1, e_2, e_3\}$ and a symmetric non-degenerate bilinear form g, the *Minkowski metric*, characterised by

$$g(e_0, e_0) = -1 \qquad g(e_i, e_i) = 1 \text{ for } i = 1,2,3 \qquad g(e_\mu, e_\nu) = 0 \ \mu,\nu \in \{0,1,2,3\},\ \mu \neq \nu.$$

We denote by $*$ the associated Hodge star and by $\delta = * \circ d \circ *$ the associated co-differential. To keep the convention most widespread in physics, we will label indices with Greek letters if they take values $\mu, \nu, ... \in \{0,1,2,3\}$ and reserve Latin indices $i, j, k, ... \in \{1,2,3\}$ for the spatial indices. The time coordinate t therefore corresponds to x^0.

By combining the components of the electrical and magnetic field into a two-form and the current and charge density into a one-form, we obtain the following definition.

Definition 1.5.3: (Electrodynamics in formulation with forms)

The *field strength tensor* is the two-form $F \in \Omega^2(\mathbb{R}^4)$

$$F = -\sum_{i=1}^{3} E_i\, dx^0 \wedge dx^i + B_1\, dx^2 \wedge dx^3 + B_2\, dx^3 \wedge dx^1 + B_3\, dx^1 \wedge dx^2.$$

Its Hodge dual is the two-form $*F \in \Omega^2(\mathbb{R}^4)$, often referred to as *dual field strength tensor*

$$*F = -\sum_{i=1}^{3} B_i\, dx^0 \wedge dx^i - E_1\, dx^2 \wedge dx^3 - E_2\, dx^3 \wedge dx^1 - E_3\, dx^1 \wedge dx^2. \tag{1.91}$$

The *four-current* $J \in \Omega^1(\mathbb{R}^4)$ is the one-form

$$J = -\rho\, dx^0 + j_1\, dx^1 + j_2\, dx^2 + j_3\, dx^3. \tag{1.92}$$

with Hodge dual $*J \in \Omega^3(\mathbb{R}^4)$

$$* J = \rho\, dx^1 \wedge dx^2 \wedge dx^3 - j_1\, dx^0 \wedge dx^2 \wedge dx^3 - j_2\, dx^0 \wedge dx^3 \wedge dx^1 - j_3\, dx^0 \wedge dx^1 \wedge dx^2.$$

Remark 1.5.4: Note that the electrical fields

$$E = \sum_{i=1}^{n} E_i\, dx^i$$

and the magnetic field

$$B = \sum_{j,k=1}^{n} \epsilon_{ijk} B^i\, dx^j \wedge dx^k$$

enter the expression for the field strength tensor in a different way. This is related to the fact that the magnetic field B is a *pseudovector*, i.e. a two-form, while the electrical field E is a *vector* or one-form.

Lemma 1.5.5: (Maxwell equations and charge conservation)

In terms of field strength tensor and four-current, the Maxwell equations take the form

$$dF = 0 \qquad \delta F = *d*F = J. \tag{1.93}$$

The charge conservation law is given by

$$\delta J = *d*J = 0. \tag{1.94}$$

Proof:

We have

$$\begin{aligned} dF = &\left(\text{div}\, \vec{B}\right) dx^1 \wedge dx^2 \wedge dx^3 + (\partial_0 B_1 + \partial_2 E_3 - \partial_3 E_2)\, dx^0 \wedge dx^2 \wedge dx^3 \\ &+ (\partial_0 B_2 + \partial_3 E_1 - \partial_1 E_3)\, dx^0 \wedge dx^3 \wedge dx^1 + (\partial_0 B_3 + \partial_2 E_1 - \partial_1 E_2)\, dx^0 \wedge dx^1 \wedge dx^2. \end{aligned}$$

$dF = 0$ is therefore equivalent to *the absence of magnetic sources* and *Faraday's law*

$$dF = 0 \qquad \Leftrightarrow \qquad \text{div}\, \vec{B} = 0, \quad \partial_0 B + \vec{\nabla} \times \vec{E} = 0. \tag{1.95}$$

Similarly, we can determine the co-differential $\delta F = *d*F$. This yields

$$\begin{aligned} *d*F = &-\text{div}\, \vec{E}\, dx^0 + (-\partial_0 E_1 + \partial_2 B_3 - \partial_3 B_2)\, dx^1 + (-\partial_0 E_2 + \partial_3 B_1 - \partial_1 B_3)\, dx^2 \\ &+ (-\partial_0 E_3 + \partial_1 B_2 - \partial_2 B_1)\, dx^3. \end{aligned}$$

The equation $\delta F = *d*F = 0$ is therefore equivalent to *Gauss law* and *Ampère's law*

$$\delta F = *d*F = J \qquad \Leftrightarrow \qquad \text{div}\, \vec{E} = \rho, \qquad \vec{\nabla} \times \vec{B} - \partial_t \vec{E} = \vec{\jmath}. \tag{1.96}$$

Finally, we determine the co-differential δJ and find that the equation $\delta J = 0$ is equivalent to the *current conservation law*

$$\delta J = *d*J = 0 \qquad \Leftrightarrow \qquad \partial_0 \rho + \text{div}\, \vec{\jmath} = 0. \tag{1.97}$$

We will now use Poincaré's lemma to obtain the gauge potentials of electrodynamics. These are the *vector potential* $\vec{A} : \mathbb{R}^4 \to \mathbb{R}^3$ and the *scalar potential* $\phi : \mathbb{R}^4 \to \mathbb{R}$ from which the electrical and magnetic field are obtained as

$$\vec{B} = \vec{\nabla} \times \vec{A} \qquad \vec{E} = \text{grad}\, \phi - \partial_0 \vec{A}, \tag{1.98}$$

or, equivalently, in coordinates,

$$B_i = \epsilon_i{}^{jk} \partial_j A_k \qquad e_i = \partial_i \phi - \partial_0 A_i.$$

Lemma 1.5.6: (Poincaré's lemma ⇒ existence of gauge potentials)

As $dF = 0$ and $\mathbb{R}^4$ star-shaped, Poincaré's lemma implies the existence of a one-form $A \in \Omega^1(\mathbb{R}^4)$, the *gauge potential*, such that $F = dA$. With the definition

$$A = -\phi\, dx^0 + \sum_{i=1}^{3} A_i dx^i = -\phi\, dx^0 + A_1\, dx^1 + A_2\, dx^2 + A_3\, dx^3, \tag{1.99}$$

where ϕ is the scalar potential and A_i are the components of the vector potential, the equation $F = dA$ reproduces (1.98)

$$F = dA \quad \Leftrightarrow \vec{B} = \vec{\nabla}\times \vec{A} \qquad \vec{E} = \mathrm{grad}\,\phi - \partial_0 \vec{A}. \tag{1.100}$$

Remark 1.5.7: Note that the *gauge potential* A is defined only up to addition of an exact one-form $d\chi \in \Omega^1(\mathbb{R}^4)$, $\chi \in \Omega^0(\mathbb{R}^4)$. As $d^2 = 0$, we have: $d(A + d\chi) = dA + d^2\chi = dA$. A transformation $A \mapsto A+d\chi$ corresponds to a transformation of the scalar and vector potential

$$\phi \mapsto \phi - \partial_0\chi \qquad \vec{A} \mapsto \vec{A} + \mathrm{grad}\chi \tag{1.101}$$

and is called a *gauge transformation*. It can be used to bring the potentials $\vec{A}$ and ϕ into a form that is suited to the problem under consideration. This is called *fixing a gauge*.

Exercise 14: (Covariant formulation of electrodynamics in local coordinates)

We express the gauge potential, the field strength tensor and the four-current current using Einstein's summation convention

$$A = A_\mu dx^\mu \qquad F = \tfrac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu \qquad J = j_\mu dx^\mu, \tag{1.102}$$

where all indices $\mu, \nu, ...$ run from 0 to 3. We set $g_{\mu\nu} = g(e_\mu, e_\nu)$ and denote by $g^{\mu\nu}$ its inverse $g_{\mu\nu} g^{\nu\rho} = \delta_\mu^\rho$.

1. Show that in these coordinates the equation $F = dA$ takes the form

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\mu A_\nu \tag{1.103}$$

   and the equation $dF = 0$ reads

$$\partial_\mu F_{\nu\rho} + \partial_\rho F_{\mu\nu} + \partial_\nu F_{\rho\mu} = 0. \tag{1.104}$$

2. Let now $\epsilon^{\mu\nu\rho\sigma}$ be the totally antisymmetric tensor in four indices with the convention $\epsilon_{0123} = 1$. Note that the totally antisymmetric tensor with four indices satisfies

$$\epsilon_{\mu\nu\rho\sigma}\epsilon^{\mu\nu}{}_{\tau\kappa} = 2(g_{\rho\tau} g_{\sigma\kappa} - g_{\rho\kappa} g_{\sigma\tau}).$$

   We denote the components of $*F$ by $*F = \frac{1}{2}\mathcal{F}_{\mu\nu} dx^\mu \wedge dx^\nu$. Show that the components of $*F$ are given by

$$\mathcal{F}_{\rho\sigma} = \tfrac{1}{2}\epsilon_{\rho\sigma\tau\kappa} g^{\tau\mu} g^{\kappa\nu} F_{\mu\nu} = \tfrac{1}{2}\epsilon_{\rho\sigma\tau\kappa} F^{\tau\kappa} \qquad F^{\tau\kappa} = g^{\tau\mu} g^{\kappa\nu} F_{\mu\nu} \tag{1.105}$$

   and that the equation $\delta F = J$ reads

$$\partial^\rho F_{\rho\sigma} = j_\sigma \qquad \text{with } \partial^\rho = g^{\mu\rho}\partial_\mu. \tag{1.106}$$

3. Show that the charge conservation law takes the form

$$\partial^\sigma j_\sigma = 0 \qquad \text{with } \partial^\rho = g^{\mu\rho}\partial_\mu. \tag{1.107}$$

Exercise 15: (Gauge fixing)

1. *Coulomb gauge and static Maxwell equations*

   Show that one can use the gauge freedom $A \mapsto A + d\chi$ to impose the *Coulomb gauge* condition

$$\text{div}\,\vec{A} = 0. \tag{1.108}$$

   Show that the residual gauge freedom is $A \mapsto A + d\chi$, where χ satisfies the Laplace equation $\Delta\chi = \sum_{i=1}^3 \partial_i^2 \chi = 0$.

   We now consider the electrostatic and magnetostatic case of the Maxwell equations, i.e. a situation where the fields and the charge density are independent of time $\partial_0 \vec{E} = \partial_0 \vec{B} = 0$, $\partial_0\rho = 0$. Show that the Maxwell equations take the form

$$\text{div}\vec{E} = \rho \qquad \text{div}\vec{B} = 0 \tag{1.109}$$

$$\vec{\nabla}\times\vec{E} = 0 \qquad \vec{\nabla}\times\vec{B} = \vec{\jmath} \tag{1.110}$$

   and tat the charge conservation law becomes $\text{div}\,\vec{\jmath} = 0$. Show that if A satisfies the condition of the Coulomb gauge, the Maxwell equations can be re-formulated as

$$\Delta\phi = -\rho \qquad \Delta\vec{A} = \sum_{i=1}^{3} \partial_i^2 \vec{A} = -\vec{\jmath}. \tag{1.111}$$

   The potentials $\phi, \vec{A}$ are thus solutions of the *inhomogeneous Laplace equation.*

2. *Lorentz gauge and electromagnetic waves*

   Show that the gauge freedom $A \mapsto A + d\chi$ can also be used to impose the *Lorentz gauge*

$$\text{div}\,\vec{A} + \partial_0\phi = 0 \tag{1.112}$$

   and that the residual gauge freedom of this gauge are gauge transformations $A \mapsto A + d\chi$ where χ satisfies the *wave equation* $\Box\chi = \partial_0^2\chi - \Delta\chi = 0$.

   We now consider the Maxwell equations without sources, i.e. a situation with vanishing charge and current density $\rho = \vec{\jmath} = 0$, in which the Maxwell equations take the form

$$\text{div}\vec{E} = 0 \qquad \text{div}\vec{B} = 0 \tag{1.113}$$

$$\vec{\nabla}\times\vec{E} = -\partial_0\vec{B} \qquad \vec{\nabla}\times\vec{B} = \partial_0\vec{E}. \tag{1.114}$$

   Show that in the Lorentz gauge the Maxwell equations

$$\text{div}\vec{E} = 0 \qquad \vec{\nabla}\times\vec{B} = \partial_0\vec{E}. \tag{1.115}$$

   can be expressed as

$$\Box\phi = 0 \qquad \Box A = 0, \tag{1.116}$$

   while the other two Maxwell equations are satisfied automatically. The potentials $\phi, \vec{A}$ are therefore solutions of the *wave equation* - they correspond to electromagnetic waves, non-trivial electrical and magnetic fields in the absence of sources and currents.

   Hint: Prove and then use the identities

$$\vec{\nabla}\times\text{grad}F = 0 \qquad \vec{\nabla}\times(\vec{\nabla}\times\vec{G}) = \text{grad}(\text{div}\,\vec{G}) - \Box\vec{G}. \tag{1.117}$$

Chapter 2

Groups and Algebras

2.1 Groups, algebras and Lie algebras

Groups and algebras play a fundamental role in physics because they are the mathematical structures which describe *symmetries* and *observables* of physical systems.

Classical symmetries such as rotation or translation symmetry in classical mechanics and Lorentz symmetry in special relativity are described by groups. This is due to the fact the composition of two symmetries should yield another symmetry, that each symmetry should have an inverse and that there should be a neutral or trivial symmetry. This leads to the definition of a group.

One distinguishes "continuous" symmetries (such as rotations, Lorentz transformations or translations), which correspond to Lie groups, and discrete symmetries such as permutations or time reversal. The former can be differentiated, which leads to the concept of Lie algebras. Lie algebras correspond to infinitesimal symmetries of physical systems and are associated with additive quantum numbers in the quantum theory.

(Associative) algebras, often equipped with additional structures, describe observables of physical systems. Examples are the Poisson algebras which correspond to phase spaces of classical physical systems theory and algebras of observables in the quantum theory.

Definition 2.1.1: (Associative unital algebra, subalgebra, abelian algebra)

1. An (associative) algebra is a vector space A over $k = \mathbb{R}$ or $k = \mathbb{C}$ together with a bilinear map $\circ : A \times A \to A$, the *multiplication*, that is *associative* $(a \circ b) \circ c = a \circ (b \circ c)$ for all $a, b, c \in A$. If there exists an element e such that $e \circ a = a \circ e = a \; \forall a \in A$, this element is called the *unit* and A is called a *unital algebra.*

2. A (unital) subalgebra of a unital associative algebra A is a linear subspace $B \subset A$ that is a (unital) algebra, i.e. that is closed under the multiplication: $b \circ b' \in B \; \forall b, b' \in B$ (and that contains the unit: $e \in B$). An algebra is called *abelian* or *commutative* if $a \circ b = b \circ a$ for all $a, b \in A$.

Remark 2.1.2: If a unit exists, it is unique.

Example 2.1.3: (Associative algebras)

1. The endomorphisms $\mathrm{End}(V)$ of a vector space V with the composition $\circ$ form an associative algebra.

2. The smooth functions on a submanifold of $\mathbb{R}^n$ form an associative unital algebra $\mathcal{C}^\infty(M)$ with pointwise multiplication, pointwise addition and multiplication with $\mathbb{R}$. This algebra is abelian.

3. The tensor algebra $T(V)$ and the exterior algebra $\Lambda(V)$ of a finite-dimensional vector space V are associative algebras with multiplication given by respectively, the tensor product and the wedge product.

4. The matrices $M(n,k)$ with the usual addition, multiplication by k and matrix multiplication form an associative unital algebra.

Definition 2.1.4: (Group)

A group is a set G with an operation $\circ : G \times G \to G$ such that

1. The group multiplication $\circ$ is associative $(u \circ v) \circ w = u \circ (v \circ w)$ for all $u, v, w \in G$.

2. There exists an element e, the *unit*, such that $u \circ e = e \circ u = u$ for all $u \in G$.

3. Any element $g \in G$ has an inverse g^{-1} satisfying $g \circ g^{-1} = g^{-1} \circ g = e$.

A subset $H \subset G$ is called a *subgroup* of G if it is a group, i.e. if $e \in H$, $h \circ h' \in H$ for all $h, h' \in H$ and $h^{-1} \in H$ for all $h \in H$.

Example 2.1.5: (Groups)

1. The permutations $S(k)$ of the set $\{1, ..., k\}$ form a group.

2. For any (finite dimensional) vector space V, the set $\mathrm{Aut}(V)$ forms a group with the composition as group multiplication.

3. $(\mathbb{Z}, +)$ and $(\mathbb{R} \setminus \{0\}, \cdot)$ are groups.

4. The set $GL(n,k)$ of *invertible* $n \times n$ matrices over k with the matrix multiplication is a group.

5. The set $SL(2,\mathbb{Z}) = \{M \in M(2,\mathbb{Z}) \mid \det M = 1\}$ of 2×2 matrices with integer entries and determinant 1 forms a group with respect to the matrix multiplication. This is called the *modular group*.

Definition 2.1.6: (Lie algebra, Lie subalgebra)

A *Lie algebra* $\mathfrak{g}$ is a vector space $\mathfrak{g}$ over k together with a bilinear map $[\,,\,] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$, the *Lie bracket*, that

1. is *antisymmetric*: $[\boldsymbol{x}, \boldsymbol{y}] = -[\boldsymbol{y}, \boldsymbol{x}]$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathfrak{g}$.

2. satisfies the *Jacobi identity*: $[[\boldsymbol{x}, \boldsymbol{y}], \boldsymbol{z}] + [[\boldsymbol{z}, \boldsymbol{x}], \boldsymbol{y}] + [[\boldsymbol{y}, \boldsymbol{z}], \boldsymbol{x}] = 0$ for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathfrak{g}$.

The coefficients of $[\,,\,]$ with respect to a basis $B = \{e_1, ..., e_n\}$ of $\mathfrak{g}$ are called the *structure constants* of $\mathfrak{g}$

$$[e_i, e_j] = \sum_{k=1}^{n} f_{ij}{}^{k} e_k. \tag{2.1}$$

A *Lie subalgebra* of $\mathfrak{g}$ is a linear subspace $\mathfrak{k} \subset \mathfrak{g}$ which is closed under the Lie bracket, i.e. $[\boldsymbol{x}, \boldsymbol{y}] \in \mathfrak{k}$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathfrak{k}$. A Lie algebra is called *abelian* if its Lie bracket vanishes $[\boldsymbol{x}, \boldsymbol{y}] = 0$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathfrak{g}$.

Example 2.1.7: (Lie algebras)

1. For a submanifold $M \subset \mathbb{R}^n$, the set $\mathcal{C}^\infty(M)$ of smooth functions on M with pointwise addition, multiplication by $\mathbb{R}$ and a Poisson bracket is an (infinite dimensional) Lie algebra.
2. Every vector space V can be made into a Lie algebra if it is given the *trivial Lie algebra structure* $[\boldsymbol{v}, \boldsymbol{w}] = 0$ for all $\boldsymbol{v}, \boldsymbol{w} \in V$.
3. The vector space $\mathbb{R}^3$ with the wedge product as Lie bracket $[\,,\,] = \wedge$, $(\boldsymbol{x} \wedge \boldsymbol{y})^a = \epsilon^a{}_{bc} x^b y^c$ is a Lie algebra.
4. Any associative algebra becomes a Lie algebra when equipped with the commutator as a Lie bracket $[a, b] = a \circ b - b \circ a$ for $a, b \in A$. In particular, this implies that the set of matrices $M(n, \mathbb{R})$ and $M(n, \mathbb{C})$ with the usual matrix commutator are Lie algebras. These Lie algebras are called $\mathfrak{gl}(n, \mathbb{R})$ and $\mathfrak{gl}(n, \mathbb{C})$.
5. The subset $\mathfrak{sl}(n, \mathbb{C}) = \{M \in M(n, k) \mid \mathrm{Tr}\,(M) = \sum_{i=1}^{n} m_{ii} = 0\} \subset \mathfrak{gl}(n, k)$ of matrices with vanishing trace form a Lie subalgebra of $M(n, k)$.

Exercise 16: Show that the examples given in Examples 2.1.3, 2.1.5, 2.1.7 are associative unital algebras, groups and Lie algebras. Think of other examples of groups, algebras and Lie algebras.

Exercise 17: (Structure constants)

Show that a bilinear map $\Phi : V \times V \to V$ defines a Lie bracket on V if and only its coefficients with respect to a basis $B = \{e_1, ..., e_n\}$ satisfy

$$\phi(e_i, e_j) = \sum_{k=1}^{n} \phi_{ij}{}^{k} e_k \qquad \phi_{ij}{}^{k} = -\phi_{ji}{}^{k}, \quad \sum_{k=1}^{n} (\phi_{ij}{}^{k} \phi_{kl}{}^{h} + \phi_{li}{}^{k} \phi_{kj}{}^{h} + \phi_{jl}{}^{k} \phi_{ki}{}^{h}) = 0 \tag{2.2}$$

Lemma 2.1.8: (Vector fields $\mathrm{Vec}(U)$ form a Lie algebra)

Let $U \subset \mathbb{R}^n$ be an open subset and $\mathrm{Vec}(U)$ the vector space of vector fields on U. The *Lie bracket* of two smooth vector fields $\boldsymbol{X} = \sum_{i=1}^{n} x^i \partial_i$, $\boldsymbol{Y} = \sum_{i=1}^{n} y^i \partial_i$ with $x^i, y^i \in \mathcal{C}^\infty(U)$ is the vector field

$$[\boldsymbol{X}, \boldsymbol{Y}] = \sum_{i,j=1}^{n} (x^j \partial_j y^i - y^j \partial_j x^i) \partial_i. \tag{2.3}$$

It satisfies

$$\boldsymbol{X}.\boldsymbol{Y}.f - \boldsymbol{Y}.\boldsymbol{X}.f = [\boldsymbol{X}, \boldsymbol{Y}].f \qquad \forall f \in \mathcal{C}^\infty(U), \boldsymbol{X}, \boldsymbol{Y} \in \mathrm{Vec}(U). \tag{2.4}$$

With this Lie bracket, the vector space $\mathrm{Vec}(U)$ becomes a Lie algebra.

Proof: It is clear from (2.3) that the bracket $[\,,\,] : \mathrm{Vec}(U) \times \mathrm{Vec}(U) \to \mathrm{Vec}(U)$ is bilinear and antisymmetric. Identity (2.4) is obtained by direct calculation

$$\boldsymbol{X}.\boldsymbol{Y}.f - \boldsymbol{Y}.\boldsymbol{X}.f = \sum_{j=1}^n x^j \partial_j \left(\sum_{i=1}^n y^i \partial_i f \right) - \sum_{j=1}^n y^j \partial_j \left(\sum_{i=1}^n x^i \partial_i f \right) = \sum_{i,j=1}^n (x^j \partial_i y^i - y^j \partial_j x^i) \partial_i f.$$

The Jacobi identity follows by direct calculation from (2.3) or (2.4)

$$\begin{aligned}
[[\boldsymbol{X},\boldsymbol{Y}],\boldsymbol{Z}]f &= \boldsymbol{X}.\boldsymbol{Y}.\boldsymbol{Z}.f - \boldsymbol{Y}.\boldsymbol{X}.\boldsymbol{Z}.f - \boldsymbol{Z}.\boldsymbol{X}.\boldsymbol{Y}.f + \boldsymbol{Z}.\boldsymbol{Y}.\boldsymbol{X}.f \\
[[\boldsymbol{Z},\boldsymbol{X}],\boldsymbol{Y}]f &= \boldsymbol{Z}.\boldsymbol{X}.\boldsymbol{Y}.f - \boldsymbol{X}.\boldsymbol{Z}.\boldsymbol{Y}.f - \boldsymbol{Y}.\boldsymbol{Z}.\boldsymbol{X}.f + \boldsymbol{Y}.\boldsymbol{X}.\boldsymbol{Z}.f \\
[[\boldsymbol{Y},\boldsymbol{Z}],\boldsymbol{X}]f &= \boldsymbol{Y}.\boldsymbol{Z}.\boldsymbol{X}.f - \boldsymbol{Z}.\boldsymbol{Y}.\boldsymbol{X}.f - \boldsymbol{X}.\boldsymbol{Y}.\boldsymbol{Z}.f + \boldsymbol{X}.\boldsymbol{Z}.\boldsymbol{Y}.f
\end{aligned} \tag{2.5}$$

□

Definition 2.1.9: (Homomorphisms, endomorphisms, isomorphisms and automorphisms)

- A *Homomorphism of (associative) algebras* A, B is a linear map $\Phi \in \mathrm{Hom}(A, B)$ that is compatible with the multiplication: for all $a, a' \in A$: $\phi(a \circ a') = \phi(a) \circ \phi(a')$. A homomorphism of algebras is called, respectively, an *endomorphism*, *isomorphism* or *automorphism of algebras* if it is an endomorphism, isomorphism and automorphism of vector spaces. If there exists an algebra isomorphism between algebras A and B, A and B are called isomorphic. The vector space of algebra endomorphisms of an algebra A is a subalgebra of the algebra $\mathrm{End}(A)$ (with composition $\circ$ as multiplication).
- A *homomorphism of Lie algebras* $\mathfrak{h}, \mathfrak{k}$ is a linear map $\phi \in \mathrm{Hom}(\mathfrak{h}, \mathfrak{k})$ that is compatible with the Lie brackets: $[\phi(\boldsymbol{x}), \phi(\boldsymbol{y})] = \phi([\boldsymbol{x}, \boldsymbol{y}])$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathfrak{h}$. It is called an an *endomorphism*, *isomorphism* or *automorphism of Lie algebras* if it is an endomorphism, isomorphism and automorphism of vector spaces. If there exists an isomorphism between Lie algebras $\mathfrak{h}$ and $\mathfrak{k}$, $\mathfrak{h}$ and $\mathfrak{k}$ are called isomorphic. The vector space of Lie algebra endomorphisms of a Lie algebra $\mathfrak{h}$ is Lie subalgebra $\mathrm{End}(\mathfrak{h})$ with the commutator as the Lie bracket.
- A *group homomorphism* is a map $\rho : G \to H$ between groups G, H that is compatible with the group multiplication: $\rho(u \cdot v) = \rho(u) \cdot \rho(v)\ \forall u, v \in G$. It is called an *endomorphism* if $H = G$, an *isomorphism* if $\phi(u) = \phi(v)$ implies $u = v$ and an *automorphism* if it is an isomorphism and $G = H$. If there exists a group isomorphism between groups G and H, G and H are called isomorphic. The set of group automorphisms of a group G forms a group with group multiplication given by composition and id_G as unit.

Lemma 2.1.10: (Adjoint action)

1. For any group G and any $g \in G$ the map $G_g : G \to G$, $h \mapsto g \circ h \circ g^{-1}$ is a group automorphism. C_g is called the *adjoint action of* G *on itself*. The map $C : G \to \mathrm{Aut}(G)$, $g \mapsto C_g$ is a group homomorphism from G to the group of group automorphisms of G.
2. For a Lie algebra $\mathfrak{g}$ and any $\boldsymbol{x} \in \mathfrak{g}$, the map $\mathrm{ad}_{\boldsymbol{x}} : \mathfrak{g} \to \mathfrak{g}$, $\boldsymbol{y} \mapsto [\boldsymbol{x}, \boldsymbol{y}]$ is a Lie algebra endomorphism. This Lie algebra endomorphism is called the *adjoint action of* $\mathfrak{g}$ *on itself*. In terms of the structure constants, it is given by

$$\mathrm{ad}_{e_i}(e_j) = \sum_{k=1}^n f_{ij}^k e_k.$$

The map $\mathfrak{g} \to \text{End}(\mathfrak{g})$, $\boldsymbol{x} \mapsto \text{ad}_{\boldsymbol{x}}$ is a Lie algebra homomorphism from $\mathfrak{g}$ to the Lie algebra of Lie algebra endomorphisms of $\mathfrak{g}$.

2.2 Lie algebras and matrix Lie groups

An important role in physics is played by the so-called matrix (Lie) groups which we now define.

Definition 2.2.1: (Matrix Lie group)

1. A *matrix Lie group*, or *classical Lie group*, G is a closed subgroup of the group $GL(n,k)$ of invertible matrices with respect to the matrix multiplication.
2. The *tangent space* T_gG at a point g in a matrix Lie group G is the subvector space of $M(n,k)$ defined as

$$T_gG = \{c'(0) \in M(n,k) \mid c :]-\epsilon, \epsilon[\to G \text{ differentiable with } c(0) = g\}.$$

Theorem 2.2.2: The tangent space $\mathfrak{g} = T_eG$ to a matrix Lie group at the identity element is a Lie algebra with Lie bracket

$$[\boldsymbol{x}, \boldsymbol{y}] = \frac{d^2}{dsdt}\bigg|_{s=t=0} c_x(t)c_y(s)(c_x(t))^{-1} = \boldsymbol{x} \cdot \boldsymbol{y} - \boldsymbol{y} \cdot \boldsymbol{x} \tag{2.6}$$

where $\boldsymbol{x}, \boldsymbol{y} \in T_eG$ and $c_x :]-\epsilon, \epsilon[\to G$, $c_y :]-\epsilon, \epsilon[\to G$ are smooth curves in G with $c_x(0) = c_y(0) = e$ and derivatives $\dot{c}_x(0) = \boldsymbol{x}$, $\dot{c}_y(0) = \boldsymbol{y}$.

Proof:

1. For all $g \in G$, the map $C_g \circ c_y :]-\epsilon, \epsilon[\to G$, $s \mapsto g \cdot c_y(s) \cdot g^{-1}$ defines a path in G with $C_g \circ c_y(0) = e$, $C_g \circ \dot{c}_y(0) \in T_eG$. For each $g \in G$, we obtain a linear map

$$\text{Ad}_g : T_eG \to T_eG \qquad \boldsymbol{x} \mapsto \tfrac{d}{dt}|_{t=0} C_g \circ c_{\boldsymbol{x}}(t) \qquad \forall \boldsymbol{x} \in \mathfrak{g}, g \in G. \tag{2.7}$$

2. This implies that for all $t \in \mathbb{R}$ $\frac{d}{ds}|_{s=0} c_x(t) \cdot c_y(s) \cdot c_x(t)^{-1} \in T_eG$. As c_x is a smooth curve in G, the derivative

$$\frac{d}{dt}\bigg|_{t=0} c_x(t)\dot{c}_y(0)c_x(t)^{-1} = \frac{d^2}{dsdt}\bigg|_{t=s=0} c_x(t)c_y(s)(c_x(t))^{-1} \tag{2.8}$$

is again an element of $\mathfrak{g} = T_eG$. Using the product rule, we find

$$\frac{d^2}{dsdt}\bigg|_{t=s=0} c_x(t)c_y(s)(c_x(t))^{-1} = \boldsymbol{x} \cdot \boldsymbol{y} - \boldsymbol{y} \cdot \boldsymbol{x}. \tag{2.9}$$

The third example in Lemma 2.1.7 states that $\frac{d^2}{dsdt}|_{t=0}\, c_x(t)c_y(s)(c_x(t))^{-1}$ has the properties of a Lie bracket. □

Corollary 2.2.3: The map $\text{Ad} : G \to \text{Aut}(\mathfrak{g})$, $g \mapsto \text{Ad}_g$ where $\text{Ad}_g(\boldsymbol{x}) = g\boldsymbol{x}g^{-1}$ for all $g \in G$, $\boldsymbol{x} \in \mathfrak{g}$ defines a group homomorphism $G \to \text{Aut}(\mathfrak{g})$, the *adjoint action of* G *on its Lie algebra* $\mathfrak{g}$.

Example 2.2.4: ($SO(3)$ and $\mathfrak{so}(3)$)

We consider the rotation group $SO(3)$, i.e. the group of 3×3 matrices M with real entries that satisfy $M^T = M^{-1}$ and $\det M = 1$. This is a submanifold of $M(3, \mathbb{R})$. Any rotation matrix can be expressed as a product of the three matrices

$$c_1(s) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos s & \sin s \\ 0 & -\sin s & \cos s \end{pmatrix} \quad c_2(t) = \begin{pmatrix} \cos t & 0 & -\sin t \\ 0 & 1 & 0 \\ \sin t & 0 & \cos t \end{pmatrix} \quad c_3(u) = \begin{pmatrix} \cos u & -\sin u & 0 \\ \sin u & \cos u & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

with $t, s, u \in [-\pi, \pi[$. When viewed as functions of the parameters t, s, u these matrices define paths in $SO(3)$ with $c_1(0) = c_2(0) = c_3(0) = e$. Their derivatives at $t = 0$, $s = 0$, $u = 0$ are

$$J_1 = \dot{c}_1(0) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \quad J_2 = \dot{c}_2(0) = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad J_3 = \dot{c}_3(0) = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Their Lie brackets are

$$[J_1, J_2] = J_1 J_2 - J_2 J_1 = J_3 \qquad [J_3, J_1] = J_3 J_1 - J_1 J_3 = J_2 \qquad [J_2, J_3] = J_2 J_3 - J_3 J_2 = J_1.$$

We find that the Lie algebra $T_e SO(3)$ is the Lie algebra $\mathfrak{so}(3)$ of anti-symmetric, traceless 3×3 matrices. The elements J_1, J_2, J_3 form a basis of $\mathfrak{so}(3)$.

Exercise 18: ($SU(2)$ and $\mathfrak{su}(2)$)

1. Consider the group $SU(2)$ of unitary 2×2 matrices with unit determinant, i.e. matrices $M \in M(2, \mathbb{C})$ satisfying $M^\dagger = \bar{M}^T = M^{-1}$, $\det(M) = 1$. Show that any element of $SU(2)$ can be expressed as

$$M = \begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix} \qquad a, b \in \mathbb{C},\ |a|^2 + |b|^2 = 1. \tag{2.10}$$

Show that any matrix of the form (2.10) can be parametrised as

$$M = \begin{pmatrix} \cos\alpha e^{i\beta} & \sin\alpha e^{i(\beta+\gamma)} \\ -\sin\alpha e^{-i(\beta+\gamma)} & \cos\alpha e^{-i\beta} \end{pmatrix} \qquad \alpha, \beta, \gamma \in [-\pi, \pi[. \tag{2.11}$$

2. Show that if $c :]-\epsilon, \epsilon[\to M(2, \mathbb{C})$

$$c(t) = \begin{pmatrix} a(t) & b(t) \\ -\bar{b}(t) & \bar{a}(t) \end{pmatrix} \qquad a, b :]-\epsilon, \epsilon[\to \mathbb{C} \text{ smooth}$$

defines a smooth path in $SU(2)$ with $c(0) = e$, we have

$$a(0) = 1 \quad b(0) = 0 \qquad \mathrm{Re}(\dot{a}(0)) = 0.$$

Conclude that the Lie algebra $\mathfrak{su}(2)$ is the set of anti-hermitian, traceless matrices in $M(2, \mathbb{C})$. Show that they form a *real* linear subspace of $M(2, \mathbb{C})$.

3. Show that the following define smooth paths $M_i :]-\pi, \pi[\to SU(2)$ with $M_i(0) = e$

$$M_1(s) = \begin{pmatrix} \cos s & \sin s \\ -\sin s & \cos s \end{pmatrix} \quad M_2(u) = \begin{pmatrix} \cos u & i\sin u \\ i\sin u & \cos u \end{pmatrix} \qquad M_3(t) = \begin{pmatrix} e^{it} & 0 \\ 0 & e^{-it} \end{pmatrix}.$$

Show that the derivatives of these paths are given by

$$K_1 = \dot{M}_1(0) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad K_2 = \dot{M}_2(0) = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \qquad K_3 = \dot{M}_3(0) = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$$

and that they form a basis of the set of traceless, anti-hermitian matrices in $M(2,\mathbb{C})$. Show that their Lie bracket is given by

$$[K_1, K_2] = K_3 \qquad [K_3, K_1] = K_2 \qquad [K_2, K_3] = K_1. \tag{2.12}$$

3. Define a linear map $\Phi : \mathfrak{su}(2) \to \mathfrak{so}(3)$ by setting $\Phi(K_1) = J_1$, $\Phi(K_2) = J_2$, $\Phi(K_3) = J_3$ where J_1, J_2, J_3 are the matrices in Example 2.2.4. Show that this map is a Lie algebra isomorphism and therefore $\mathfrak{so}(3) \cong \mathfrak{su}(2)$ as Lie algebras.

Remark: As they can be expressed as Lie algebras with structure constants in $\mathbb{R}$, both $\mathfrak{so}(3)$ and $\mathfrak{su}(2)$ are Lie algebras over $\mathbb{R}$. Note that this does *not* mean that the matrices corresponding to these elements are matrices with entries in $\mathbb{R}$. The anti-hermitian matrices are only a real liner subspace, but not a complex linear subspace of $M(n,\mathbb{C})$.

Remark: As a submanifold of $\mathbb{R}^4 \cong \mathbb{C}^2$, the group $SU(2)$ is diffeomorphic to the three-sphere S^3. The identification is given by

$$M = \begin{pmatrix} \cos\alpha e^{i\beta} & \sin\alpha e^{i(\beta+\gamma)} \\ -\sin\alpha e^{-i(\beta+\gamma)} & \cos\alpha e^{-i\beta} \end{pmatrix} \mapsto (\cos\alpha, \sin\alpha\cos\beta, \sin\alpha\sin\beta\cos\gamma, \sin\alpha\sin\beta\sin\gamma)$$

$\alpha, \beta \in [-\pi, \pi[, \gamma \in [0, 2\pi[$.

Theorem 2.2.5: (Exponential map)

Let $\mathfrak{g}$ be a Lie algebra that is a Lie subalgebra of the Lie algebra $\mathfrak{gl}(n,k)$. Then, the series

$$\exp(M) = \sum_{k=0}^{\infty} \frac{M^k}{k!} \tag{2.13}$$

converges absolutely for all $M \in M(n,k)$ and hence defines a map $\exp : \mathfrak{g} \to M(n,k)$. The image $\exp(\mathfrak{g})$ is a matrix Lie group.

Proof:

1. We first need to show that the exponential map $\exp : M(n,k) \to M(n,k)$ is well-defined, i.e. that the series converges absolutely. For this, we consider the norm $||M|| = \sqrt{\mathrm{Tr}\,(M^\dagger M)} = \sqrt{\sum_{i,j=1}^n |m_{ij}|^2}$ introduced in Exercise 1.1.15. Using the Cauchy Schwartz inequality $|\mathrm{Tr}\left(A^\dagger B\right)| \leq ||A||||B||$, one can show by induction that $||M^k|| \leq ||M||^k$ for all $k \in \mathbb{N}$. This implies

$$\sum_{k=0}^{n} \frac{M^k}{k!} \leq \sum_{k=0}^{n} \frac{||M||^k}{k!} \tag{2.14}$$

As $\sum_{n=0}^{\infty} \frac{x^k}{k!}$ converges for all $x \in \mathbb{R}$, the series $\sum_{k=0}^{n} \frac{M^k}{k!}$ converges absolutely, and the exponential map for matrices is well-defined.

2. We now need to show that the image of exp has the structure of a group. For this, we note that $\exp(0) = 1_n$, where 1_n is the $n \times n$ unit matrix. Furthermore, we have

$$\exp(M) \cdot \exp(-M) = \sum_{k,l=0}^{\infty} (-1)^l \frac{M^k}{k!} \frac{M^l}{l!} \tag{2.15}$$

After a change of variables $s = k + l$ and reordering the sum (which can be done because it converges absolutely), we obtain

$$\exp(M) \exp(-M) = \sum_{s=0}^{\infty} \frac{M^s}{s!} \left(\sum_{l=0}^{s} \frac{s!}{l!(s-l)!} (-1)^l \right) = \sum_{s=0}^{\infty} \frac{M^s}{s!} \left(\sum_{l=0}^{s} (-1)^l \begin{pmatrix} s \\ l \end{pmatrix} \right) \tag{2.16}$$

Using the binomial formula $(x+y)^s = \sum_{l=0}^{s} \begin{pmatrix} s \\ l \end{pmatrix} x^l y^{s-l}$, we find

$$\exp(M) \cdot \exp(-M) = \sum_{s=0}^{\infty} \frac{M^s}{s!} (1 + (-1))^s = 1_n. \tag{2.17}$$

This proves the existence of inverses. It remains to show that the image of exp is closed under the multiplication, i.e. if $g = \exp(M)$, $h = \exp(N)$ there exists a $P \in \mathfrak{g}$ such that $g \cdot h = \exp(P)$. This can be shown using the formula of Baker Campbell and Hausdorff, who showed that such a P exists and gave an explicit expression for P in terms of M and N

$$P = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \sum_{\substack{k_1, j_1, ..., k_n, j_n = 0 \\ k_i + j_i > 0 \ \forall i = 1,...,n}}^{\infty} \frac{1}{\left(\sum_{l=1}^{n} (k_l + j_l)\right) k_1! j_1! \cdots k_n! j_n!} [M^{k_1} N^{j_1} M^{k_2} N^{j_2} \cdots M^{k_n} N^{j_n}]$$

$$[M^{k_1} N^{j_1} M^{k_2} N^{j_2} \cdots M^{k_n} N^{j_n}] := \underbrace{[M, [M, ..., [M}_{k_1 \times}, \underbrace{[N, [N, ..., [N}_{j_1 \times}, ... \underbrace{[M, [M, ...[M}_{k_n \times}, \underbrace{[N, [N, ...N]]]...]}_{j_n \times}.$$

Note that this multiple commutator vanishes f $j_n > 1$ or $j_n = 0$ and $k_n > 1$. □

Remark 2.2.6: If $\mathfrak{g} \subset M(n, k)$ is a Lie subalgebra, the Lie algebra of the Lie group $G = \exp(\mathfrak{g})$ is $T_e G = \mathfrak{g}$. This follows from the fact that for all $g \in \exp(\mathfrak{g})$, there exists a $M \in \mathfrak{g}$ such that $g = \exp(M)$ and the paths $c_M :]-\epsilon, \epsilon[\to G$, $c_M(t) = \exp(tM)$ satisfy $c_M(0) = e$, $\dot{c}_M(0) = M$. However, given a matrix Lie group $G \subset M(n, k)$ with Lie algebra $\mathfrak{g} = T_e G$, the exponential map $\exp : \mathfrak{g} \to G$ is not necessarily surjective. In general, its image is either G or a proper *subgroup* of G.

Example 2.2.7: ($\mathfrak{sl}(n, k)$ and $SL(n, k)$)

We consider the set $\mathfrak{sl}(n, k)$ of matrices with vanishing trace and entries in k. We claim that the image of this Lie algebra under the exponential map $\exp : M(n, K) \to GL(n, k)$ is the subgroup $SL(n, k) = \{g \in GL(n, k) \mid \det(M) = 1\}$. To show this, we prove the identity

$$\det(\exp(K)) = e^{\mathrm{Tr}(K)} \qquad \forall K \in M(n, k). \tag{2.18}$$

We set $M = tK$, $t \in \mathbb{R}$, $K = (k_{ij})_{i,j=1,...,n} \in M(n, k)$ and note that

$$\frac{d}{dt} e^{\mathrm{Tr}(tK)} = \mathrm{Tr}\,(K)\, e^{\mathrm{Tr}(K)}.$$

Using the formula for the determinant, we obtain

$$\begin{aligned}\frac{d}{dt}\det e^{tK} =&\frac{d}{dt}\sum_{\sigma\in S_n}(-1)^{\mathrm{sgn}(\sigma)}(e^{tK})_{1\sigma(1)}\cdots(e^{tK})_{n\sigma(n)}\\ =&\sum_{l,m=1}^{n}\sum_{\sigma\in S_n}\frac{k_{lm}(e^{tK})_{m\sigma(l)}}{(e^{tK})_{l\sigma(l)}}(-1)^{\mathrm{sgn}(\sigma)}(e^{tK})_{1\sigma(1)}\cdots(e^{tK})_{n\sigma(n)}.\end{aligned}$$

From the antisymmetry of the determinant it follows that all terms except the ones with $m = l$ vanish. This yields

$$\frac{d}{dt}\det e^{tK} = \sum_{l=1}^{n}\sum_{\sigma\in S_n}\frac{k_{ll}(e^{tK})_{l\sigma(l)}}{(e^{tK})_{l\sigma(l)}}(-1)^{\mathrm{sgn}(\sigma)}(e^{tK})_{1\sigma(1)}\cdots(e^{tK})_{n\sigma(n)} = \det e^{tK}\,\mathrm{Tr}\,(K).$$

Hence, we have

$$\frac{d}{dt}\ln\left(\det e^{tK}\right) = \frac{\mathrm{Tr}\,(K)\cdot\det e^{tK}}{\det e^{tK}} = \mathrm{Tr}\,(K) = \frac{\mathrm{Tr}\,(K)\cdot e^{t\mathrm{Tr}(K)}}{e^{t\mathrm{Tr}(K)}} = \frac{d}{dt}\ln\left(e^{t\mathrm{Tr}(K)}\right).$$

This implies $\det e^{tK} = C\cdot e^{t\mathrm{Tr}(K)}$, where $C\in\mathbb{C}$ constant. As we have $\det e^0 = \det(1_n) = 1 = e^0$, the contant is $c = 1$, which proves identity (2.18).

Using (2.18), one then finds that the image of a traceless matrix under the exponential map is a matrix with unit determinant. This implies that $\exp(\mathfrak{sl}(n,k))$ is either $SL(n,k)$ a proper subgroup of $SL(n,k)$. To show that it is the whole group, one can select a set of multiplicative generators of $SL(n,k)$, i.e. a set of matrices such that any element of $SL(n,k)$ can be expressed as a product in these generators and show that these generators can be obtained via the exponential map. (See Exercise 19) □.

Exercise 19: ($M(n,\mathbb{R})$ and $GL(n,\mathbb{R})$)

We consider the set $GL(n,\mathbb{R})$ of *invertible* matrices in $M(n,\mathbb{R})$, i.e. matrices $M\in M(n,\mathbb{R})$ with $\det M\neq 0$. As this is an open subset of $M(n,\mathbb{R})$, it is a submanifold of $M(n,\mathbb{R})$ and therefore a Lie group.

1. To determine $T_eGL(n,\mathbb{R})$, we consider the matrices E_{ij} whose only non-zero entry is a 1 in row i and column j: $(E_{ij})_{kl} = \delta_{i,k}\delta_{j,l}$. Show that for $i\neq j$, $M_{ij}(t) = 1_n + tE_{ij}$ is invertible with $M_{ij}^{-1}(t) = 1 - tE_{ij}$. Show that for all $i = 1,...,n$, $M_{ii}(t) = 1_n + (e^t - 1)E_{ii}$ is invertible with $M_{ii}^{-1}(t) = 1_n + (e^{-t} - 1)E_{ii}$. Conclude that the matrices M_{ij} define paths $M_{ij} :]-\epsilon,\epsilon[\to GL(n,\mathbb{R})$ with $M_{ij}(0) = 1_n$, $\dot{M}_{ij}(0) = E_{ij}$ for all $i,j = 1,...,n$. Show that the elements $\dot{M}_{ij}(0) = E_{ij}$ form a basis of $M(n,\mathbb{R})$ and therefore $T_eG = M(n,\mathbb{R})$.

Hint: Show first that $E_{ij}^2 = 0$ for $i\neq j$, $E_{ii}^2 = E_{ii}$.

2. We now consider the exponential map $\exp: M(n,\mathbb{R})\to M(n,\mathbb{R})$. Use the identity

$$\det(e^M) = \exp(\mathrm{Tr}\,(M))$$

to conclude that $\exp(M(n,\mathbb{R}))$ is a subgroup of the group of invertible matrices $M\in GL(n,k)$ with $det(M) > 0$. Although $T_eGL(n,\mathbb{R}) = M(n,\mathbb{R})$, the image of the exponential map $\exp: M(n,\mathbb{R})\to M(n,\mathbb{R})$ is therefore *not* $GL(n,\mathbb{R})$ but only a subgroup of it.

Example 2.2.8: Some common matrix groups and their Lie algebras

We denote by $M(n,k)$ the matrices of dimension $n \times n$ with entries in $k = \mathbb{R}$ or $k = \mathbb{C}$ and write $\text{diag}(p,q)$ for the diagonal matrix with the first p entries equal to 1 and the remaining q entries equal to -1. Some common Lie groups arising in physics and the associated Lie algebras can be listed as follows.

Matrix Lie group G	**Lie algebra** $\mathfrak{g} = T_e G$
$GL(n,k)$ general linear group invertible matrices in $M(n,k)$	$\mathfrak{gl}(n,k) = M(n,k)$
$SL(n,k)$ special linear group matrices in $GL(n,k)$ with $\det = 1$ preserves determinant: $\det(AB) = \det(B)\ \forall A \in SL(n,k), B \in M(n,k)$	$\mathfrak{sl}(n,k)$ traceless matrices matrices in $M(n,k)$ with $\text{tr} = 0$
$U(n)$ unitary group matrices $M \in M(n,\mathbb{C})$ with $M^\dagger = M^{-1}$ preserves standard hermitian form on $\mathbb{C}^n$	$\mathfrak{u}(n)$ anti-hermitian matrices matrices $m \in M(n,\mathbb{C})$ with $m^\dagger = -m$
$SU(n)$ special unitary group matrices $M \in U(n)$ with $\det M = 1$ preserves standard hermitian form on $\mathbb{C}^n$ and determinant	$\mathfrak{su}(n)$ traceless anti-hermitian matrices $m \in \mathfrak{u}(n)$ with $\text{Tr}\,(m) = 0$
$O(n)$ orthogonal group matrices $M \in M(n,\mathbb{R})$ with $M^T = M^{-1}$ preserves standard scalar product on $\mathbb{R}^n$	$\mathfrak{o}(n)$ anti-symmetric matrices $m \in M(n,\mathbb{R})$ with $m^T = -m$
$SO(n)$ special orthogonal group matrices $M \in O(n)$ with $\det M = 1$ preserves standard scalar product on $\mathbb{R}^n$ and determinant	$\mathfrak{so}(n)$ traceless antisymmetric matrices $m \in \mathfrak{o}(n)$ with $\text{Tr}\,(m) = 0$
$O(p,q)$ matrices $M \in M(n,\mathbb{R})$ with $M^T \cdot \text{diag}(p,q) \cdot M = \text{diag}(p,q)$ preserves bilinear form $\text{diag}(p,q)$ on $\mathbb{R}^n$	$\mathfrak{o}(p,q)$
$SO(p,q)$ matrices $M \in O(p,q)$ with $\det M = 1$ preserves bilinear form $\text{diag}(p,q)$ on $\mathbb{R}^n$ and determinant	$\mathfrak{so}(p,q)$ matrices $m \in \mathfrak{o}(p,q)$ with $\text{Tr}\,(m) = 0$

2.3 Representations

2.3.1 Representations and maps between them

Definition 2.3.1: (Representations)

1. A *representation of an (associative, unital) algebra* A is an algebra homomorphism $\rho : A \to \text{End}(V)$ from A into the algebra of endomorphisms of a vector space V.

2. A *representation of a Lie algebra* $\mathfrak{g}$ is a Lie algebra homomorphism from $\mathfrak{g}$ into the Lie algebra $\text{End}(V)$ with the commutator as Lie bracket.

3. A *representation of a group* G is a group homomorphism from G into the group of automorphisms $\mathrm{Aut}(V)$ of a vector space V.

The *dimension* of a representation is the dimension of the associated vector space.

Generally, the vector space on which a group, algebra or Lie algebra is represented can be either finite dimensional or infinite-dimensional. In the infinite dimensional case, one often needs additional assumptions such as continuity. In this chapter, we will assume that all representations are representations on *finite dimensional* vector spaces.

Lemma 2.3.2: (Adjoint representations)

1. Every matrix Lie group G has a representation on its Lie algebra $\mathfrak{g}$, the *adjoint representation* $\mathrm{Ad}: G \to \mathrm{Aut}(\mathfrak{g})$, given by
$$\mathrm{A}d(g)\boldsymbol{x} = g \cdot \boldsymbol{x} \cdot g^{-1} \qquad \forall g \in G, \boldsymbol{x} \in \mathfrak{g}. \tag{2.19}$$

2. Every Lie algebra $\mathfrak{g}$ has a representation on itself, the *adjoint representation* $\mathrm{ad}: \mathfrak{g} \to \mathrm{End}(\mathfrak{g})$, given by
$$\mathrm{ad}(\boldsymbol{x})\boldsymbol{y} = [\boldsymbol{x}, \boldsymbol{y}] \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathfrak{g}. \tag{2.20}$$
In terms of a basis $B = \{e_1, ..., e_n\}$ of $\mathfrak{g}$ with structure constants $[e_i, e_j] = \sum_{k=1}^n f_{ij}{}^k e_k$, the adjoint representation is given by the matrix
$$\mathrm{ad}(\boldsymbol{x})\boldsymbol{y} = \sum_{j,k=1}^n \mathrm{ad}(\boldsymbol{x})_j{}^k y^j e_k \quad \text{with} \quad \mathrm{ad}(\boldsymbol{x})_j{}^k = \sum_{i=1}^n x^i f_{ij}{}^k.$$

Lemma 2.3.3: (Representations of matrix Lie groups $\Rightarrow$ representations of Lie algebras)

Let G be a matrix Lie group with Lie algebra $\mathfrak{g} = T_e G$ and let $\rho_G : G \to \mathrm{Aut}(V)$ be a representation of G on a vector space V over k. Then, a representation $\rho_{\mathfrak{g}} : \mathfrak{g} \to \mathrm{End}(V)$ of the Lie algebra $\mathfrak{g}$ on V is given by
$$\rho_{\mathfrak{g}}(\boldsymbol{x})\boldsymbol{v} = \frac{d}{dt}\Big|_{t=0} \rho_G(c_x(t))\boldsymbol{v} \qquad \forall \boldsymbol{v} \in V, \boldsymbol{x} \in \mathfrak{g}, \tag{2.21}$$
where $c_x :]-\epsilon, \epsilon[\to G$ is a smooth path in G with $c_x(0) = e$ and $\dot{c}_x(0) = \boldsymbol{x}$.

Proof: By definition, for all $\boldsymbol{x} \in \mathfrak{g}$ there exists a path $c_x :]-\epsilon, \epsilon[\to G$ with $c(0) = e$, $\dot{c}(0) = \boldsymbol{x}$. For al $\boldsymbol{v} \in V$, the derivative
$$\frac{d}{dt}\Big|_{t=0} \rho_G(c_x(t))\boldsymbol{v}$$
does not depend on the choice of the path but only on its derivative $\dot{c}_x(0) = \boldsymbol{x}$. We can therefore define a map $\rho_{\mathfrak{g}} : \mathfrak{g} \to \mathrm{End}(V)$ by setting
$$\rho_{\mathfrak{g}}(\boldsymbol{x})\boldsymbol{v} = \frac{d}{dt}\Big|_{t=0} \rho_G(c_x(t))\boldsymbol{v} \qquad \forall \boldsymbol{v} \in V, \boldsymbol{x} \in \mathfrak{g}.$$
For all $\boldsymbol{x}, \boldsymbol{y} \in \mathfrak{g}$ with associated paths $c_x :]-\epsilon, \epsilon[\to G$, $c_y :]-\epsilon, \epsilon[\to G$, $c_x(0) = c_y(0) = e$, $\dot{c}_x(0) = \boldsymbol{x}$, $\dot{c}_y(0) = \boldsymbol{y}$ and all $\boldsymbol{v} \in V$, we have
$$\rho_{\mathfrak{g}}([\boldsymbol{x}, \boldsymbol{y}])\boldsymbol{v} = \frac{d}{dtds}\Big|_{t=s=0} \rho_G(c_x(t) \cdot c_y(s) \cdot c_x(t))\boldsymbol{v} = \rho_{\mathfrak{g}}(\boldsymbol{x})\rho_{\mathfrak{g}}(\boldsymbol{y})\boldsymbol{v} - \rho_{\mathfrak{g}}(\boldsymbol{y})\rho_{\mathfrak{g}}(\boldsymbol{x})\boldsymbol{v} = [\rho_{\mathfrak{g}}(\boldsymbol{x}), \rho_{\mathfrak{g}}(\boldsymbol{y})]\boldsymbol{v},$$
which implies that the map $\rho_{\mathfrak{g}} : \mathfrak{g} \to \mathrm{End}(V)$ is a representation of $\mathfrak{g}$ on V. □

Exercise 20: Show that the following matrices define a representation of the permutation group $S(3)$ on $\mathbb{R}^3$

$$\rho(\pi_{12}) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \rho(\pi_{13}) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad \rho(\pi_{23}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \tag{2.22}$$

where $\pi_{ij} : (1, ..., i, ...j, ...n) \to (1, ..., j, ..., i, ...n) \in S(n)$ denote the elementary exchanges

$$\pi_{ij}(k) = k \quad \forall k \neq i, j \qquad \pi_{ij}(i) = j \qquad \pi_{ij}(j) = i.$$

Exercise 21: (The Heisenberg algebra has no finite-dimensional representations)

We consider the associative algebra generated by two elements X, P and a unit e with commutator $[X, P] = e$. This is an infinite-dimensional vector space with basis $\{e_{kl} = P^k \circ X^l \mid k, l \in \mathbb{N}_0\}$, where $P^0 \circ X^l = X^l$, $P^k \circ X^0 = P^k$, $P^0 \circ X^0 = e$. Its algebra multiplication law is given by the relation

$$X \circ P - P \circ X = e. \tag{2.23}$$

Show that this algebra does *not* have a non-trivial *finite dimensional* representation, i.e. a non-trivial representation by matrices.

Hint: Suppose it has a matrix representation and use the trace to obtain a contradiction.

In physics, one often restricts attention to *unitary* representations of groups on vector spaces V which carry a hermitian product $\langle\, , \rangle$. These are the representations $\rho : G \to \mathrm{End}(V)$ which leave the hermitian product invariant. The motivation for this is that a physical symmetry represented on the Hilbert space should leave transition amplitudes invariant, i.e. preserve the absolute value of the hermitian form.

$$|\langle \rho(g)\boldsymbol{x}, \rho(g)\boldsymbol{y}\rangle| = |\langle \boldsymbol{x}, \boldsymbol{y}\rangle| \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in V, g \in G$$

Wigner's theorem states that such a representation is either *unitary*

$$\langle \rho(g)\boldsymbol{x}, \rho(g)\boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in V, g \in G$$

or *anti-unitary*

$$\langle \rho(g)\boldsymbol{x}, \rho(g)\boldsymbol{y}\rangle = \overline{\langle \boldsymbol{x}, \boldsymbol{y}\rangle} \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in V, g \in G.$$

This provides a motivation for the investigation of unitary representations. We start with a definition.

Definition 2.3.4: (Unitarity)

A representation $\rho : G \to \mathrm{End}(V)$ of a group G on a vector space V over $\mathbb{C}$ with hermitian product $\langle\, , \rangle : V \times V \to \mathbb{C}$ is called *unitary* if

$$\langle \rho(g)\boldsymbol{v}, \rho(g)\boldsymbol{w}\rangle = \langle \boldsymbol{v}, \boldsymbol{w}\rangle \quad \forall \boldsymbol{v}, \boldsymbol{w} \in V, g \in G. \tag{2.24}$$

This is equivalent to the statement that for any basis $B = \{e_1, ..., e_n\}$, the representing matrices $\rho(g)$, $g \in G$ are unitary: $\rho(g)^\dagger = \rho(g)^{-1}$.

Lemma 2.3.5:

1. A unitary representation $\rho_G : G \to \mathrm{Aut}(V)$ of a matrix Lie group G on a complex vector space V induces a representation $\rho_{\mathfrak{g}} : \mathfrak{g} \to \mathrm{End}(V)$ of the associated Lie algebra $\mathfrak{g} = T_e G$ by anti-hermitian matrices.
2. The elements $i\boldsymbol{x}$ with $\boldsymbol{x} \in \mathfrak{g}$ are represented by hermitian matrices. Consequently, for each $\boldsymbol{x} \in \mathfrak{g}$ there exists a basis with respect to which $\rho(i\boldsymbol{x})$ is diagonal with *real* eigenvalues and $\rho(\boldsymbol{x})$ is diagonal with *imaginary* eigenvalues.

Remark 2.3.6:

1. Note that in this lemma $\mathfrak{g}$ is considered as a *real Lie algebra*, i.e. a Lie algebra over $\mathbb{R}$. Linear combinations of elements in $\mathfrak{g}$ with complex coefficients are not necessarily anti-hermitian nor are they necessarily diagonalisable.
2. Although all representing matrices $\rho(i\boldsymbol{x})$, $\rho(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathfrak{g}$ can be diagonalised, in general this cannot be achieved *simultaneously* for all $\boldsymbol{x} \in \mathfrak{g}$. Different elements $\boldsymbol{x} \in \mathfrak{g}$ are diagonalised by different choices of bases. For two elements $\boldsymbol{x}, \boldsymbol{y} \in \mathfrak{g}$, the representation matrices can be diagonalised simultaneously if and only if $\rho(\boldsymbol{x})$ and $\rho(\boldsymbol{y})$ commute.

Proof: Suppose $\rho : G \to \mathrm{Aut}(V)$ is a unitary representation of G on a complex vector space V. Then according to Lemma 2.3.3, the associated representation $\rho_{\mathfrak{g}} : \mathfrak{g} \to \mathrm{End}(V)$ of $\mathfrak{g}$ on V is given by

$$\rho_{\mathfrak{g}}(\boldsymbol{x})\boldsymbol{v} = \frac{d}{dt}\Big|_{t=0} \rho_G(c_x(t))\boldsymbol{v} \qquad \forall \boldsymbol{v} \in V, \boldsymbol{x} \in \mathfrak{g},$$

where $c_x :]-\epsilon, \epsilon[\to G$ is a smooth path with $c(0) = e$, $\dot{c}(0) = \boldsymbol{x}$. To show that $\rho_{\mathfrak{g}}$ is anti-hermitian, we differentiate the product $\rho_G \circ c_x \cdot \rho_G^\dagger \circ c_x$, which by unitarity of the representations is equal to 1

$$\frac{d}{dt}\Big|_{t=0} \underbrace{\rho_G(c_x(t)) \cdot \rho_G^\dagger(c_x(t))}_{=1 \text{ for all } t \in]-\epsilon,\epsilon[} = \rho_{\mathfrak{g}}(\boldsymbol{x}) + \rho_{\mathfrak{g}}(\boldsymbol{x})^\dagger = 0.$$

This shows that $\rho_{\mathfrak{g}}(\boldsymbol{x})$ is anti-hermitian for all $\boldsymbol{x} \in \mathfrak{g}$. □

As we have seen in the previous sections, a homomorphism of vector spaces V, W is a *linear* map between V and W, i.e. a map that is compatible with the structure of a vector space. A homomorphism of algebras A, B is a vector space homomorphism $\phi : A \to B$ that, additionally, is compatible with the algebra multiplication $\phi(a \cdot b) = \phi(a) \cdot \phi(b)$ for all $a, b \in A$. Similarly, a homomorphisms of groups G, H is a map $\phi : G \to H$ that is compatible with the group structure: $\phi(u \cdot v) = \phi(u) \cdot \phi(v)$ for all $u, v \in G$.

This provides a motivation for studying *homomorphisms of representations* or *intertwiners*. These are maps between the representation spaces that are compatible with the action of the groups, algebras or Lie algebras represented on them.

Definition 2.3.7: (Intertwiners - homomorphisms of representations)

An *intertwiner* between two representations $\rho_V : X \to \mathrm{End}(V)$, $\rho_W : X \to \mathrm{End}(W)$, where X is either an associative algebra, a group or a Lie algebra, is a vector space homomorphism $\phi \in \mathrm{Hom}(V, W)$ that is compatible with the action of X on V and W

$$\rho_W(x)\phi(\boldsymbol{v}) = \phi(\rho_V(x)\boldsymbol{v}) \qquad \forall \boldsymbol{v} \in V, x \in X. \tag{2.25}$$

Two representations $\rho_V : X \to \mathrm{End}(V)$, $\rho_W : X \to \mathrm{End}(W)$ are called *isomorphic*, if there exists an intertwiner between them that is an *isomorphism* of vector spaces.

2.3.2 (Ir)reducibility

Definition 2.3.8: (Reducibility and irreducibility)

1. A representation $\rho_V : X \to \mathrm{End}(V)$, where X is either an associative algebra, a group or a Lie algebra, is called *irreducible* if there is no proper linear subspace $W \subset V$, $W \neq \{0\}, V$ that is invariant under the action of X

$$\rho(X)W \subset W \quad \Rightarrow \quad W = 0 \text{ or } W = V. \tag{2.26}$$

   Otherwise, the representation is called *reducible*.

2. A reducible representation $\rho_V : X \to \mathrm{End}(V)$ is called *fully reducible* if there exist vector spaces $V_1, ..., V_n \subset V$, $V_i \cap V_i = \{0\}$ for $i \neq j$ such that

   (a) V_i is invariant under the action of X: $\rho(X)V_i \subset V_i$

   (b) $\rho_{V_i} = \rho_V\big|_{V_i}$ is irreducible

   (c) $V = V_1 \oplus ... \oplus V_n$.

   This is equivalent to the existence of a basis of V in which the representation matrix takes a block diagonal form and all blocks correspond to irreducible representations.

Example 2.3.9: (Spin 1/2 system)

The matrices in Example 2.2.4 define an irreducible representation of $\mathfrak{su}(2)$ on $\mathbb{R}^3$ which agrees with the adjoint representation.

Example 2.3.10: The representation $\rho : \mathfrak{su}(2) \to \mathrm{End}(\mathbb{C}^4)$ given by the representation matrices

$$\rho(K_1) = \begin{pmatrix} K_1 & 0 \\ 0 & 0 \end{pmatrix} \qquad \rho(K_2) = \begin{pmatrix} K_2 & 0 \\ 0 & 0 \end{pmatrix} \qquad \rho(K_3) = \begin{pmatrix} K_3 & 0 \\ 0 & 0 \end{pmatrix}, \tag{2.27}$$

where K_i, $i = 1, 2, 3$ are the matrices from Example 18 is reducible because the subspace $\mathrm{Span}(\{e_3, e_4\})$ is invariant.

Lemma 2.3.11: (Schur's lemma)

An intertwiner between two irreducible representations of a Lie algebra $\mathfrak{g}$, an algebra A or a group G is either an isomorphism or zero.

Proof:

Suppose $\Phi : V \to W$ is an intertwiner between representations $\rho_V : X \to \mathrm{End}(V)$ and $\rho_W : X \to \mathrm{End}(W)$ and consider the kernel $\ker(\Phi) = \{v \in V : \Phi(v) = 0\}$. Because Φ is linear, this is a subspace of V. Because Φ is compatible with the representation ρ_V, it is an *invariant subspace* of V. As ρ_V is irreducible, this implies $\ker(\Phi) = V$ or $\ker(\Phi) = \{0\}$. In the first case, Φ is zero, in the second case, it is injective. Similarly, the image $\mathrm{Im}(\Phi) = \{w \in W : \exists v \in V : w = \Phi(v)\}$ is a subspace of W since Φ is linear. It is invariant under ρ_W because Φ is an intertwiner. This implies $\mathrm{Im}(\Phi) = \{0\}$ or $\mathrm{Im}(\Phi) = W$. In the first case, Φ vanishes. In the second case, Φ is surjective. Hence, Φ is either zero or an isomorphism. □

Corollary 2.3.12: (Schur's lemma 2)

Let $\rho_V : X \to \mathrm{End}(V)$ be an irreducible representation of an algebra, a Lie algebra or a group on a *finite dimensional* vector space V over $\mathbb{C}$. Then, any self-intertwiner $\Phi \in \mathrm{End}(V)$ is a multiple of the identity, i.e. there exists $\lambda \in \mathbb{C}$ such that $\Phi(v) = \lambda v$ for all $v \in V$.

Proof: As V is a finite-dimensional vector space over $\mathbb{C}$, Φ has at least one eigenvector with eigenvalue $\lambda \in \mathbb{C}$. This implies that $\ker(\Phi - \lambda 1_{\dim(V)}) \neq \{0\}$. But $\Phi - \lambda 1_{\dim(V)}$ is also an intertwiner. Schur's lemma then implies $\Phi - \lambda 1_{\dim(V)} = 0$. □

Corollary 2.3.13: (Schur's lemma 3)

Let $\rho_V : X \to \mathrm{End}(V)$ be an irreducible representation of an algebra, a Lie algebra or a group on a n-dimensional vector space V over $\mathbb{C}$. A matrix $M \in M(n, \mathbb{C})$ that commutes with all representation matrices $(\rho_V(X)_i{}^j)_{i,j=1,...,n}$ with respect to a basis $B = \{e_1, ..., e_n\}$ of V is a multiple of the identity matrix 1_n.

Proof: Any element $M \in M(n, \mathbb{C})$ defines an endomorphism $\Phi \in \mathrm{End}(V)$ via the choice of a basis $B = \{e_1, ..., e_n\}$ and the identity

$$\Phi(e_i) = \sum_{j=1}^{n} M_i{}^j e_j. \tag{2.28}$$

If M commutes with all representation matrices $(\rho_V(x)_i{}^j)_{i,j=1,...,n}$, $x \in X$, Φ is a self-intertwiner of ρ_V. Corollary 2.3.12 then implies that Φ is a multiple of the identity and M therefore a multiple of the identity matrix. □

Remark 2.3.14: Although Lemma 2.3.11 and Corollaries 2.3.12, 2.3.13 can all be found under the name "Schur's lemma" in the literature, they are *not* equivalent. Lemma 2.3.11 is stronger than Corollary 2.3.12 which is in turn stronger than Corollary 2.3.13.

Remark 2.3.15: Corollary 2.3.13 is useful to determine if a given representation is irreducible. If it is possible to find a matrix that is not proportional to the identity matrix and commutes with all representation matrices, the representation is reducible.

2.3.3 (Semi)simple Lie algebras, Casimir operators and Killing form

The question if the adjoint representation of a Lie algebra on itself is reducible leads to the notion of *simple* and *semisimple* Lie algebras, which play a prominent role in the classification of Lie algebras. They also are very important in the classification of their representations.

Definition 2.3.16: (Ideal, simple, semisimple)

An *ideal* of a Lie algebra is a linear subspace $\mathfrak{i} \subset \mathfrak{g}$ such that $[\boldsymbol{x}, \boldsymbol{y}] \in \mathfrak{i}$ for all $\boldsymbol{x} \in \mathfrak{g}$ and $\boldsymbol{y} \in \mathfrak{i}$. A Lie algebra is called *simple* if it is *not abelian* and it has *no non-trivial ideals*, i.e. it has no ideals $\mathfrak{i} \neq \{0\}, \mathfrak{g}$. A Lie algebra is called *semisimple* if it is the direct sum of simple Lie algebras, i.e. there exist simple Lie subalgebras $\mathfrak{g}_i \subset \mathfrak{g}$, $i = 1, ..., k$, such that $\mathfrak{g} = \mathfrak{g}_1 \oplus ... \oplus \mathfrak{g}_k$ as a vector space and the Lie bracket takes the form

$$[\boldsymbol{x}, \boldsymbol{y}]_{\mathfrak{g}} = 0 \text{ for } \boldsymbol{x} \in \mathfrak{g}_i, \boldsymbol{y} \in \mathfrak{g}_j, i \neq j \qquad [\boldsymbol{x}, \boldsymbol{y}]_{\mathfrak{g}} = [\boldsymbol{x}, \boldsymbol{y}]_{\mathfrak{g}_i} \text{ for } \boldsymbol{x}, \boldsymbol{y} \in \mathfrak{g}_i. \tag{2.29}$$

Corollary 2.3.17:

The adjoint representation of a Lie algebra $\mathfrak{g}$ is irreducible if and only if $\mathfrak{g}$ is *simple*.

Theorem 2.3.18: A Lie algebra is semisimple if and only if it has no non-trivial *abelian* ideals.

Proof: The proof can be found in many textbooks on Lie algebras.

We now construct a class of Lie algebras which are neither simple nor semisimple, the Lie algebras associated with *semi-direct product groups*.

Example 2.3.19: (Semidirect products)

1. Let G be a matrix Lie group with Lie algebra $\mathfrak{g} = T_e G$ and denote by $[\,,\,]_{\mathfrak{g}}$ the Lie bracket on $\mathfrak{g}$. Let $\rho_G : G \to \mathrm{Aut}(V)$ be a representation of G on a vector space V over k and $\rho_{\mathfrak{g}} : \mathfrak{g} \to \mathrm{End}(V)$ the associated representation of $\mathfrak{g}$.

   We consider the set $G \times V = \{(g, \boldsymbol{v}) \mid g \in G, \boldsymbol{v} \in V\}$ and define a multiplication $\cdot : (G \times V) \times (G \times V) \to (G \times V)$ by setting

   $$(g, \boldsymbol{v}) \cdot (h, \boldsymbol{w}) = (gh, \boldsymbol{v} + \rho_G(g)\boldsymbol{w}) \qquad \forall g, h \in G, \boldsymbol{v}, \boldsymbol{w} \in V.$$

   Then, the set $G \times V$ with the multiplication $\cdot$ is a group, called the *semidirect product* of G and V and denoted $G \ltimes V$.

2. Its Lie algebra, denoted $\mathfrak{g} \ltimes V$, is the set $\mathfrak{g} \times V = \{(\boldsymbol{x}, \boldsymbol{v}) \mid \boldsymbol{x} \in \mathfrak{g}, \boldsymbol{v} \in V\}$ with the vector space structure of the direct sum

   $$(\boldsymbol{x}, \boldsymbol{v}) + (\boldsymbol{y}, \boldsymbol{w}) = (\boldsymbol{x} + \boldsymbol{y}, \boldsymbol{v} + \boldsymbol{w}) \qquad t(\boldsymbol{x}, \boldsymbol{v}) = (t\boldsymbol{x}, t\boldsymbol{v}) \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathfrak{g}, \boldsymbol{v}, \boldsymbol{w} \in V, t \in k.$$

   and with Lie bracket

   $$[(\boldsymbol{x}, \boldsymbol{v}), (\boldsymbol{y}, \boldsymbol{w})] = ([\boldsymbol{x}, \boldsymbol{y}]_{\mathfrak{g}}, \rho_{\mathfrak{g}}(\boldsymbol{x})\boldsymbol{w} - \rho_{\mathfrak{g}}(\boldsymbol{y})\boldsymbol{v}) \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathfrak{g}, \boldsymbol{v}, \boldsymbol{w} \in V.$$

3. This Lie algebra is neither simple nor semi-simple. The linear subspaces $V = \{(0, \boldsymbol{v}) \mid \boldsymbol{v} \in V\} \subset \mathfrak{g} \ltimes V$ and $\mathfrak{g} = \{(\boldsymbol{x}, 0) \mid \boldsymbol{x} \in \mathfrak{g}\}$ are ideals of $\mathfrak{g}$

   $$\begin{aligned} &[(\boldsymbol{x}, \boldsymbol{v}), (0, \boldsymbol{w})] = (0, \rho_{\mathfrak{g}}(\boldsymbol{x})\boldsymbol{w}) && \forall \boldsymbol{x} \in \mathfrak{g}, \boldsymbol{v}, \boldsymbol{w} \in V \\ &[(\boldsymbol{x}, \boldsymbol{v}), (\boldsymbol{y}, 0)] = ([\boldsymbol{x}, \boldsymbol{y}]_{\mathfrak{g}}, 0) && \forall \boldsymbol{x}, \boldsymbol{y} \in \mathfrak{g}, \boldsymbol{v} \in V. \end{aligned}$$

   This implies that $\mathfrak{g} \ltimes V$ cannot be simple, unless G is the trivial Lie group: $G = \{e\}$ or V is the trivial vector space of dimension 0: $V = \{0\}$. Moreover, one finds that the vector space $V \subset \mathfrak{g} \ltimes V$ is an *abelian* ideal of $\mathfrak{g} \ltimes V$

   $$[(0, \boldsymbol{v}), (0, \boldsymbol{w})] = 0 \qquad \forall \boldsymbol{v}, \boldsymbol{w} \in V.$$

   By Theorem 2.3.18, a semisimple Lie algebra cannot have any non-trivial *abelian* ideals, which implies that the semidirect product $\mathfrak{g} \ltimes V$ cannot be semisimple. □

Remark 2.3.20: Many groups that are important in physics are semidirect products. Examples are the Euclidean group $SO(3) \ltimes \mathbb{R}^3$ and the Poincaré group $SO(3,1) \ltimes \mathbb{R}^4$.

Exercise 22: (Semidirect products)

1. Let G be a matrix Lie group with Lie algebra $\mathfrak{g} = T_eG$ and denote by $[\,,\,]_{\mathfrak{g}}$ the Lie bracket on $\mathfrak{g}$. Let $\rho_G : G \to \mathrm{Aut}(V)$ be a representation of G on a vector space V over k and $\rho_{\mathfrak{g}} : \mathfrak{g} \to \mathrm{End}(V)$ the associated representation of $\mathfrak{g}$. We consider the set $G \times V = \{(g, \boldsymbol{v}) \mid g \in G, \boldsymbol{v} \in V\}$ with multiplication law

$$(g, \boldsymbol{v}) \cdot (h, \boldsymbol{w}) = (gh, \boldsymbol{v} + \rho_G(g)\boldsymbol{w}) \qquad \forall g, h \in G, \boldsymbol{v}, \boldsymbol{w} \in V.$$

   Show that this multiplication law gives $G \times V$ the structure of a group, which we denote by $G \ltimes V$. Give the unit element of the group. What is the inverse of an element $(g, \boldsymbol{v})$ with $g \in G$, $\boldsymbol{v} \in V$?

2. Consider smooth paths $c_{(x,v)} :]-\epsilon, \epsilon[\to G \ltimes V$ with $c_{x,v}(0) = e$ parametrised as

$$c_{x,v}(t) = (g_x(t), \boldsymbol{v}(t)) \qquad g(t) \in G, c_v(t) \in V,$$

   where $g_x :]-\epsilon, \epsilon[\to G$ is a smooth path in G with $g_x(0) = e_G$, $\dot{g}_x(0) = \boldsymbol{x}$ and $c_v :]-\epsilon, \epsilon[\to V$ is a smooth path in V with $c_v(0) = 0$, $\dot{c}_v(0) = \boldsymbol{v}$. Express the product $c_{x,v}(t) \cdot c_{y,w}(s) \cdot c_{x,v}(t)^{-1}$ in terms of the paths $g_x, g_y :]-\epsilon, \epsilon[\to G$ and $c_v, c_w :]-\epsilon, \epsilon[\to V$. Show that its derivative at $t = s = 0$ is given by

$$\left.\frac{d^2}{dsdt}\right|_{t=s=0} c_{x,v}(t) \cdot c_{y,w}(s) \cdot c_{x,v}(t)^{-1} = ([\boldsymbol{x}, \boldsymbol{y}]_{\mathfrak{g}}, \rho_{\mathfrak{g}}(\boldsymbol{x})\boldsymbol{w} - \rho_{\mathfrak{g}}(\boldsymbol{y})\boldsymbol{v}).$$

   Show by direct calculation that

$$[(\boldsymbol{x}, \boldsymbol{v}), (\boldsymbol{y}, \boldsymbol{w})] := ([\boldsymbol{x}, \boldsymbol{y}]_{\mathfrak{g}}, \rho_{\mathfrak{g}}(\boldsymbol{x})\boldsymbol{w} - \rho_{\mathfrak{g}}(\boldsymbol{y})\boldsymbol{v}) \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathfrak{g}, \boldsymbol{v}, \boldsymbol{w} \in V.$$

   defines a Lie bracket on the vector space $\mathfrak{g} \oplus V$.

An important tool in the classification of Lie algebras and in the investigation of their properties is the *Killing form*, a canonical symmetric bilinear form on a Lie algebra.

Definition 2.3.21: (Killing form)

The *Killing form* of a finite dimensional Lie algebra $\mathfrak{g}$ over k is the bilinear form $\kappa : \mathfrak{g} \times \mathfrak{g} \to k$ defined by

$$\kappa(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{Tr}\left(\mathrm{ad}_{\boldsymbol{x}} \circ \mathrm{ad}_{\boldsymbol{y}}\right) \tag{2.30}$$

where $\mathrm{ad}_{\boldsymbol{x}} : \boldsymbol{y} \mapsto [\boldsymbol{x}, \boldsymbol{y}]$ is the adjoint representation. In terms of a basis $B = \{e_1, .., e_n\}$ of $\mathfrak{g}$ and the associated structure constants it takes the form

$$\kappa_{ij} = \kappa(e_i, e_j) = \sum_{k,l=1}^{n} f_{il}{}^k f_{jk}{}^l. \tag{2.31}$$

Example 2.3.22: (Killing form of $\mathfrak{su}(2)$)

In terms of the generators J_i, $i = 1, 2, 3$ in example 3, the Killing form of $su(2)$ takes the form $\kappa(J_i, J_j) = \delta_{ij}$.

Lemma 2.3.23: (Properties of the Killing form)

The Killing-form is

1. *bilinear*
2. *symmetric*: $\kappa(\boldsymbol{x}, \boldsymbol{y}) = \kappa(\boldsymbol{y}, \boldsymbol{x})$ for all $\boldsymbol{a}, \boldsymbol{y} \in \mathfrak{g}$
3. ad-*invariant*: $\kappa([\boldsymbol{y}, \boldsymbol{x}], \boldsymbol{z}) + \kappa(\boldsymbol{x}, [\boldsymbol{y}, \boldsymbol{z}]) = 0$ for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathfrak{g}$.

Proof: Exercise.

Exercise 23:

1. Consider the Killing form κ on a Lie algebra $\mathfrak{g}$ defined by

$$\kappa(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{Tr}\left(\mathrm{ad}_{\boldsymbol{x}} \circ \mathrm{ad}_{\boldsymbol{y}}\right) \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathfrak{g},$$

where for $\boldsymbol{x} \in \mathfrak{g}$, $\mathrm{ad}_{\boldsymbol{x}} : \mathfrak{g} \to \mathfrak{g}$ is the bilinear map defined by

$$\mathrm{ad}_{\boldsymbol{x}}(\boldsymbol{y}) = [\boldsymbol{x}, \boldsymbol{y}] \qquad \forall \boldsymbol{y} \in \mathfrak{g}.$$

Let $B = \{e_1, ..., e_n\}$ be a basis of $\mathfrak{g}$ with structure constants $[e_i, e_j] = \sum_{k=1}^n f_{ij}{}^k e_k$. Use the expression for the adjoint action in terms of the structure constants

$$\mathrm{ad}_{\boldsymbol{x}}(e_j) = \sum_{k=1}^n (\mathrm{ad}_{\boldsymbol{x}})_j{}^k e_k \qquad (\mathrm{ad}_{\boldsymbol{x}})_j{}^k = \sum_{i=1}^n x^i f_{ij}{}^k$$

to show that κ takes the form

$$\kappa(\boldsymbol{x}, \boldsymbol{y}) = \sum_{k,l=1}^n x^i y^j f_{li}{}^k f_{kj}{}^l = \sum_{k,l=1}^n x^i y^j f_{il}{}^k f_{jk}{}^l.$$

2. Show that κ is bilinear and use the conjugation invariance of the trace to show that κ is symmetric. Use the Jacobi identity to show

$$\mathrm{ad}_{[\boldsymbol{x},\boldsymbol{y}]}(\boldsymbol{z}) = \mathrm{ad}_{\boldsymbol{x}} \circ \mathrm{ad}_{\boldsymbol{y}}(\boldsymbol{z}) - \mathrm{ad}_{\boldsymbol{y}} \circ \mathrm{ad}_{\boldsymbol{x}}(\boldsymbol{z}) \qquad \forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathfrak{g}.$$

Use this to demonstrate that κ is ad-invariant.

Theorem 2.3.24: (Killing form)

1. If $\mathfrak{g}$ is a *simple*, finite dimensional Lie algebra over $\mathbb{C}$, the Killing form is unique up to scalar multiplication: Any bilinear form on $\mathfrak{g}$ with the properties in Lemma 2.3.23 is of the form $\alpha\kappa$ with $\alpha \in \mathbb{C} \setminus \{0\}$.
2. The Killing form of a finite dimensional Lie algebra $\mathfrak{g}$ over $\mathbb{C}$ is non-degenerate if and only if $\mathfrak{g}$ is *semisimple*.

Proof:

1. The first statement follows from Schur's lemma. If $\beta : \mathfrak{g} \times \mathfrak{g} \to k$ is a bilinear form with the properties of Lemma 2.3.23, then its representing matrix $\beta_{ij} = \beta(e_i, e_j)$ with respect to a basis $B = \{e_1, ..., e_n\}$ is symmetric and commutes with all matrices $(\mathrm{ad}_{\boldsymbol{x}})_i{}^k$. By Schur's lemma, it is therefore a multiple of the identity.

2. The second statement can be proved as follows: To show that the Killing form is non-degenerate if $\mathfrak{g}$ is semisimple, we express $\mathfrak{g}$ as a direct sum $\mathfrak{g} = \mathfrak{g}_1 \oplus ... \oplus \mathfrak{g}_k$, where $\mathfrak{g}_i$, $k = i, .., k$ are simple subalgebras and $[\mathfrak{g}_i, \mathfrak{g}_j] = 0$ for $i \neq j$. We select bases $B_i = \{e^i_1, ..., e^i_{s_i}\}$ of $\mathfrak{g}_i$. Then the representing matrix $(\kappa_{ij})_{i,k=1,...,n}$ of κ with respect to the basis $B = \bigcup_{i=1}^k B_i$ takes a block diagonal form with blocks corresponding to the restrictions $\kappa_i = \kappa\big|_{\mathfrak{g}_i}$. As $\mathfrak{g}_i$ is simple, κ_i is a multiple of the identity matrix $\kappa_i = \lambda_i 1_{s_i}$ with $\lambda_i \neq 0$. Hence, the matrix $(\kappa_{ij})_{i,k=1,...,n}$ is diagonal with eigenvalues $\lambda_i \neq 0$. This implies that κ is non-degenerate.

3. To show that $\mathfrak{g}$ is semisimple if its Killing form is non-degenerate, we suppose that $\mathfrak{i} \subset \mathfrak{g}$ is an ideal in $\mathfrak{g}$. Then $\mathfrak{i}^\perp = \{\boldsymbol{x} \in \mathfrak{g} \mid \kappa(\boldsymbol{x}, \boldsymbol{y}) = 0 \;\forall \boldsymbol{y} \in \mathfrak{i}\}$ is also an ideal of $\mathfrak{g}$ since $\kappa([\boldsymbol{y}, \boldsymbol{x}], \boldsymbol{z}) = -\kappa(\boldsymbol{x}, [\boldsymbol{y}, \boldsymbol{z}]) = 0$ if $\boldsymbol{x} \in \mathfrak{i}^\perp, \boldsymbol{z} \in \mathfrak{g}$, $\boldsymbol{y} \in \mathfrak{i}$. Hence $\mathfrak{g} = \mathfrak{i} \oplus \mathfrak{i}^\perp$, where both $\mathfrak{i}, \mathfrak{i}^\perp$ are ideals and therefore subalgebras of $\mathfrak{g}$. If they are both simple, the proof is complete. Otherwise, one of them, say $\mathfrak{i}$ contains an ideal $\mathfrak{j} \subset \mathfrak{i}$. We then repeat the argument and find $\mathfrak{i} = \mathfrak{j} \oplus \mathfrak{j}^\perp$, where the orthogonal complement is taken with respect to the restriction of the Killing form to $\mathfrak{i}$. After a finite number of steps, we obtain a decomposition of $\mathfrak{g}$ as a direct sum of simple subalgebras. □

Definition 2.3.25: (Quadratic Casimir)

The quadratic Casimir for a finite-dimensional simple matrix Lie algebra $\mathfrak{g}$ with basis $B = \{e_1, ..., e_n\}$ is the matrix

$$C = \sum_{i,j=1}^n \kappa^{ij} e_i \cdot e_j \quad \text{where } \kappa_{ij} = \kappa(e_i, e_j) \;\; \kappa_{ij}\kappa^{jk} = \delta_i^j, \tag{2.32}$$

where κ is the Killing form and the product is the matrix multiplication. Note that the quadratic Casimir is *not* an element of the Lie algebra $\mathfrak{g}$.

Example 2.3.26: (Quadratic Casimir for $\mathfrak{su}(2)$)

The quadratic Casimir of $\mathfrak{su}(2)$ is given by

$$J^2 = J_1^2 + J_2^2 + J_3^2, \tag{2.33}$$

where J_i are the generators of $\mathfrak{su}(2)$ in example 3.

Lemma 2.3.27: The quadratic Casimir is independent of the choice of basis and its matrix commutator with all elements of $\mathfrak{g}$ vanishes.

Proof:

When expressed in terms of structure constants, the ad-invariance of the Killing form reads

$$\sum_{j=1}^n \kappa_{ij} f_{kl}^{\;\;j} + \kappa_{lj} f_{ki}^{\;\;j} = 0, \tag{2.34}$$

which implies for its inverse

$$\sum_{j=1}^{n} \kappa^{ij} f_{jk}^{\ \ l} + \kappa^{jl} f_{jk}^{\ \ i} = 0. \tag{2.35}$$

The Lie bracket of the quadratic Casimir element with an element of the basis is given by

$$[C, e_k] = \sum_{i,j,l=1}^{n} \kappa^{ij} (f_{jk}^{\ \ l} e_i e_l + f_{ik}^{\ \ l} e_l e_j) = \sum_{i,j,l=1}^{n} e_i e_l (\kappa^{ij} f_{jk}^{\ \ l} + \kappa^{jl} f_{jk}^{\ \ i}) = 0. \tag{2.36}$$

Under a change of basis $e'_i = \sum_{j=1}^{n} a_i^{\ j} e_j$, $A \in GL(n, k)$ the Killing form changes according to

$$\kappa'_{ij} = \kappa(e'_i, e'_j) = \sum_{k,l=1}^{n} A_i^{\ k} A_j^{\ l} \kappa(e_k, e_l) = \sum_{k,l=1}^{n} A_i^{\ k} A_j^{\ l} \kappa_{kl}.$$

Its inverse therefore transforms according to

$$\kappa'^{ij} = \sum_{k,l=1}^{n} (A^{-1})_k^{\ i} (A^{-1})_l^{\ j} \kappa^{kl}.$$

The quadratic Casimir is therefore invariant and does not depend on the choice of basis

$$\begin{aligned} C' = \sum_{i,j=1}^{n} \kappa'^{ij} e'_i e'_j &= \sum_{i,j,k,l,m,n=1}^{n} (A^{-1})_k^{\ i} (A^{-1})_l^{\ j} \kappa^{kl} A_i^{\ m} A_j^{\ n} e_n \cdot e_m = \sum_{i,j,k,l,m,n=1}^{n} \delta_k^m \delta_l^n \kappa^{kl} e_m e_n \\ &= \sum_{k,l=1}^{n} \kappa^{kl} e_k e_l = C. \end{aligned}$$

□

Corollary 2.3.28: In any finite-dimensional, irreducible representation of a simple Lie algebra on a vector space over $\mathbb{C}$, the matrix representing the quadratic Casimir element is a multiple of the identity.

Proof: This follows from Schur's lemma (version 3) □

Example 2.3.29: (Irreducible unitary representations of $\mathfrak{su}(2)$)

We determine the finite-dimensional irreducible unitary representations of $\mathfrak{su}(2)$. Let $\rho : \mathfrak{su}(2) \to \text{End}(V)$ be a representation of $\mathfrak{su}(2)$ on a vector space V over $\mathbb{C}$ with hermitian product $\langle\, ,\, \rangle$. The Lie algebra $\mathfrak{su}(2)$ has a basis $B = \{J_1, J_2, J_3\}$ with Lie bracket

$$[J_i, J_j] = i \sum_{k=1}^{3} \epsilon_{ij}^{\ \ k} J_k. \tag{2.37}$$

Unitarity of the representation means that the adjoint of the action of the generators J_i on V agrees with the action of J_i

$$\rho(J_i)^\dagger = \rho(J_i) \quad \text{or, equivalently} \quad \langle \rho(\boldsymbol{x}) w, v \rangle = \langle w, \rho(\boldsymbol{x}) v \rangle \; \forall v, w \in V. \tag{2.38}$$

1. We introduce the new basis

$$J_\pm = J_1 \pm iJ_2 \qquad J_z = J_3 \tag{2.39}$$

in which the Lie bracket and unitarity condition take the form

$$[J_z, J_\pm] = \pm J_\pm \qquad [J_+, J_-] = 2J_z \qquad \rho(J_z)^\dagger = \rho(J_z), \quad \rho(J_\pm)^\dagger = \rho(J_\mp). \tag{2.40}$$

2. We now consider a maximal abelian and semisimple Lie subalgebra of $\mathfrak{su}(2)$, also called *Cartan subalgebra* of $\mathfrak{su}(2)$, i.e. a maximal set of matrices that can be diagonalised simultaneously. The Lie bracket implies that this maximal abelian subalgebra can only be one-dimensional. We select it to be the one-dimensional subspace spanned by J_z. As $\rho(J_z)$ is hermitian, it can be diagonalised over $\mathbb{C}$ with real eigenvalues. Hence, we obtain a basis $B_V = \{v_1, ..., v_n\}$ of V spanned by eigenvectors

$$\rho(J_z)v_k = \lambda_k v_k \qquad \lambda_k \in \mathbb{R}. \tag{2.41}$$

We now note the identity

$$\rho(J_z)\rho(J_\pm)^k v = (\lambda \pm k)\rho(J_\pm)^k v, \tag{2.42}$$

which can be proved by induction using the Lie bracket of the generators $J_\pm$ and J_z:

$$\rho(J_z)\rho(J_\pm)v = \rho(J_\pm)\rho(J_z)v \pm \rho(J_\pm)v = (\lambda \pm 1)\rho(J_\pm)v. \tag{2.43}$$

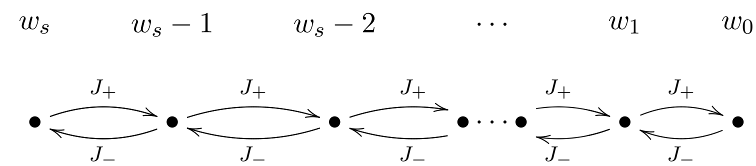
Identity (2.42) implies that for any eigenvector v of $\rho(J_z)$ with eigenvalue λ, $\rho(J_\pm)v$ is either an eigenvector with eigenvalue $\lambda \pm 1$ or vanishes. Suppose now that the basis is ordered such that λ_n is the biggest eigenvalue or *highest weight* of $\rho(J_z)$. Then $\rho(J_+)v_n = 0$, and we obtain a sequence of eigenvectors $w_k = \rho(J_-^k)v_n$ satisfying $\rho(J_z)w_k = (\lambda_n - k)w_k$. As the representation is finite dimensional, there exists a $s \in \mathbb{N}$ such that $\rho(J_-)^k v_n = 0$ for $k > s$, $\rho(J_-^s)v_n \neq 0$.

3. We will now show that the vectors $\rho(J_-^k)v_n$ for $0 \leq k \leq s$ form a basis of V. That they are linearly independent follows from the fact that they are eigenvectors with different eigenvalues. To show that they span V, we use the identity

$$\rho(J_+)w_k = (2k\lambda_n - k(k-1))w_{k-1}, \tag{2.44}$$

which can be proved by induction:

$$\begin{aligned}
\rho(J_+)w_1 &= \rho(J_+)\rho(J_-)v_n = 2J_z v_n = 2\lambda_n v_n = 2\lambda_n w_0 \\
\rho(J_+)w_{k+1} &= \rho(J_-)\rho(J_+)w_k + 2J_z w_k = (2k\lambda_n - k(k-1))\rho(J_-)w_{k-1} + 2(\lambda_n - k)w_k \\
&= (2k\lambda_n - k(k-1) + 2(\lambda_n - k))w_k = (2(k+1)\lambda_n - (k+1)k)w_k
\end{aligned}$$



Hence, we have

$$\rho(J_z)w_k = (\lambda_n - k)w_k \qquad \rho(J_-)w_k = w_{k+1} \qquad \rho(J_+)w_k = (2k\lambda_n - k(k-1))w_{k-1}, \tag{2.45}$$

which implies that $\text{Span}(w_0, ..., w_s)$ is an invariant subspace of V. As ρ is irreducible, $\text{Span}(w_0, ..., w_s) = V$ and $\dim(V) = s + 1$. the irreducible representations are labelled by their dimension $s \in \mathbb{N}$.

4. Normalisation of eigenstates. From the basis $\{w_0, ..., w_s\}$ of V, we construct an *orthonormal* basis, whose elements we denote by $|j, m\rangle$ with $j = \lambda_n \in \mathbb{Z}/2$, $m \in \{-j, -j+1, ..., j-1, j\}$. Supposing $v_n = |j, j\rangle$ is normalised such that $\langle j, j|j, j\rangle = 1$, we find

$$\langle w_{k+1}|w_{k+1}\rangle = \langle w_k|\rho(J_+)w_{k+1}\rangle = 2jk - k(k-1)\langle w_k|w_k\rangle \tag{2.46}$$

This implies after some further calculations that the action of $\mathfrak{su}(2)$ on the normalised states is given by

$$J_z|j, m\rangle = m|j, m\rangle \qquad J_\pm|j, m\rangle = \sqrt{j(j+1) - m(m \pm 1)}\,|j, m \pm 1\rangle \tag{2.47}$$

5. Finally, we determine the eigenvalues of the quadratic Casimir $\rho(J^2)$ of $\mathfrak{su}(2)$

$$J^2 = J_1^2 + J_2^2 + J_3^2 = J_z^2 + \tfrac{1}{2}(J_+J_- + J_-J_+). \tag{2.48}$$

In any irreducible, finite dimensional representation $\rho(J^2)$ must be proportional to the unit matrix, $\rho(J^2) = \tau 1$ with a proportionality constant τ that characterises the representation. This implies

$$\rho(J^2)w_0 = \rho(J_z)^2 w_0 + \tfrac{1}{2}\rho(J_+)\rho(J_-)w_0 = \lambda_n^2 w_0 + \tfrac{1}{2}\rho(J_+)w_1 = \lambda_n(\lambda_n + 1)w_0 \tag{2.49}$$
$$\rho(J^2)w_s = \rho(J_z)^2 w_s + \tfrac{1}{2}\rho(J_-)\rho(J_+)w_s = (\lambda_n - s)^2 w_s + s\lambda_n - \tfrac{1}{2}s(s-1)w_s \tag{2.50}$$

We have

$$\tau = \lambda_n^2 + \lambda_n = (\lambda_n - s)^2 + s\lambda_n - \tfrac{1}{2}s(s-1) \quad \Rightarrow \quad \lambda_n = \tfrac{1}{2}s. \tag{2.51}$$

and find that the eigenvalues of $\rho(J_z)$ are $\frac{s}{2}, \frac{s}{2} - 1, ..., -\frac{s}{2} + 1, -\frac{s}{2}$.

In physics, Lie algebra symmetries such as the $\mathfrak{su}(2)$ symmetry in this example correspond to observables. Therefore, their eigenvalues in a representation are quantum numbers. The eigenvalues j of the Casimir operator J^2, which correspond to the absolute value of the angular momentum or spin vector are called the principal angular momentum or *spin quantum number*. The eigenvalues m of J_z, which correspond to the z-component of the angular momentum, are referred to as *magnetic quantum numbers*.

2.4 Duals, direct sums and tensor products of representations

2.4.1 Groups and Lie groups

Given a representation of a Lie algebra on a vector space V, one can construct new representations using the associated vector spaces investigated in Chapter 1 - the dual V^*, the direct sum $V \oplus W$ and tensor product $V \otimes W$. We obtain the following definition.

Definition 2.4.1: (Representations of Lie algebras - constructions)

1. *Dual of a representation*: The dual of the representation ρ_V on a on vector spaces V is the representation $\rho_{V^*}: \mathfrak{g} \to \mathrm{End}(V^*)$ on the dual vector space V^* defined via

$$\rho_{V^*}(\boldsymbol{x})\alpha = \alpha \circ \rho_V(-\boldsymbol{x}) \qquad \forall \alpha \in V^*, \boldsymbol{x} \in \mathfrak{g} \tag{2.52}$$

or, in terms of a basis $B_V = \{e_1, ..., e_n\}$ and its dual

$$\rho_{V^*}(\boldsymbol{x})e^i = \sum_{k=1}^{n} (\rho_V(-\boldsymbol{x}))_k{}^i e^k \qquad \forall \boldsymbol{x} \in \mathfrak{g}. \tag{2.53}$$

2. *Direct sum of representations*: The direct sum of the representations $\rho_V : \mathfrak{g} \to \mathrm{End}(V)$, $\rho_W : \mathfrak{g} \to \mathrm{End}(W)$ is the representation $\rho_{V\oplus W} : \mathfrak{g} \to \mathrm{End}(V \oplus W)$ on the direct sum of the associated vector spaces defined by

$$\rho_{V\oplus W}(\boldsymbol{x})(v + w) = \rho_V(\boldsymbol{x})v + \rho_W(\boldsymbol{x})w \qquad \forall \boldsymbol{x} \in \mathfrak{g}, v \in V \subset V \oplus W, w \in W \subset V \oplus W.$$

3. *Tensor product of representations*: The tensor product of the representations $\rho_V : \mathfrak{g} \to \mathrm{End}(V)$, $\rho_W : \mathfrak{g} \to \mathrm{End}(W)$ is the representation $\rho_{V\otimes W} : \mathfrak{g} \to \mathrm{End}(V \otimes W)$ on the tensor product of the associated vector spaces defined by

$$\rho_{V\otimes W}(\boldsymbol{x})(v \otimes w) = (\rho_V(\boldsymbol{x})v) \otimes w + v \otimes (\rho_W(\boldsymbol{x})w) \qquad \forall \boldsymbol{x} \in \mathfrak{g}, v \in V, w \in W. \tag{2.54}$$

Remark 2.4.2: In particle physics, elementary particles are described by *irreducible* representations of Lie algebras. Examples are quarks, electrons, neutrinos etc. The *duals* of these representations describe the associated *anti-particle*. *Composite particles* such as the hadrons (including the proton and the neutron) and the mesons are described by *tensor products* of irreducible representations.

It can be shown that for a semisimple Lie algebra, the tensor product of irreducible, finite dimensional representations is always fully reducible, i.e. decomposable into irreducible representations. An irreducible representation that arises in such a decomposition is called a *multiplet* or, if the dimension of the representation is one, a *singlet*. More precisely, one speaks of a *doublet* if the associated irreducible representation has dimension two, a *triplet* if it has dimension three, a *quadruplet* for dimension four etc. The composite particles which correspond to such multiplets have similar masses and similar behaviour with respect to the fundamental interactions modelled by these representations.

The construction of a representations on dual vector spaces and tensor products of vector spaces for Lie algebras can be performed analogously for groups if we replace $\mathfrak{g} \leftrightarrow G$, $\mathrm{End}(\cdot) \leftrightarrow \mathrm{Aut}(\cdot)$, $-\boldsymbol{x} \in \mathfrak{g} \leftrightarrow g^{-1} \in G$ and sums with products in (2.54) in Def. 2.4.1.

This is corresponds to the concepts of *additive* and *multiplicative* quantum numbers. *Additive quantum numbers* are eigenvalues of Lie algebra elements in representations on a Hilbert space. If one considers two copies of a quantum mechanical system, which corresponds to taking the tensor product of two representations, the corresponding eigenvalue in the tensor product of the representations is the *sum* of the eigenvalues in the individual representations.

Multiplicative quantum numbers are eigenvalues of group elements in a representation of a group on a Hilbert space. When considering two copies of the system, i.e. taking the tensor product, the corresponding eigenvalue in the tensor product of the representations is the *product* of the eigenvalues in the individual representations.

Definition 2.4.3: (Representations of groups - constructions)

1. *Dual of a representation*: The dual of the representation $\rho_V : G \to \text{Aut}(V)$ on a vector space V is the representation $\rho_{V^*} : G \to \text{Aut}(V^*)$ on the dual vector space V^* defined via

$$\rho_{V^*}(g)\alpha = \alpha \circ \rho_V(g^{-1}) \qquad \forall \alpha \in V^*, g \in G \tag{2.55}$$

or, in terms of a basis $B_V = \{e_1, ..., e_n\}$ and its dual,

$$\rho_{V^*}(g)e^i = \sum_{k=1}^{n} (\rho_V(g^{-1}))_k{}^i e^k \qquad \forall g \in G. \tag{2.56}$$

2. *Direct sum of representations*: The direct sum of the representations $\rho_V : G \to \text{Aut}(V)$, $\rho_W : G \to \text{Aut}(W)$ is the representation $\rho_{V\oplus W} : G \to \text{Aut}(V \oplus W)$ on the direct sum of the associated vector spaces defined by

$$\rho_{V\oplus W}(g)(v + w) = \rho_V(g)v + \rho_W(g)w \qquad \forall g \in G, v \in V \subset V \oplus W, w \in W \subset V \oplus W.$$

3. *Tensor product of representations*: The tensor product of the representations $\rho_V : G \to \text{Aut}(V)$, $\rho_W : G \to \text{Aut}(W)$ is the representation $\rho_{V\otimes W} : G \to \text{Aut}(V \otimes W)$ on the tensor product of the associated vector spaces defined by

$$\rho_{V\otimes W}(g)(v \otimes w) = (\rho_V(g)v) \otimes (\rho_W(g)w) \qquad \forall g \in G, v \in V, w \in W. \tag{2.57}$$

Exercise 24:

1. Show that the maps $\rho_{V\otimes W} : G \to \text{Aut}(V \otimes W)$, $\rho_{V^*} : G \to \text{Aut}(V^*)$ defined by

$$\begin{aligned} &\rho_{V\otimes W}(g)(v \otimes w) = (\rho_V(g)v) \otimes (\rho_W(g)w) && \forall g \in G, v \in V, w \in W \\ &\rho_{V^*}(g)\alpha = \alpha \circ \rho_V(g^{-1}) && \forall \alpha \in V^*, g \in G \end{aligned} \tag{2.58}$$

and the maps $\rho_{V\otimes W} : \mathfrak{g} \to \text{End}(V \otimes W)$, $\rho_{V^*} : \mathfrak{g} \to \text{End}(V^*)$ defined by

$$\begin{aligned} &\rho_{V\otimes W}(\boldsymbol{x})(v \otimes w) = \rho_V(\boldsymbol{x})v \otimes w + v \otimes \rho_W(w)w && \forall \boldsymbol{x} \in \mathfrak{g}, v \in V, w \in W \\ &\rho_{V^*}(\boldsymbol{x})\alpha = \alpha \circ \rho_V(-\boldsymbol{x}) && \forall \boldsymbol{x} \in \mathfrak{g}, \alpha \in V^* \end{aligned} \tag{2.59}$$

are indeed representations of G and $\mathfrak{g}$ on $V \otimes W$ and V^* if ρ_V and ρ_W are representations of G and $\mathfrak{g}$ on V and W. Why are the minus sign in (2.59) and the inverse in (2.58) needed?

2. Suppose now that G is a Lie group with Lie algebra $\mathfrak{g} = T_eG$. Show that expression (2.59) for the representations of $\mathfrak{g}$ on $V \otimes W$ and V^* follows from (2.58).

   Hint: Consider a path $c_x :]-\epsilon, \epsilon[\to G$ with $c_x(0) = e$, $\dot{c}_x(0) = \boldsymbol{x}$ and differentiate.

2.4.2 Hopf algebras

We are now ready to investigate the case of associative unital algebras. Clearly, one can construct the direct sum of two representations just as for Lie algebras and groups.

Definition 2.4.4: (Representations of associative algebras - direct sums)

Let A be an associative algebra and $\rho_V : A \to \mathrm{End}(V)$, $\rho_W : A \to \mathrm{End}(W)$ representations of A on vector spaces V, W over k. Then the *direct sum* of the representations ρ_V, ρ_W is the representation $\rho_{V\oplus W} : A \to \mathrm{End}(V \oplus W)$ on the direct sum $V \oplus W$ defined by

$$\rho_{V\oplus W}(a)(\boldsymbol{v} + \boldsymbol{w}) = \rho_V(a)\boldsymbol{v} + \rho_W(a)\boldsymbol{w} \qquad \forall a \in A, \boldsymbol{v} \in V \subset V \oplus W, \boldsymbol{w} \in W \subset V \oplus W.$$

However, when attempting to construct a representation on A on duals or on tensor products of representation spaces, one encounters difficulties. As inverses do not need to exist for elements of an associative algebra A, we cannot use equation (2.55) to define a representation on the dual. Using equation (2.52) instead would not be sufficient to guarantee that the resulting map $\rho_{V^*} : A \to \mathrm{End}(V^*)$ is an algebra homomorphism satisfying $\rho^*_V(a \cdot b) = \rho_{V^*}(a) \cdot \rho_{V^*}(b)$ for all $a, b \in A$.

Similar difficulties arise for the tensor product of the representations. Setting $\rho_{V\otimes W}(a)(v \otimes w) = \rho_V(a)v \otimes \rho_W(a)w$ for all $a \in A$ leads to a contradiction because for a linear combination $a = \sum_{j=1}^m \lambda_j a_j$, $a_j \in A$, $\lambda_j \in k$ we would obtain

$$\rho_{V\otimes W}(a)(v \otimes w) = \sum_{i,j=1}^n \lambda_i\lambda_j \rho_V(a_i)v \otimes \rho_W(a_j)w \neq \sum_{i=1}^n \lambda_i \rho_{V\otimes W}(a_i)v \otimes w.$$

Setting $\rho_{V\otimes W}(a)v\otimes w = \rho_V(a)v\otimes w + v\otimes\rho_W(a)w$ as in the case of a Lie algebra would not be sufficient to guarantee to compatibility with the multiplication $\rho_{V\otimes W}(a{\cdot}b) = \rho_{V\otimes W}(a)\rho_{V\otimes W}(b)$.

Finally, we note that both a Lie group G and a Lie algebra $\mathfrak{g}$ have a canonical representation, the trivial representation on the field k which is given by

$$\rho_k(g) = 1 \quad \forall g \in G \qquad \rho_k(\boldsymbol{x}) = 0 \quad \forall \boldsymbol{x} \in \mathfrak{g}.$$

This again is not the case for an associative unital algebra since such a representation would have to satisfy $\rho_k(e) = 1$ for the unit and $\rho_k(0) = 0$ and there is no canonical way of achieving this.

The structure of an associative unital algebra is therefore not sufficient for the existence of representations on the duals and tensor products of representation spaces and for the existence of a trivial representation on k. To obtain representations on the duals and tensor products of representation spaces, an associative algebra needs to be equipped with additional structure. To derive this structure, we note that we can formalise the definitions for the groups and Lie algebras by introducing maps

$$\begin{aligned} &S_G : G \to G,\ g \mapsto S_G(g) = g^{-1} && \Delta_G : G \to G \otimes G,\ g \mapsto \Delta_G(g) = g \otimes g \\ &S_{\mathfrak{g}} : \mathfrak{g} \to \mathfrak{g}, \boldsymbol{x} \mapsto S_{\mathfrak{g}}(\boldsymbol{x}) = -\boldsymbol{x} && \Delta_{\mathfrak{g}} : \mathfrak{g} \to \mathfrak{g} \otimes \mathfrak{g}, \boldsymbol{x} \mapsto \Delta_{\mathfrak{g}}(\boldsymbol{x}) = 1 \otimes \boldsymbol{x} + \boldsymbol{x} \otimes 1 \end{aligned} \tag{2.60}$$

In terms of these maps, we can unify the definitions of duals and tensor products of representations for groups and Lie algebras as

$$\begin{aligned} &\rho_{V^*}(x)\alpha = \alpha \circ \rho_V \circ S(x) \\ &\rho_{V\otimes W}(x)(\boldsymbol{v} \otimes \boldsymbol{w}) = (\rho_V \otimes \rho_W) \circ \Delta(x)(\boldsymbol{v} \otimes \boldsymbol{w}) \quad \forall \alpha \in V^*, \boldsymbol{v} \in V, \boldsymbol{w} \in W, \end{aligned} \tag{2.61}$$

where x is either an element of a group G or of a Lie algebra $\mathfrak{g}$, and S and Δ are the maps given by (2.60). Note that in order to yield *representations*, of groups and Lie algebras, the maps S and Δ must satisfy

$$S(x \diamond y) = S(y) \diamond S(x) \qquad \Delta(x \diamond y) = \Delta(x) \diamond \Delta(y),$$

where $x \diamond y = [x, y]$ for $x, y \in \mathfrak{g}$ and $x \diamond y = x \cdot y$ for $x, y \in G$. In other words, Δ must be a *homomorphism* of groups or Lie algebras and S an *anti-endomorphism* of groups or Lie algebras. Note also that the fact that the map Δ consistently defines representations on multiple tensor products $U \otimes V \otimes W$ is due to the fact that for both, the case of a group and a Lie algebra, it satisfies the *co-associativity condition*

$$(\Delta \otimes \mathrm{id}) \circ \Delta = (\mathrm{id} \otimes \Delta) \circ \Delta.$$

Moreover, we note that in both, the case of a group and the Lie algebra, the *trivial representation* on the field k is given by a map

$$\begin{aligned} &\epsilon_{\mathfrak{g}} : \mathfrak{g} \to k && \epsilon(\boldsymbol{x}) = 0 \quad \forall \boldsymbol{x} \in \mathfrak{g} \\ &\epsilon_G : G \to k && \epsilon(g) = 1 \quad \forall g \in G. \end{aligned} \tag{2.62}$$

Note that in order to give rise to *representations*, the map ϵ must be a homomorphism of groups or Lie algebras, i.e. satisfy

$$\epsilon(x \diamond y) = \epsilon(x) \cdot \epsilon(y),$$

where again $x \diamond y = [x, y]$ for $x, y \in \mathfrak{g}$ and $x \diamond y = x \cdot y$ for $x, y \in G$. Moreover, the map ϵ which defines the trivial representations must be compatible with with the fact that $k \otimes V \cong V \cong V \otimes k$ for all vector spaces V over k. More precisely, the canonical isomorphisms $\phi_L : k \otimes V \to V$, $t \otimes \boldsymbol{v} \mapsto t\boldsymbol{v}$ and $\phi_R : V \otimes k \to V$, $\boldsymbol{v} \otimes t \mapsto t\boldsymbol{v}$ should be intertwiners between the representations on $k \otimes V$, $V \otimes k$ constructed from a representation on V via the maps ϵ and Δ and the representation on V. This translates into the condition

$$(\epsilon \otimes \mathrm{id}) \circ \Delta = (\mathrm{id} \otimes \epsilon) \circ \Delta = \mathrm{id},$$

which is satisfied trivially for both the case of the group and the Lie algebra.

Finally, we note that for both, groups and Lie algebras the evaluation maps $\mathrm{ev}_L : V^* \otimes V \to k$, $\alpha \otimes \boldsymbol{v} \mapsto \alpha(\boldsymbol{v})$ and $\mathrm{ev}_R : V \otimes V^* \to k$, $\boldsymbol{v} \otimes \alpha \mapsto \alpha(\boldsymbol{v})$ are *intertwiners* between the representations on $V \otimes V^*$ and $V^* \otimes V$ constructed via the maps Δ and S and between the trivial representation given by the map ϵ. The former are given by

$$\rho_{V^* \otimes V}(x) \alpha \otimes \boldsymbol{v} = (\rho_{V^*} \otimes \rho_V)(\Delta(x)) \alpha \otimes \boldsymbol{v} \quad \rho_{V \otimes V^*}(x) \boldsymbol{v} \otimes \alpha = (\rho_V \otimes \rho_{V^*})(\Delta(x)) \boldsymbol{v} \otimes \alpha$$

where the representation ρ_{V^*} is defined as in (2.61), $\alpha \in V^*$ and $\boldsymbol{v} \in V$. The requirement that the evaluation map is an intertwiner between these representations and the trivial representation on k defined via the map ϵ translates into the condition

$$m \circ (S \otimes \mathrm{id}) \circ \Delta(x) = m \circ (\mathrm{id} \otimes S) \circ \Delta(x) = \epsilon(x) e,$$

where e is the unit (e=1 for groups and $e = 0$ for Lie algebras). The map $m : x \otimes y \mapsto x \diamond y$ is the group multiplication $m : g \otimes h \mapsto g \cdot h$ in the case of a group G and the Lie bracket

$m : \boldsymbol{x} \otimes \boldsymbol{y} \mapsto [\boldsymbol{x}, \boldsymbol{y}]$ in the case of a Lie algebra. Clearly, the maps (2.60) and (2.62) satisfy this condition.

We can summarise this discussion as follows:

We consider a structure X which is either a group, a Lie algebra or an associative, unital algebra with a multiplication map $\diamond : X \times X \to X$ which is either the group multiplication, or the Lie bracket or the multiplication of the associative algebra and with a unit e satisfying $e \diamond x = x \diamond e = x$ for all $x \in X$. Representations of X are X-homomorphisms $\rho_V : X \to \mathrm{End}(V)$ into the set of endomorphisms of a vector space V.

In order to construct

1. representations $\rho_{V\otimes W} : X \to \mathrm{End}(V \otimes W)$ on the tensor products $V \otimes W$ of representation spaces for all representations $\rho_V : X \to \mathrm{End}(V)$, $\rho_W : X \to \mathrm{End}(W)$

2. a trivial representation on the field k

such that the following consistency conditions are satisfied

1. The representations $\rho_{U\otimes(V\otimes W)}, \rho_{(U\otimes V)\otimes W} : X \to \mathrm{End}(U \otimes V \otimes W)$ are identical

2. The canonical isomorphisms $k \otimes V \cong V$, $V \otimes k \cong V$ are intertwiners between the representations $\rho_{k\otimes V} : X \to \mathrm{End}(k \otimes V)$, $\rho_{V\otimes k} : X \to \mathrm{End}(V \otimes k)$ and ρ_V

we need X-homomorphisms $\Delta : X \to X \otimes X$ and $\epsilon : X \to k$ and an anti-X-endomorphism $S : X \to X$ that satisfy the following conditions

1. *Co-associativity*: $(\Delta \otimes \mathrm{id}_X) \circ \Delta = (\mathrm{id}_X \otimes \Delta) \circ \Delta$.

2. *Compatibility between ϵ and Δ*: $(\epsilon \otimes \mathrm{id}_X) \circ \Delta = (\mathrm{id}_X \otimes \epsilon) \circ \Delta = \mathrm{id}_X$

If additionally we require

3. The existence of representations $\rho_{V^*} : X \to \mathrm{End}(V^*)$ on the dual vector spaces V^* of all representations $\rho_V : X \to \mathrm{End}(V)$.

such that

3. The evaluation maps $\mathrm{ev}_L : V^* \otimes V \to k$, $\alpha \otimes \boldsymbol{v} \mapsto \alpha(\boldsymbol{v})$ and $\mathrm{ev}_R : V \otimes V^* \to k$, $\boldsymbol{v} \otimes \alpha \mapsto \alpha(\boldsymbol{v})$ are *intertwiners* between the representations $\rho_{V^*\otimes V} : X \to \mathrm{End}(V^* \otimes V)$, $\rho_{V\otimes V^*} : X] \to \mathrm{End}(V^* \otimes V)$ and $\rho_k : X \to \mathrm{End}(k)$

we also need an anti-X-endomorphism $S : X \to X$ that is compatible with Δ, ϵ and $\diamond$: $m \circ (S \otimes \mathrm{id}_X) \circ \Delta = m \circ (\mathrm{id}_X \otimes S) \circ \Delta = \epsilon \cdot e$, where $m : x \otimes y \to x \diamond y$ corresponds to the multiplication map and e is the unit.

By applying these requirements to an associative algebra, we obtain the definition of a Bialgebra and of a Hopf algebra.

Definition 2.4.5: (Bialgebra, Hopf algebra)

A *Bialgebra* is an associative algebra $(A, +, \cdot, \bullet)$ with unit e that carries the following additional structure:

1. A *co-product* $\Delta : A \to A \otimes A$, which is an *algebra homomorphism* with respect to the algebra structure on A and the algebra structure on $A \otimes A$ defined by $(a \otimes b) \bullet (c \otimes d) = ac \otimes bd$ for all $a, b, c, d \in A$. It is required to satisfy the *co-associativity condition* $(\Delta \otimes \mathrm{id}_A) \circ \Delta = (\mathrm{id}_A \otimes \Delta) \circ \Delta : A \to A \otimes A \otimes A$.

2. A *co-unit* $\epsilon : A \to k$, which is an *algebra homomorphism* with respect to the algebra structure on A and the algebra structure on k and which is compatible with the co-product
$$(\epsilon \otimes \mathrm{id}_A) \circ \Delta = (\mathrm{id}_A \otimes \epsilon) \circ \Delta = \mathrm{id}_A$$

A is called a *Hopf algebra* if it is a bialgebra and there exists a bilinear map $S : A \to A$, the *antipode*, which is an *anti-algebra homomorphism*: $S(a \bullet b) = S(b) \bullet S(a)$ $\forall a, b \in A$ and satisfies
$$m \circ (S \otimes \mathrm{id}_A) \circ \Delta(a) = m \circ (\mathrm{id}_A \otimes S) \circ \Delta(a) = \epsilon(a) \cdot e,$$
where m denotes the algebra multiplication $m(a \otimes b) := a \bullet b$ for all $a, b \in A$. In other words, if we express $\Delta(a)$ in a basis $\{e_i\}_{i \in I}$ of A, $\Delta(a) = \sum_{i,j \in I} c_{ij}(a) e_i \otimes e_j$, $c_{ij(a) \in k}$ we have
$$\sum_{i,j \in I} c_{ij}(a) S(e_i) \bullet e_j = \sum_{i,j \in I} c_{ij}(a) e_i \bullet S(e_j) = \epsilon(a) e.$$

The discussion above can be encoded in the following theorem.

Theorem 2.4.6: (Bialgebras and Hopf algebras - construction of representations)

Let A be a bialgebra and $\rho_V : A \to \mathrm{End}(V)$, $\rho_W : A \to \mathrm{End}(W)$ representations of A on vector spaces V, W. Then
$$\rho_{V \otimes W}(a)(\boldsymbol{v} \otimes \boldsymbol{w}) = (\rho_V \otimes \rho_W) \circ \Delta(a)(\boldsymbol{v} \otimes \boldsymbol{w}) \qquad \forall a \in A, \boldsymbol{v} \in V, \boldsymbol{w} \in W.$$
defines a representation of A on the tensor product $V \otimes W$ and co-unit defines a representation of A on the field k
$$\rho_k(a) = \epsilon(a) \qquad \forall a \in A.$$
If A is a Hopf algebra, a representation on the dual vector space V^* of V is given by
$$\rho_{V^*}(a)\alpha = \alpha \circ \rho_V(S(a)) \qquad \forall a \in A, \alpha \in V^*.$$
These representations satisfy $\rho_{(U \otimes V) \otimes W} = \rho_{U \otimes (V \otimes W)}$ for all representation spaces U, V, W and are such that the canonical isomorphisms $k \otimes V \cong V$, $V \otimes k \cong V$ and the evaluation maps $V \otimes V^* \to k$, $V^* \otimes V \to k$ are intertwiners.

Proof: The idea of the proof is given in the discussion above. Reformulating it in a concise way is left as an exercise. □

2.4.3 Tensor product decomposition

Example 2.4.7: (Tensor product for $\mathfrak{su}(2)$, Clebsch Gordan coefficients)

We consider the tensor product of two representations $\rho_V : \mathfrak{su}(2) \to \mathrm{End}(V)$, $\rho_W : \mathfrak{su}(2) \to \mathrm{End}(W)$. We characterise these representations by the dimensions d_V, d_W or, equivalently, by the eigenvalue of the quadratic Casimir operator $\rho_V(J^2) = \frac{1}{2}j_V(j_V+1)\cdot 1_V$, $\rho_W(J^2) = \frac{1}{2}j_W(j_W+1)\cdot 1_W$ with $d_V = 2j_V+1$, $d_W = 2j_W+1$.

A basis of V is then given by the eigenvectors of J_z: $B_V = \{|j_V, m_V\rangle \mid m_V = -j_V, ..., j_V\}$, $\rho_V(J_z)|j_V, m_V\rangle = m_V|j_V, m_V\rangle$ and similarly for W $B_W = \{|j_W, m_W\rangle \mid m_W = -j_W, ..., j_W\}$, $\rho_W(J_z)|j_W, m_W\rangle = m_W|j_V, m_W\rangle$.

1. Tensor product of representations

All elements of the tensor product basis $|j_V, m_V\rangle \otimes |j_W, m_W\rangle$ are eigenstates of $\rho_{V\otimes W}(J_z)$

$$\begin{aligned}\rho_{V\otimes W}(J_z)|j_V, m_V\rangle \otimes |j_W, m_W\rangle =&(\rho_V(J_z)|j_V, m_V\rangle) \otimes |j_W, m_W\rangle + |j_V, m_V\rangle \otimes (\rho_W(J_z)|j_W, m_W\rangle)\\ =&(m_V + m_W)|j_V, m_V\rangle \otimes |j_W, m_W\rangle\end{aligned}$$

However, they are in general *not* eigenstates of the quadratic Casimir operator:

$$\begin{aligned}&\rho_{V\otimes W}(J^2)|j_V, m_V\rangle \otimes |j_W, m_W\rangle =\\ &= \sum_{i=x,y,z} (\rho_V(J_i)\otimes 1 + 1\otimes \rho_W(J_i))(\rho_V(J_i)\otimes 1 + 1\otimes\rho_W(J_i))|j_V, m_V\rangle \otimes |j_W, m_W\rangle \qquad (2.63)\\ &= \tfrac{1}{2}(j_V(j_V+1) + j_W(j_W+1) + 2m_W m_V)|j_V, m_V\rangle \otimes |j_W, m_W\rangle\\ &\quad + \tfrac{1}{2}\sqrt{j_V(j_V+1) - m_V(m_V+1)}\sqrt{j_W(j_W+1) - m_w(m_w-1)}|j_V, m_V+1\rangle \otimes |j_W, m_W-1\rangle\\ &\quad + \tfrac{1}{2}\sqrt{j_V(j_V+1) - m_V(m_V-1)}\sqrt{j_W(j_W+1) - m_w(m_w+1)}|j_V, m_V-1\rangle \otimes |j_W, m_W+1\rangle\end{aligned}$$

With Schur's lemma, this implies that the representation $\rho_{V\otimes W}$ is reducible.

2. Tensor product decomposition

We now want to decompose the tensor product $\rho_{V\otimes W}$ into irreducible representations. In other words, we want to find a basis of $V\otimes W$ that consists of eigenvectors $|J, M\rangle$ of $\rho_{V\otimes W}(J^2)$ and $\rho_{V\otimes W}(J_z)$ satisfying

$$\rho_{V\otimes W}(J^2)|J, M\rangle = \tfrac{1}{2}J(J+1)|J, M\rangle \qquad \rho_{V\otimes W}|J, M\rangle = M|J, M\rangle \qquad M =\in \{-J, ..., J\}.$$

From the action of J^2, one can see that the states in the tensor product basis are eigenstates of J^2 if and only if either $m_V = j_V$, $m_W = j_W$ or $m_V = -j_V$, $m_W = -j_W$ and the corresponding eigenvalues are $M = \pm(j_V + j_W)$, $J = j_V + j_W$.

As the states in the tensor product basis are eigenstates of $\rho_{V\otimes W}(J_z)$ with eigenvalues $M = m_V + m_W$ and $m_V \in \{-j_V, ..., j_V - 1, J_V\}$, $m_W \in \{-j_W, ..., j_W - 1, J_W\}$, we have $M \in \{-(j_V + j_W), ..., j_V + j_W - 1, j_V + j_W\}$. Moreover, equation (2.63) implies

$$\langle j_V, m_V| \otimes \langle j_W m_W|\rho_{V\otimes W}(J^2)|j_V, m_V\rangle \otimes |j_W, m_W\rangle = \tfrac{1}{2}(j_V(j_V+1) + j_W(j_W+1) + 2m_V m_W)$$

and therefore

$$\tfrac{1}{2}(|j_V - j_W|)(|j_V - j_W| + 1) \le \tfrac{1}{2}(j_V(j_V+1) + j_W(j_W+1) + 2m_V m_W) \le \tfrac{1}{2}(j_V + j_W)(j_V + j_W + 1).$$

This implies that the expectation value of J^2 in the representation $\rho_{V\otimes W}$ takes values in

$$\langle J, M|\rho_{V\otimes W}(J^2)J, M\rangle \in [\tfrac{1}{2}(|j_V - j_W|)(|j_V - j_W| + 1), \tfrac{1}{2}(j_V + j_W)(j_V + j_W + 1)],$$

and therefore $J \in \{-|j_V - j_W|, -|j_V + j_W| + 1, ..., j_V + j_W - 1, j_V + j_W\}$. One can show that each value $J \in \{-|j_V - j_W|, -|j_V + j_W| + 1, ..., j_V + j_W - 1, j_V + j_W\}$ arises exactly once and the tensor product decomposition of the representations ρ_V and ρ_W is given by

$$V \otimes W = \bigoplus_{J=-|j_V-j_W|}^{j_V+j_W} V_J,$$

where V_J are linear subspaces of $V \otimes W$ which are invariant under the action $\rho_{V\otimes W}$ of $\mathfrak{su}(2)$ and carry an irreducible representation of $\mathfrak{su}(2)$ of dimension $d_J = 2J + 1$.

3. Clebsch Gordan coefficients

We will now determine the basis transformation from the tensor product basis $\{|j_V, m_V\rangle \otimes |j_W, m_W\rangle\}$ to the basis $\{|J, M\rangle\}$. As the states in the tensor product basis are eigenstates of $\rho_{V\otimes W}(J_z)$, the eigenstate $|J, M\rangle$ must be given as a linear combination of vectors in

$$B_{J,M} = \{|j_V, m_V\rangle \otimes |j_W, m_W\rangle \mid m_V \in \{-j_V, ..., j_V\}, m_W \in \{-j_W, ..., j_W\}, m_V + m_W = M\}.$$

This implies in particular that the eigenvalue of the quadratic Casimir operator in the tensor product takes values $\frac{1}{2}J(J + 1)$ with

$$J \in \{|j_V - j_W|, |j_V + j_W| + 1, ..., j_V + j_W - 1, j_V + j_W\}.$$

The coefficients in this linear combination are called *Clebsch-Gordan coefficients* and are usually denoted $\langle j_V, m_V, j_W, m_W|J, M\rangle$.

$$|J, M\rangle = \sum_{m_V=-j_V}^{j_V} \sum_{m_W=-j_W}^{j_W} \langle j_V, m_V, j_W, m_W|J, M\rangle\ |j_V, m_V\rangle \otimes |j_W, m_W\rangle. \tag{2.64}$$

Clearly, the Clebsch Gordan coefficients satisfy

$$\begin{aligned}&\langle j_V, m_V, j_W, m_W|J, M\rangle = \delta_{M,m_V+m_W}\langle j_V, m_V, j_W, m_W|J, m_V + m_W\rangle\\ &\langle j_V, m_V, j_W, m_W|J, M\rangle = 0 \quad \text{if} \quad J \notin \{|j_V - j_W|, |j_V - j_W| + 1, ..., j_V + j_W\}.\end{aligned}$$

By applying $\rho_{V\otimes W}(J_\pm)$ to both sides of the equation, we find the following *recursion relation*

$$\begin{aligned}\sqrt{J(J+1) - M(M\pm 1)}|J, M\pm 1\rangle =&\sqrt{j_V(j_V+1) - m_V(m_V \mp 1)}\langle j_V, m_V \mp 1, j_W, m_W|J, M\rangle\\ &+\sqrt{j_W(j_W+1) - m_W(m_W \mp 1)}\langle j_V, m_V, j_W, m_W \mp 1|J, M\rangle.\end{aligned}$$

Together with a *phase convention*, namely that the Clebsch Gordan coefficients $\langle j_V, j_V, j_W, j_V - J|JJ\rangle$ are real and positive for all $J \in \{|j_V - j_W|, |j_V - j_W| + 1, ..., j_V + j_W\}$, this recursion relation determines the Clebsch Gordan coefficients uniquely, and one finds that they are all real.

We also note that the Clebsch-Gordan coefficients are a basis transformation between two orthonormal bases of the representation space $V \otimes W$, the tensor product basis $\{|j_V, m_V\rangle \otimes |j_W, m_W\rangle\}$ and the basis $\{|J, M\rangle\}$. They are therefore coefficients of a *unitary matrix* which, by the choice of the phase convention, has only real entries. The relation $A^\dagger A = A^T A = 1$ for a unitary matrix with real entries is equivalent to the *orthogonality relations* for Clebsch-Gordan coefficients

$$\sum_{m_V=-j_V}^{j_V} \sum_{m_W=-j_W}^{j_W} \langle j_V, m_V, j_W, m_W | J, M\rangle\langle j_V, m_V, j_W, m_W | J', M'\rangle = \delta_{JJ'}\delta_{MM'}$$

$$\sum_{J=|j_V-j_W|}^{j_V+j_W} \sum_{M=-J}^{J} \langle j_V, m_V, j_W, m_W | J, M\rangle\langle j_V, m'_V, j_W, m'_W | J, M\rangle = \delta_{m_V,m'_V}\delta_{m_W,m'_W}.$$

Chapter 3

Special Relativity

3.1 Minkowski metric and Lorentz group

Definition 3.1.1: (Minkowski metric)

We consider the vector space $\mathbb{R}^4$ with standard basis $B = \{e_0, e_1, e_2, e_3\}$

$$e_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \qquad e_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \qquad e_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \qquad e_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

The *Minkowski metric* is the non-degenerate bilinear form $g : \mathbb{R}^4 \times \mathbb{R}^4 \to \mathbb{R}$ of signature $(-1, 1, 1, 1)$ given by

$$g(e_0, e_0) = -1 \qquad g(e_i, e_i) = 1\ \forall i \in \{1, 2, 3\} \qquad g(e_i, e_j) = 0\ \forall i, j \in \{0, 1, 2, 3\}, i \neq j. \quad (3.1)$$

In the following, we will use the notation M_g for the matrix with entries $(M_g)_{ij} = g(e_i, e_j)$. We also write $\boldsymbol{x} \cdot \boldsymbol{y}$ for $g(\boldsymbol{x}, \boldsymbol{y})$ and $\boldsymbol{x}^2$ for $g(\boldsymbol{x}, \boldsymbol{x})$.

Definition 3.1.2: (Timelike, spacelike lightlike, future directed, past directed)

A vector $\boldsymbol{x} \in \mathbb{R}^4$ is called *timelike* if $g(\boldsymbol{x}, \boldsymbol{x}) < 0$, *lightlike* if $g(\boldsymbol{x}, \boldsymbol{x}) = 0$, $\boldsymbol{x} \neq 0$ and *spacelike* if $g(\boldsymbol{x}, \boldsymbol{x}) > 0$. A timelike or lightlike vector $\boldsymbol{x} \in \mathbb{M}^4$ is called *future directed* (*past directed*) if $g(\boldsymbol{x}, e_0) < 0$ ($g(\boldsymbol{x}, e_0) > 0$).

Exercise 25: (Cauchy Schwartz inequality for Lorentzian signature)

1. Show that for any two future directed, timelike vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^4$, we have $\boldsymbol{x} \cdot \boldsymbol{y} < 0$. $|\boldsymbol{x} \cdot \boldsymbol{y}| \geq \sqrt{|\boldsymbol{x}^2|}\sqrt{|\boldsymbol{y}^2|}$ and $|\boldsymbol{x} \cdot \boldsymbol{y}| = \sqrt{|\boldsymbol{x}^2|}\sqrt{|\boldsymbol{y}^2|}$ if and only if $\boldsymbol{x} = \alpha \boldsymbol{y}$, $\alpha \in \mathbb{R}^+$.

2. Show that for any two spacelike vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^4$, the Cauchy Schwartz inequality holds in its usual form $|\boldsymbol{x} \cdot \boldsymbol{y}| \leq \sqrt{\boldsymbol{x}^2}\sqrt{\boldsymbol{y}^2}$ and equality holds if and only if $\boldsymbol{x}, \boldsymbol{y}$ are linearly dependent $\boldsymbol{x} = \alpha \boldsymbol{y}$, $\alpha \in \mathbb{R} \setminus \{0\}$.

3. Show that for any two lightlike vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^4$, we have $\boldsymbol{x} \cdot \boldsymbol{y} \leq 0$ if and only if both are future directed or both are past directed and $\boldsymbol{x} \cdot \boldsymbol{y} \geq 0$ if and only if one is past directed and one future directed.

Definition 3.1.3: (Lorentz group)

The *Lorentz group* $SO(3,1)$ is the group of matrices $B \in M(4,\mathbb{R})$ that preserve the Minkowski metric

$$g(B\boldsymbol{x}, B\boldsymbol{y}) = g(\boldsymbol{x}, \boldsymbol{y}) \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}. \tag{3.2}$$

Lemma 3.1.4: Any element of the Lorentz group can be expressed uniquely as

$$\begin{aligned} M^{\pm}(R,\vec{w}) &= \begin{pmatrix} \pm\sqrt{1+|\vec{w}|^2} & \pm(R^T\vec{w})^T \\ \vec{w} & \left(\frac{\sqrt{1+|\vec{w}|^2}-1}{|\vec{w}|^2}\vec{w}\cdot\vec{w}^T + 1_3\right)\cdot R \end{pmatrix} \\ &= \begin{pmatrix} \pm\sqrt{1+|\vec{w}|^2} & \pm\vec{w}^T \\ \vec{w} & \left(\frac{\sqrt{1+|\vec{w}|^2}-1}{|\vec{w}|^2}\vec{w}\cdot\vec{w}^T + 1_3\right) \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & R \end{pmatrix} \\ &\text{with } R \in O(3),\ \vec{w} \in \mathbb{R}^3,\ |\vec{w}|^2 = \sum_{i=1}^3 (w^i)^2. \end{aligned} \tag{3.3}$$

1. An element of the Lorentz group parametrised as in (3.3) is called *time orientation preserving* if it has $m_{00} = \sqrt{1+|\vec{w}|^2}$. Otherwise it is called *time orientation reversing.* The element $T = \text{diag}(-1,1,1,1)$ is called *time reversal.*

2. An element of the Lorentz group parametrised as in (3.3) is called *orientation preserving* if $\det R = 1$, i.e. $R \in SO(3)$. Otherwise it is called *orientation reversing.* The element $P = \text{diag}(1,-1,-1,-1)$ is called *parity transformation.*

3. The elements which are both orientation and time orientation preserving form a subgroup $SO^{+,\uparrow}(3,1) \subset SO(3,1)$ which is a connected component of $SO(3,1)$. This subgroup is called the *proper orthochronous Lorentz group.*

4. An element $M^+(1,\vec{w}) \in SO^{+,\uparrow}(3,1)$ parametrised as in (3.3) with $R = 1$ is called a *boost in the direction of* $\vec{w}$. The parameter θ defined by $|\vec{w}| = \sinh\theta$ is called the *rapidity* of the boost.

5. An element $M^+(R,0) \in SO^{+,\uparrow}(3,1)$ parametrised as in (3.3) is called a *rotation* if $\vec{w} = 0$ and $R \in SO(3)$. The rotations form the subgroup $SO(3) \subset SO^{+,\uparrow}(3,1)$.

6. Any element $L \in SO(3,1)$ of the Lorentz group can be expressed uniquely as $L = T^p P^q BR$, where $p, q \in \{0,1\}$, T is the time reversal, P the parity transformation B a boost and $R \in SO(3)$ a rotation.

7. Any element $L \in SO^{+,\uparrow}(3,1)$ of the proper orthochronous Lorentz group can be expressed uniquely as a product $L = BR$, where B is a boost and R a rotation. This is called the *polar decomposition.*

Proof: Exercise □

Remark 3.1.5: The Lorentz group is *not compact.* In physics, this statement corresponds to the fact that one cannot boost a massive particle to the velocity of light. By considering the norm $||M|| = \text{Tr}\left(M^\dagger M\right) = \sqrt{\sum_{i,j=0}^3 m_{ij}^2}$ and a sequence of boosts $B_n = M^+(1, n\cdot e_1) \in$

$SO^{+,\uparrow}(3,1)$, one finds $||B_n||^2 = 4\sqrt{1+n^2}$. Hence, the sequence cannot have a convergent subsequence, and the Lorentz group is non-compact. The non-compactness of the Lorentz group has important consequences for its representation theory.

Exercise 26: (The Lorentz group, boosts and rotations)

1. Show that if $\boldsymbol{x} = \sum_{i=0}^{3} x^i e_i$, $\boldsymbol{y} = \sum_{i=0}^{3} y^i e_i$, the metric $g(\boldsymbol{x}, \boldsymbol{y})$ is given by

$$g(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^T \cdot M_g \cdot \boldsymbol{y} = \begin{pmatrix} x^0 & x^1 & x^2 & x^3 \end{pmatrix} \cdot \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} y^0 \\ y^1 \\ y^2 \\ y^3 \end{pmatrix}.$$

Show that a matrix $B \in M(4, \mathbb{R})$ is an element of the Lorentz group if and only if

$$M^T \cdot M_g \cdot M = M_g. \tag{3.4}$$

2. Consider a matrix M of the form

$$M = \begin{pmatrix} m_{00} & \vec{v}^T \\ \vec{w} & S \end{pmatrix} \quad \text{with } S \in M(3, \mathbb{R}), \vec{v}, \vec{w} \in \mathbb{R}^3, m_{00} \in \mathbb{R}. \tag{3.5}$$

Show that this matrix is an element of the Lorentz group if and only if

$$m_{00}^2 = 1 + |\vec{w}|^2 \quad S^T S = 1 + \vec{v} \cdot \vec{v}^T \quad -m_{00}\vec{v} + S^T \vec{w} = 0 \quad \text{where } |\vec{w}|^2 = \sum_{i=1}^{3} (w^i)^2.$$

3. Consider the 3×3-matrix

$$B = \lambda \vec{w} \cdot \vec{w}^T + \mu 1_3 = \lambda \begin{pmatrix} w_1^2 & w_1 w_2 & w_1 w_3 \\ w_1 w_2 & w_2^2 & w_2 w_3 \\ w_1 w_3 & w_2 w_3 & w_3^2 \end{pmatrix} + \mu \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

where $\vec{w} \in \mathbb{R}^3$, $\lambda, \mu \in \mathbb{R}$. Show that it satisfies $B^T B = 1$ if and only if $\mu = 1$ and $\lambda = 0$ or $\mu = 1$ and $\lambda = -\frac{2\mu}{|\vec{w}|^2}$. Hint: Use the associativity of the matrix multiplication to show $\vec{w}^T \cdot (\vec{w} \cdot \vec{w}^T) = |\vec{w}|^2 \vec{w}^T$, $(\vec{w} \cdot \vec{w}^T) \cdot \vec{w} = |\vec{w}|^2 \vec{w}$

4. Set $S = (\lambda \vec{w} \cdot \vec{w}^T + \mu 1_3) \cdot R$ and show that the matrix M in (3.5) is an element of the Lorentz group if and only if

$$R \in O(3) \qquad \vec{v} = R^T \vec{w} \qquad \lambda = \frac{\sqrt{|\vec{w}|^2 + 1} - 1}{|\vec{w}|^2} \qquad \mu = 1 \quad m_{00} = \sqrt{1 + |\vec{w}|^2}. \tag{3.6}$$

Exercise 27: (Orbits of the Lorentz group)

1. For a representation $\rho : G \to \mathrm{End}(V)$ of a group G on a vector space V, the *orbit* of a vector $\boldsymbol{x} \in V$ is the set $O_{\boldsymbol{x}} = \{\rho(g)\boldsymbol{x} \in V \mid g \in G\}$. Show that the orbits of the representations of the proper orthochronous Lorentz group $SO^{+,\uparrow}(3,1)$ on $\mathbb{R}^4$ are given by:

1. $\{0\}$

2. $O_T^{\pm} = \{\boldsymbol{y} \in \mathbb{R}^4 : y_0^2 - y_1^2 - y_2^2 - y_3^2 = T^2\ ,\ \pm y^0 > 0\ T > 0\}$

3. $O_0^{\pm} = \{\boldsymbol{y} \in \mathbb{R}^4 : y_0^2 - y_1^2 - y_2^2 - y_3^2 = 0\ ,\ \pm y^0 > 0\}$
4. $O_{iT} = \{\boldsymbol{y} \in \mathbb{R}^4 : y_0^2 - y_1^2 - y_2^2 - y_3^2 = -T^2\ ,\ T > 0\}$

Draw these orbits. The orbits $O_T^{\pm}$ are called *two shell hyperboloids*. The orbits O_{iT} are called *single shell hyperboloids*. The orbits O_0^+ and O_0^- are called, respectively, *future lightcone* and *past lightcone*.

2. Show that the tangent space to $O_T^{\pm}$ in $\boldsymbol{x} \in O_T^{\pm}$ is given by $\boldsymbol{x}^{\perp} = \{\boldsymbol{y} \in \mathbb{R}^4 \mid g(\boldsymbol{x}, \boldsymbol{y}) = 0\}$. Show that any tangent vector on $O_T^{\pm}$ is *spacelike*.

3. We consider the intersection $O_T^+ \cap \mathrm{Span}(\{e_0, \vec{w}\})$. Show that this is a hyperbola parametrised by $h(t) = \cosh(t) e_0 + \sinh(t) \vec{w}$. Show that for any $\boldsymbol{x} \in \mathrm{Span}(\{e_0, \vec{w}\})$, the rapidity of the boost $M(\vec{w}, R = 1)$ is equal to the length of the segment from $\boldsymbol{x}$ to $M\boldsymbol{x}$ on the hyperbola h.

Hint: The length of a curve $c : [0, 1] \to \mathbb{R}^4$, $c(0) = \boldsymbol{x}$, $c(1) = \boldsymbol{y}$ is given by

$$l(c) = \int_0^1 dt \sqrt{g(\dot{c}(t), \dot{c}(t))} \tag{3.7}$$

Lemma 3.1.6: (Lorentz algebra)

The Lie algebra of the Lorentz group is a six-dimensional Lie algebra $\mathfrak{so}(3,1)$ over $\mathbb{R}$, in which the generators B_i, $i = 1, 2, 3$, corresponding to the boosts are given by

$$B_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad B_2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad B_3 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \tag{3.8}$$

and the generators corresponding to rotations take the form

$$R_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix} \quad R_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \quad R_3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{3.9}$$

Its Lie bracket is given by

$$[B_i, B_j] = \epsilon_{ijk} R_k \qquad [R_i, R_j] = -\epsilon_{ijk} R_k \qquad [B_i, R_j] = -\epsilon_{ijk} B_k \qquad i, j, k \in \{1, 2, 3\}, \tag{3.10}$$

where ϵ_{ijk} is the totally antisymmetric tensor with $\epsilon_{123} = 1$. The exponential map $\exp : \mathfrak{so}(3,1) \to SO^{+,\uparrow}(3,1)$ is surjective. The Lorentz group has two quadratic Casimir operators, which are given by

$$R^2 + B^2 = \sum_{i=1}^{3} R_i^2 + B_i^2 \qquad\qquad RB = \sum_{i=1}^{3} R_i B_i. \tag{3.11}$$

Proof: To determine the Lie algebra of the Lorentz group, we parametrise elements of $SO^{+,\uparrow}(3,1)$ as in (3.3). we consider the paths in $SO^{+,\uparrow}(3,1)$ which are given by

$$b_i(t) = M^+(1, t \cdot e_i) \qquad r_i(t) = M^+(R_i(t), 0) \tag{3.12}$$

$$R_1(t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos t & \sin t \\ 0 & -\sin t & \cos t \end{pmatrix} \; R_2(t) = \begin{pmatrix} \cos(t) & 0 & -\sin t \\ 0 & 1 & 0 \\ \sin t & 0 & \cos t \end{pmatrix} \; R_3(t) = \begin{pmatrix} \cos t & \sin t & 0 \\ -\sin t & \cos t & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Clearly, $b_i(0) = r_i(0) = 1$ for all $i = 1, 2, 3$. Setting $B_i = \frac{d}{dt}|_{t=0} b_i(t)$ and $R_i = \frac{d}{dt}|_{t=0} b_i(t)$, we obtain (3.8) and (3.9). The Lie bracket then follows straightforwardly by calculating the commutators of these generators. The Casimir operators can be determined by straightforward calculation. □

Exercise 28: (Lorentz group and $SL(2,\mathbb{C})$)

1. Show that any hermitian matrix $X \in M(2,\mathbb{C})$ can be parametrised uniquely as

$$X = x^0 \cdot 1_2 + \sum_{i=1}^{3} x^i \sigma_i = \begin{pmatrix} x^0 + x^3 & x^1 - ix^2 \\ x^1 + ix^2 & x^0 - x^3 \end{pmatrix} \qquad x^i \in \mathbb{R}, \tag{3.13}$$

where σ_i, $i = 1, 2, 3$ are the Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3.14}$$

Conclude that the map $\phi : \mathbb{R}^4 \to H(2,\mathbb{C})$

$$\boldsymbol{x} = \sum_{i=0}^{3} x^i e_i \mapsto \phi(\boldsymbol{x}) = \begin{pmatrix} x^0 + x^3 & x^1 - ix^2 \\ x^1 + ix^2 & x^0 - x^3 \end{pmatrix}$$

from $\mathbb{R}^4$ into the *real* vector space of hermitian 2×2 matrices is bijective. Show that with this parametrisation

$$\det(\phi(\boldsymbol{y})\phi(\boldsymbol{x})) = x^0 y^0 - x^1 y^1 - x^2 y^2 - x^3 y^3 = -\boldsymbol{x}\boldsymbol{y}.$$

2. Show that for any $L \in GL(2,\mathbb{C})$ and any $X \in H(2,\mathbb{C})$, the matrix $LXL^\dagger$ is hermitian. Show that $\det(LXL^\dagger) = \det(X)$ for all $X \in H(2,\mathbb{C})$ implies $\det(L) = 1$. Show that $LXL^\dagger = X$ for all $X \in H(2,\mathbb{C})$ implies $L = \pm 1$.

3, Conclude that the the group

$$PSL(2,\mathbb{C}) = SL(2,\mathbb{C})/\{\pm 1\} = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \,\middle|\, a, b, c, d \in \mathbb{C},\; ad - bc = 1 \right\} / \{\pm 1\}$$

is isomorphic to a subgroup of the Lorentz group $SO(3,1)$. Show that $\phi(P\boldsymbol{x}) = \det(\phi(\boldsymbol{x})) \cdot \phi(\boldsymbol{x})^{-1}$ and that for all $L \in SL(2,\mathbb{C})$ we have $\mathrm{Tr}\left(LXL^\dagger\right) > 0$ if and only if $\mathrm{Tr}\,(X) > 0$. Conclude that the group $PSL(2,\mathbb{C})$ is isomorphic to a subgroup of the proper orthochronous Lorentz group $SO^{+,\uparrow}(3,1)$.

4. Show that the transformations

$$X(\boldsymbol{x}) \mapsto BX(\boldsymbol{x})B^\dagger \qquad B = \exp(\tfrac{1}{2} n^i \sigma_i) \in SL(2,\mathbb{C}), \vec{n} \in \mathbb{R}^3$$

correspond to boosts acting on $\boldsymbol{x} \in \mathbb{R}^4$ and the transformations

$$X(\boldsymbol{x}) \mapsto RX(\boldsymbol{x})R^\dagger \qquad R = \exp(-\tfrac{i}{2} n^i \sigma_i) \in SL(2,\mathbb{C}), \vec{n} \in \mathbb{R}^3$$

correspond to rotations acting on $\boldsymbol{x}$. Conclude that the proper orthochronous Lorentz group $SO^{+,\uparrow}(3,1)$ is isomorphic to $PSL(2,\mathbb{C})$.

3.2 Minkowski space

Definition 3.2.1: (Affine space)

An *affine space* $\mathbb{A}$ over a vector space V is a set $\mathbb{A}$ together with a vector space V and a map $\phi : \mathbb{A} \times V \to \mathbb{A}$, $(a, \boldsymbol{v}) \mapsto \phi(a, \boldsymbol{v})$ such that

1. For all $a \in \mathbb{A}$, the map $\phi^a = \phi(a, \cdot) : V \to \mathbb{A}$, $\boldsymbol{v} \mapsto \phi(a, \boldsymbol{v})$ is a bijection.
2. For all $a \in \mathbb{A}$, $\boldsymbol{v}, \boldsymbol{w} \in V$, $\phi(\phi(a, \boldsymbol{v}), \boldsymbol{w}) = \phi(a, \boldsymbol{v} + \boldsymbol{w})$. In other words: the map $\phi : \mathbb{A} \times V \to \mathbb{A}$ gives rise to an action of the abelian group $(V, +)$ on $\mathbb{A}$.

Lemma 3.2.2: (Vector spaces and affine spaces)

1. Let $\mathbb{A}, \mathbb{B}$ affine spaces over a vector space V with maps $\phi_A : \mathbb{A} \times V \to \mathbb{A}$ and $\phi_B : \mathbb{B} \times V \to \mathbb{B}$. Then there exists a bijection $f : \mathbb{A} \to \mathbb{B}$ such that
$$\phi_B(f(a), \boldsymbol{v}) = f(\phi_A(a, \boldsymbol{v})) \qquad \forall a \in \mathbb{A}, \boldsymbol{v} \in V. \tag{3.15}$$
2. For any vector space V, V becomes an affine space over itself if we set $\phi : V \times V \to V$, $\phi(\boldsymbol{v}, \boldsymbol{w}) = \boldsymbol{v} + \boldsymbol{w}$ for all $\boldsymbol{v}, \boldsymbol{w} \in V$.

Proof: Exercise.

Remark 3.2.3: (Affine spaces and vector spaces)

The first condition in Definition 3.2.1 tells us that the choice of an element $a \in \mathbb{A}$ allows us to identify $\mathbb{A}$ and V via the map $\phi(a, \cdot) : V \to \mathbb{A}$. As a consequence of this, any two affine spaces over a vector space V can be identified via bijections defined as in (3.15). Note, however, that this identification is not canonical.

The second statement in Lemma 3.2.2 then tells us that we can identify any affine space with the set $\mathbb{R}^n$ or $\mathbb{C}^n$ viewed as an affine space over the vector space $\mathbb{R}^n$ or $\mathbb{C}^n$. The fundamental difference between a vector space and an affine space is that a vector space has a *distinguished element*, the null vector, whereas an affine space does not. Roughly speaking: "an affine space is a vector space who has forgotten about its origin".

In the following, we omit the map $\phi : \mathbb{A} \times V \to \mathbb{A}$ and write $a + \boldsymbol{v}$ for $\phi(a, \boldsymbol{v})$ for any $a \in \mathbb{A}$, $\boldsymbol{v} \in V$. We write $a - b$ for the unique element $\boldsymbol{v} \in V$ such that $b = \phi(a, \boldsymbol{v})$.

Remark 3.2.4: Via its identification with the underlying vector space and the identification of the vector space with $\mathbb{R}^n$, any affine space $\mathbb{A}$ over a vector space V, $\dim(V) = n$ can be viewed as a manifold isomorphic to $\mathbb{R}^n$. The tangent space at any point $a \in \mathbb{A}$ is isomorphic to $V \cong \mathbb{R}^n$.

Definition 3.2.5: (Affine subspaces, hyperplanes, lines)

Let $\mathbb{A}$ be an affine space over a vector space V. An *affine subspace* or *k-dimensional affine hyperplane* of $\mathbb{A}$ is a subset $\mathbb{B} \subset \mathbb{A}$ of the form

$$\mathbb{B} = a + W = \{a + \boldsymbol{w} \mid a \in \mathbb{A}, \boldsymbol{w} \in W \subset V\}$$

where $W \subset V$ is a linear subspace of V of dimension k. An affine subspace is called *straight line* if it is an affine hyperplane of dimension 1, i.e. $W = \text{Span}(\{\boldsymbol{w}\})$, $\boldsymbol{w} \in V \setminus \{0\}$.

The transformations which preserve the structure of a vector space V are *linear maps* $f : V \to V$. The transformations which preserve affine spaces are *affine maps* or *affine transformations.*

Definition 3.2.6: (Affine transformation)

Let $\mathbb{A}$ be an affine space over a vector space V. An *affine transformation* is a map $f : \mathbb{A} \to \mathbb{A}$ for which there exists an $a \in \mathbb{A}$, $M \in \text{End}(V)$ and $\boldsymbol{v} \in V$ such that

$$f(b) = a + \boldsymbol{v} + M(b - a) \qquad \forall b \in \mathbb{A}. \tag{3.16}$$

Bijective linear transformations of a vector space V are characterised by the fact that they map linear subspaces of dimension one to linear subspaces of dimension one. A similar characterisation exists for bijective affine transformations.

Theorem 3.2.7: A map $f : \mathbb{A} \to \mathbb{A}$ is called a *collineation* if for all $a, b, c \in \mathbb{A}$ which lie on a straight line, $f(a), f(b), f(c)$ lie on a straight line. A map $f : \mathbb{A} \to \mathbb{A}$ is a bijective collineation if and only if f is a bijective affine transformation, i.e. there exists an $a \in \mathbb{A}$, $M \in \text{Aut}(V)$, $\boldsymbol{v} \in V$ such that

$$f(b) = a + \boldsymbol{v} + M(b - a) \qquad \forall b \in \mathbb{A}.$$

Remark 3.2.8: Note that for any affine transformation characterised as in (3.16), we have $f(a) = a + \boldsymbol{v}$. Given another element $c \in \mathbb{A}$, we can reexpress (3.16) as

$$f(b) = a + \boldsymbol{v} + M(b - a) = c + \boldsymbol{v}' + M(b - c) \qquad \boldsymbol{v}' = M(c - a) - (c - a) + \boldsymbol{v} \qquad \forall b \in \mathbb{A}.$$

Hence it is possible to choose a fixed element $a \in \mathbb{A}$ and express *all affine transformations* with respect to this element. With the choice of such an element $a \in \mathbb{A}$ an affine transformation is then characterised uniquely by the associated linear map $M \in \text{End}(V)$ and the vector $\boldsymbol{v} \in V$. Hence, we can identify the set of affine transformations with the set of tuples $(M, \boldsymbol{v})$, $M \in \text{End}(V)$, $\boldsymbol{v} \in V$.

We are now ready to investigate the *structure* of the set of bijective affine transformations and find that they form a group.

Lemma 3.2.9: The affine bijections form a group with respect to their composition, namely the semidirect product $\text{Aut}(V) \ltimes V$. In other words: given an element $a \in \mathbb{A}$ and two affine transformations $f_1, f_2 : \mathbb{A} \to \mathbb{A}$ characterised with respect to $a \in \mathbb{A}$ by linear maps $M_1, M_2 \in \text{Aut}(V)$ and $\boldsymbol{v}_1, \boldsymbol{v}_2 \in V$, the composition $f_1 \circ f_2 : \mathbb{A} \to \mathbb{A}$ is given by

$$f_1 \circ f_2(b) = a + (M_1 M_2)(b - a) + \boldsymbol{v}_1 + M_1 \boldsymbol{v}_2 \qquad \forall b \in \mathbb{A}.$$

Denoting f_i by a tuple $(M_i, \boldsymbol{v}_i)$, we can express the multiplication law as

$$(M_1, \boldsymbol{v}_1) \circ (M_2, \boldsymbol{v}_2) = (M_1 \circ M_2, \boldsymbol{v}_1 + M_1 \boldsymbol{v}_2).$$

Affine spaces can be equipped with additional structure if additional structures such as bilinear forms are present on the underlying vector spaces. Affine bijections then should be restricted to those affine transformations which preserve this additional structure. If one considers an affine space over the vector space $\mathbb{R}^4$ equipped with the Minkowski metric, one obtains Minkowski space. The affine bijections which preserve the structure of Minkowski space are the Poincaré transformations.

Definition 3.2.10: (Minkowski space)

Minkowski space $\mathbb{M}^4$ is an affine space over the vector space $\mathbb{R}^4$ equipped with the Minkowski metric. The *Poincaré group* is the set of bijective affine transformations which preserve the Minkowski metric. It is the semidirect product $P_3 = SO(3,1) \ltimes \mathbb{R}^4$ with group multiplication

$$(L_1, \boldsymbol{w}_1) \circ (L_2, \boldsymbol{w}_2) = (L_1 \cdot L_2, \boldsymbol{w}_1 + L_1 \boldsymbol{w}_2) \qquad \forall L_1, L_2 \in SO(3,1), \boldsymbol{w}_1, \boldsymbol{w}_2 \in \mathbb{R}^4. \tag{3.17}$$

Straight lines and, more generally, paths $c : \mathbb{R} \to \mathbb{M}^4$ in Minkowski space can now be investigated with respect to their behaviour with respect to the Minkowski metric. In particular, we have the following definitions.

Definition 3.2.11: (Worldline, eigentime, arclength, geodesic)

1. A smooth curve $\boldsymbol{x} : \mathbb{R} \to \mathbb{M}^4$ is called *future directed* (*past directed*) if $\dot{\boldsymbol{x}}(t)$ is future directed (past directed) for all $t \in \mathbb{R}$.

2. It is called timelike (lightlike, spacelike) if for all $t \in \mathbb{R}$ the velocity vector $\dot{\boldsymbol{x}}(t) \in \mathbb{R}^4$ is timelike (lightlike, spacelike).

3. It is called *geodesic* if its image is a straight line in Minkowski space.

4. A future directed, timelike curve $\boldsymbol{x} : \mathbb{R} \to \mathbb{M}^4$ is called a *worldline*. It is called *parametrised according to eigentime* if $\dot{\boldsymbol{x}}(t)^2 = -c^2$ for all $t \in \mathbb{R}$, where $c \in \mathbb{R}^+$ is a fixed constant. The vector
$$\boldsymbol{v}(t) = c \frac{\dot{\boldsymbol{x}}(t)}{\sqrt{|\dot{\boldsymbol{x}}(t)^2|}}$$
is called *four velocity* of the worldline $\boldsymbol{x}$ at time t.

Remark 3.2.12:

Any smooth lightlike curve is a geodesic and either future directed or past directed.

3.3 The theory of special relativity

The axioms of special relativity are usually stated in the following form.

Axiom 1 *Spacetime is homogeneous and isotropic*

Axiom 2 *Relativity: The laws of physics take the same form for all inertial observers. With respect to inertial observers, particles which are not subject to forces move on straight lines and with constant speed.*

Axiom 3 *There exists a universal velocity $c \in \mathbb{R}$ which is the same for all inertial observers and cannot be reached by any massive observer.*

One can show that, together and interpreted correctly, these axioms uniquely characterise Minkowski space. We will not give the full derivation, but show how these axioms can be recast in a mathematically precise form.

We start with the first axiom. The statement that spacetime is homogeneous means that it is described by an affine space $\mathbb{A}^4$ over the vector space $\mathbb{R}^4$. Additionally, to give rise to a notion of lengths, angles and time, one requires that $\mathbb{R}^4$ is equipped with a non-degenerate symmetric bilinear form g.

Spacetime transformations or symmetry transformations must form a subgroup of the group of affine bijections $GL(4,\mathbb{R}) \ltimes \mathbb{R}^4$. As time and space are homogenous, this subgroup must contain all translations and therefore be of the form $G \ltimes \mathbb{R}^4$. The requirement that spacetime is isotropic amounts to imposing that the rotation group $SO(3)$ should be contained in the group of symmetry transformations: $SO(3) \subset G$. Moreover, we require that spacetime transformations are orientation and volume preserving. In other words: the group G should be a subgroup $G \subset SL(4,\mathbb{R})$. We can thus rephrase the first axiom as follows:

Axiom 1' Time and space are described by an affine space $\mathbb{A}^4$ over the vector space $\mathbb{R}^4$ with a non-degenerate, symmetric bilinear form g. Spacetime symmetries which relate different observers form a subgroup $G \ltimes \mathbb{R}^4$ of the group of affine bijections with $SO(3) \subset G \subset SL(4,\mathbb{R})$.

In order to give a precise formulation of the second axiom we must develop a notion of *inertial* observer and of forces. Intuitively, it seems plausible to translate this statement into the requirement that inertial observers and particles which are not subjected to forces should be associated with certain lines in the affine space $\mathbb{A}^4$. The set of all inertial observers should be invariant under spacetime symmetries, and any two inertial observers should be related by a spacetime symmetry. (Otherwise, there would be several non-equivalent types of inertial observers). This yields the following reformulation of Axiom 2:

Axiom 2' 1: Inertial observers and the motion of massive particles not subjected to forces are described by certain lines in $\mathbb{A}^4$. The set of inertial observers forms an orbit under the action of the group of symmetry transformations $S = G \ltimes \mathbb{R}^4$. In other words: a line in $\mathbb{A}^4$ corresponds to an observer if and only if it is of the form $a + \boldsymbol{w}$ in $\mathbb{A}^4$, where $a \in \mathbb{A}^4$, $\boldsymbol{w} = Le_0$, $L \in G$ and e_0 is a fixed vector in $\mathbb{R}^4$.

In order to determine precisely which subgroup $G \subset SL(4,\mathbb{R})$ describes the spacetime symmetries between observers, we need to make more precise statements about the form of these transformations. While the rotation group $SO(3)$ is contained in G as a subgroup, G should contain additional transformations, the *velocity transformations*, which relate observers moving with constant velocities with respect to each other.

We require that these velocity transformations are parametrised uniquely by velocity vectors $\vec{v} \in \mathbb{R}^3$, that $\vec{v} = 0$ corresponds to the identity and that the dependence on $\vec{v}$ is continuous. This amounts to the statement that the group of symmetry transformations is a six-dimensional Lie subgroup of G. Moreover, one requires that the velocity transformations are compatible with rotations which encode the isotropy of spacetime and that velocity transformations in a fixed direction form a subgroup of G. This leads to the following condition.

Axiom 2' 2: The group of symmetry transformations is of the form $G \ltimes \mathbb{R}^4$ where $G \subset SL(4,\mathbb{R})$ is a six-dimensional Lie group which contains $SO(3)$. In addition to rotations, G contains velocity transformations $L(\vec{v}) \in G$ which satisfy

$$L(-\vec{v}) = L(\vec{v})^{-1} \qquad R \cdot L(\vec{v}) \cdot R^{-1} = L(R\vec{v}) \qquad \forall R \in SO(3).$$

Velocity transformations $L(\alpha\boldsymbol{w})$, $\alpha \in \mathbb{R}$ in a fixed direction $\boldsymbol{w} \in \mathbb{R}^4$ form a subgroup of G.

Axiom 3 can be reformulated as follows

Axiom 3' *The set of all admissible velocity vectors* $\vec{v}$ *is the closed three-ball* $B_c(0) \subset \mathbb{R}^3$*. For any* $v, v' \in [-c, c]$ *and any* $\vec{e} \in S^2 \subset \mathbb{R}^3$*, there exists a unique* $v'' \in [-c, c]$ *such that*

$$L(v\vec{e}) \cdot L(v'\vec{e}) = L(v''\vec{e}).$$

It can be shown that together, Axioms 1, 2 and 3 fix the group $G \subset SL(4, \mathbb{R})$ uniquely. They imply $G = SO^{+,\uparrow}(3, 1)$ and that the affine space which describes the spacetime is Minkowski space $\mathbb{M}^4$. We can summarise the resulting description of space and time, i.e. the theory of special relativity in the following postulates.

Postulate 1: (Spacetime and observers)

Space and time are described by Minkowski space $\mathbb{M}^4$. Each inertial observer corresponds to a timelike straight line in Minkowski space. The admissible symmetry transformations which relate different observers form the group of affine transformations $SO(3, 1)^{+,\uparrow} \ltimes \mathbb{R}^4$, i.e. the proper orthochronous Poincaré group.

Postulate 2: (Time and space as perceived by observers)

For an inertial observer whose motion is given by a timelike line $a + \text{Span}(\boldsymbol{w})$, the *velocity unit vector* is the unique future directed, timelike vector $\hat{\boldsymbol{w}}$ satisfying $\hat{\boldsymbol{w}}^2 = -1$ and $\text{Span}(\boldsymbol{w}) = \text{Span}(\hat{\boldsymbol{w}})$.

The time interval elapsed between two events $x, y \in \mathbb{M}^4$ as perceived by this observer is given by $\Delta t_{\hat{\boldsymbol{w}}}(x, y) = |\hat{\boldsymbol{w}}(x - y)|$, where $\hat{\boldsymbol{w}}$ is the velocity unit vector of the observer. The events x, y are perceived as simultaneous if $\hat{\boldsymbol{w}}(x - y) = 0$. Event x occurs before event y if $\hat{\boldsymbol{w}}(x - y) < 0$ and after event y if $\hat{\boldsymbol{w}}(x - y) > 0$. The set of events occurring simultaneously with an event $x \in \mathbb{M}^4$ with respect to the observer given by $\hat{\boldsymbol{w}}$ is the affine plane $x + \hat{\boldsymbol{w}}^\perp$. This gives rise to a foliation of spacetime by affine hyperplanes $b + \hat{\boldsymbol{w}}^\perp$, $b \in a + \text{Span}(\boldsymbol{w})$ through points on the observer's worldline.

The relative position between two simultaneous events $x, y \in a + \hat{\boldsymbol{w}}^\perp$ as perceived by the observer is the spacelike vector $x - y$.

Postulate 3: (Motion of massive particles and of light)

Point particles of mass $m \neq 0$ move on worldlines. The eigentime of the worldline coincides with the time perceived by the particle, i.e. the time shown on a clock moving with the particle. Particles with mass $m \neq 0$ that are not subjected to any forces moves on future directed, timelike geodesics. Light moves along future directed, lightlike geodesics.

In order to formulate the fourth postulate, we need to introduce a notion of future and past.

Definition 3.3.1: (Future, past, lightcone)

The *future* $I^+(x)$ and *past* $I^-(x)$ of a point $x \in \mathbb{M}^4$ are the sets

$$I^\pm(x) = \{y \in \mathbb{M}^4 \mid (y - x)^2 \leq 0\ ,\ y - x \,\text{future (past) directed}\}.$$

They are called the future (past) *lightcone* at $x \in \mathbb{M}^4$.

The future (past) of a subset $S \in \mathbb{M}^4$ is the set

$$I^\pm(S) = \{y \in \mathbb{M}^4 \mid \exists x \in S :\ (x - y)^2 \leq 0\ ,\ y - x \text{ future (past) directed}\} = \bigcup_{x \in S} I^\pm(x).$$

Using this notion of future and past, we can state the fourth postulate as follows:

Postulate 4: (Causality)

Each physical event corresponds to a unique point in Minkowski space $\mathbb{M}^4$. An event $\boldsymbol{x} \in \mathbb{M}^4$ can influence an event $\boldsymbol{y} \in \mathbb{M}^4$ or a signal can be sent from $\boldsymbol{x}$ to $\boldsymbol{y}$ if and only if $\boldsymbol{y}$ lies in the future of $\boldsymbol{x}$. If $\boldsymbol{x} \notin I^+(\boldsymbol{y})$ and $\boldsymbol{y} \notin I^+(\boldsymbol{x})$, $\boldsymbol{x}$ and $\boldsymbol{y}$ are called *causally disconnected* and no signal can be passed between them.

We now have to show that Postulate 2 gives rise to a sensible notion of time for each inertial observer, which is compatible with the notion of past, future and causality introduced in Postulate 4. This is guaranteed by the following lemma.

Lemma 3.3.2: For any inertial observer whose worldline is characterised by a velocity unit vector $\hat{\boldsymbol{w}}$ and any events $x, y, z \in \mathbb{M}^4$, we have:

1. x and y simultaneous implies $y - x$ spacelike
2. $x \in I^+(y)$ implies that y happens before x with respect to the observer, $x \in I^-(y)$ implies that x happens before y.
3. If x happens before y and y happens before z, then x happens before z.

Proof: Exercise.

This lemma shows that for a single observer, the assignment of time makes sense and is compatible with the notion of past, future and causality. However, it is clear that it depends on the observer. Time and space are no longer absolute as in Newtonian mechanics but become observer-dependent.

In order to ensure that the notion of time is physically sensible, we now need to make sure that assignments of time made by different observers are compatible and that the second axiom holds. This is guaranteed by the content of the following lemma.

Lemma 3.3.3: (Relativity of simultaneity)

For any two inertial observers whose velocity unit vectors $\hat{\boldsymbol{w}}$, $\hat{\boldsymbol{v}}$, $\hat{\boldsymbol{w}} \neq \hat{\boldsymbol{v}}$ are not equal, there exist events x, y such that x happens before y with respect to the observer characterised by $\hat{\boldsymbol{w}}$ and y happens before x with respect to the observer characterised by $\hat{\boldsymbol{v}}$. Such events are always causally disconnected.

Proof: Exercise. □

Remark 3.3.4: Note that the existence of events whose order in time is reversed with respect to two observers is not in contradiction with Axiom 2. As such events are always causally disconnected, neither of them can affect the other. Although two observers could disagree about which event happens before the other, they would agree on the fact that neither of the two events can influence the other or, equivalently, that there is no light ray emitted at event x and received at event y or vice versa. The fact that two observers assign different times to events therefore does not mean that they perceive different laws of physics.

Exercise 29: (Relativity of simultaneity)

1. Consider two inertial observers with velocity unit vectors $\hat{\boldsymbol{w}}$, $\hat{\boldsymbol{v}}$, $\hat{\boldsymbol{w}} \neq \hat{\boldsymbol{v}}$. Show that there exist events $x, y \in \mathbb{M}^4$ such that $\hat{\boldsymbol{w}}(x-y) < 0$, $\hat{\boldsymbol{v}}(x-y) > 0$, i.e. events $x, y \in \mathbb{M}^4$ such that x happens before y with respect to the observer characterised by $\hat{\boldsymbol{w}}$ and x happens after y for the observer characterised by $\hat{\boldsymbol{v}}$. Hint: it is sufficient to show that any two affine, spacelike planes which are not parallel intersect.

2. Show that such events are always causally disconnected, i.e. that $\hat{\boldsymbol{w}}(x-y) < 0$, $\hat{\boldsymbol{v}}(x-y) > 0$ implies $x - y$ spacelike.

3. Show that the future of an event $x \in \mathbb{M}^4$ is the set of points

$$I^+(x) = \{y \in \mathbb{M}^4 \mid \hat{\boldsymbol{v}}(y-x) < 0 \; \forall \hat{\boldsymbol{v}} \in \mathbb{R}^4, \hat{\boldsymbol{v}}^2 = -1\}, \tag{3.18}$$

i.e. the set of events that happen after x for *all inertial observers*.

This lemma guarantees that the time measured by different observers is compatible with the notion of causality and is sensible for all observers. However, the observer-dependence of time and space has consequences for the measurement of time intervals and lengths. The first consequence of the relativity of simultaneity is *time dilatation* - moving clocks slow down with respect to resting ones.

Example 3.3.5: (Time dilatation)

Consider two inertial observers with velocity unit vectors $\hat{\boldsymbol{w}}$, $\hat{\boldsymbol{v}}$ and two events $x, y \in \mathbb{M}^4$ which occur in the momentum rest frame of the observer characterised by $\hat{\boldsymbol{w}}$, i.e. $x - y = \hat{\boldsymbol{w}}$. This could for instance be the ticking of a clock travelling with the observer. Then the time interval between these events measured by this observer is

$$\Delta t_{\hat{\boldsymbol{w}}}(x,y) = |x-y| = \sqrt{-(x-y)^2}.$$

The time elapsed between these events with respect to the observer characterised by $\hat{\boldsymbol{v}}$ is given by $\Delta t_{\hat{\boldsymbol{v}}}(x,y) = |\hat{\boldsymbol{v}}(x-y)|$. We have

$$x - y = -\Delta t_{\hat{\boldsymbol{v}}}(x,y)\hat{\boldsymbol{v}} + \boldsymbol{a},$$

where $\boldsymbol{a} = (x-y) + ((x-y) \cdot \hat{\boldsymbol{v}})\boldsymbol{v}$ is a spacelike vector orthogonal to $\hat{\boldsymbol{v}}$: $\hat{\boldsymbol{v}} \cdot \boldsymbol{a} = 0$ as shown in figure **??**. This implies

$$\Delta t_{\hat{\boldsymbol{w}}}(x,y)^2 = |x-y|^2 = \Delta t_{\hat{\boldsymbol{v}}}(x,y)^2 - \boldsymbol{a}^2 < \Delta t_{\hat{\boldsymbol{v}}}(x,y)^2$$

Hence, we have $\Delta t_{\hat{\boldsymbol{v}}}(x,y) > \Delta t_{\hat{\boldsymbol{w}}}(x,y)$. The observer who moves with respect to the clock therefore measures a bigger time interval. Moved clocks are slowed.

To obtain an explicit formula, we note that the length of the vector $\boldsymbol{a}$ is given by $\boldsymbol{a}^2 = (|\hat{\boldsymbol{v}}\hat{\boldsymbol{w}}|^2 - 1)\Delta t^2_{\hat{\boldsymbol{w}}}(x,y)$. This implies

$$\Delta t_{\hat{\boldsymbol{v}}}(x,y) = \Delta t_{\hat{\boldsymbol{w}}}(x,y)|\hat{\boldsymbol{v}}\hat{\boldsymbol{w}}|$$

While the observer associated with $\hat{\boldsymbol{w}}$ measures a time interval $\Delta t_{\hat{\boldsymbol{w}}}$ between the two events x, y, the observer associated with $\hat{\boldsymbol{v}}$ measures a time interval $\Delta t_{\hat{\boldsymbol{v}}} = \Delta t_{\hat{\boldsymbol{w}}} \cdot |\hat{\boldsymbol{w}}\hat{\boldsymbol{v}}|$. As $\hat{\boldsymbol{w}}, \hat{\boldsymbol{v}} \in \mathbb{R}^4$ are both timelike and future directed unit vectors, Exercise 25 implies $\hat{\boldsymbol{w}}\hat{\boldsymbol{v}} < 0$, $|\hat{\boldsymbol{w}}\hat{\boldsymbol{v}}| \geq 1$ and $|\hat{\boldsymbol{w}}\hat{\boldsymbol{v}}| > 1$ if $\hat{\boldsymbol{w}} \neq \hat{\boldsymbol{v}}$. Hence, we have $\Delta t_{\hat{\boldsymbol{v}}}(x,y) > \Delta t_{\hat{\boldsymbol{w}}}(x,y)$.

The second consequence of the relativity of simultaneity is the Lorentz contraction - moving rods appear shortened.

Lemma 3.3.6: (Lorentz contraction)

Consider two inertial observers with velocity unit vectors $\hat{\boldsymbol{w}}$, $\hat{\boldsymbol{v}}$ and a rod which is at rest with respect to the observer $\hat{\boldsymbol{w}}$. The worldlines of its ends are therefore of the form

$$e_1(t) = t\hat{\boldsymbol{w}} + z_1 \qquad e_2(t) = t\hat{\boldsymbol{w}} + z_2 \qquad z_1, z_2 \in \mathbb{M}^4, \hat{\boldsymbol{w}} \cdot (z_1 - z_2) = 0 \tag{3.19}$$

The relative position $\vec{p}$ of the ends of the rod as measured by observer $\hat{\boldsymbol{w}}$ is therefore given by $\vec{p} = z_1 - z_2$ and its length is $(z_1 - z_2)^2$.

When the observer $\hat{\boldsymbol{v}}$ measures the relative position of the ends of the rod, the result is the vector $\boldsymbol{u}$ between intersection of the worldlines associated with the ends of the rod with an affine plane $a + \hat{\boldsymbol{v}}^\perp$ of events simultaneous with respect to $\hat{\boldsymbol{v}}$

$$\boldsymbol{u} = z_2 - z_1 - \frac{(z_2 - z_1) \cdot \hat{\boldsymbol{v}}}{\hat{\boldsymbol{v}}\hat{\boldsymbol{w}}}\hat{\boldsymbol{w}}$$

If we decompose the vector $z_1 - z_2$ into a component $\boldsymbol{a}$ orthogonal to both $\hat{\boldsymbol{v}}$ and $\hat{\boldsymbol{w}}$ and into a component $\boldsymbol{c} = \alpha(\hat{\boldsymbol{v}} + (\hat{\boldsymbol{v}}\hat{\boldsymbol{w}})\hat{\boldsymbol{w}})$ in the plane spanned by $\hat{\boldsymbol{v}}, \hat{\boldsymbol{w}}$ and orthogonal to $\hat{\boldsymbol{w}}$ but not to $\hat{\boldsymbol{v}}$, we have

$$\boldsymbol{u} = \boldsymbol{a} + \boldsymbol{c} - \frac{\hat{\boldsymbol{v}}\boldsymbol{c}}{\hat{\boldsymbol{v}}\hat{\boldsymbol{w}}}\hat{\boldsymbol{w}}.$$

The length of $\boldsymbol{u}$ is given by

$$\boldsymbol{u}^2 = \boldsymbol{a}^2 + \boldsymbol{c}^2 - \frac{(\hat{\boldsymbol{v}}\boldsymbol{c})^2}{(\hat{\boldsymbol{v}}\hat{\boldsymbol{w}})^2} = \boldsymbol{a}^2 + \frac{\boldsymbol{c}^2}{(\hat{\boldsymbol{v}}\hat{\boldsymbol{w}})^2}.$$

The observer $\hat{\boldsymbol{v}}$ therefore assigns to the rod the length $l(\hat{\boldsymbol{v}})$ with $l(\hat{\boldsymbol{v}})^2 = \boldsymbol{u}^2 = \boldsymbol{a}^2 + \boldsymbol{c}^2/(\hat{\boldsymbol{v}}\hat{\boldsymbol{w}})^2$, while the observer associated with $\hat{\boldsymbol{w}}$ measures its length as $l(\hat{\boldsymbol{w}}) = \boldsymbol{a}^2 + \boldsymbol{c}^2$.

From Exercise 25, we have $|\hat{\boldsymbol{w}}\hat{\boldsymbol{v}}| > 1$ if $\hat{\boldsymbol{v}} \neq \hat{\boldsymbol{w}}$, which implies $l(\hat{\boldsymbol{v}}) \leq l(\hat{\boldsymbol{w}})$ and $l(\hat{\boldsymbol{v}}) < l(\hat{\boldsymbol{w}})$ if the vector $\boldsymbol{c} \neq 0$. The rod therefore appears shortened for an observer who moves with respect to the rod and whose velocity relative to the rod is not orthogonal to it. The shortening effect is maximal if the observer moves parallel to the rod ($\boldsymbol{a} = 0$).

Exercise 30:

1. Consider an observer with worldline

$$e(t) = t\boldsymbol{x} + \boldsymbol{x}_0 \qquad \text{with } \boldsymbol{x}_0 \in \mathbb{M}^4, \boldsymbol{x} \in \mathbb{R}^4, \boldsymbol{x}^2 = -c^2, e_0 \cdot \boldsymbol{x} < 0$$

and a future directed geodesic $\boldsymbol{y} : \mathbb{R} \to \mathbb{M}^4$. Use the time intervals and position vectors with respect to observers to show that the relative velocity of $\boldsymbol{y}$ with respect to the observer with worldline e is given by

$$\vec{v}(t) = -c\,\frac{\dot{\boldsymbol{y}}(t) + (\hat{\boldsymbol{x}} \cdot \dot{\boldsymbol{y}}(t))\hat{\boldsymbol{x}}}{\hat{\boldsymbol{x}} \cdot \dot{\boldsymbol{y}}(t)}.$$

Show that $\vec{v}(t)^2 < c^2$ if $\boldsymbol{y}$ is timelike and $\vec{v}(t)^2 = c^2$ if $\boldsymbol{y}$ is lightlike. Show that the velocity four vector associated with $\boldsymbol{y}$ can be written as

$$\dot{\boldsymbol{y}}(t) = \frac{1}{\sqrt{1 - \vec{v}^2(t)/c^2}}(c \cdot e_0 + \vec{v}(t)),$$

where $\vec{v}(t)$ is the velocity of $\boldsymbol{y}$ with respect to the observer with worldline $e(t) = te_0 + \boldsymbol{x}_0$.

2. Show that for an inertial observer with four velocity $\boldsymbol{x} = \frac{1}{\sqrt{1-\vec{v}^2(t)/c^2}}(ce_0 + \vec{v})$, $\vec{v}e_0 = 0$ and an internal observer with four velocity $\boldsymbol{y} = ce_0$, the formulas for time dilatation and Lorentz contraction take the form

$$\Delta t' = \frac{\Delta t}{\sqrt{1-\vec{v}^2/c^2}} \qquad l' = l\sqrt{1-\vec{v}^2/c^2}.$$

Exercise 31: (Addition of velocities)

Consider two boosts in the e_1 direction.

$$B(\theta_1) = \begin{pmatrix} \cosh\theta_1 & \sinh\theta_1 & 0 & 0 \\ \sinh\theta_1 & \cosh\theta_1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad B(\theta_2) = \begin{pmatrix} \cosh\theta_2 & \sinh\theta_2 & 0 & 0 \\ \sinh\theta_2 & \cosh\theta_2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Show that the product $B_1(\theta_1) \cdot B(\theta_2) = B(\theta_2) \cdot B(\theta_1) = B(\theta_3)$ is another boost in the e_1-direction with

$$\theta_3 = \theta_1 + \theta_2.$$

Show that the velocity of the observer with four velocity $c \cdot B(\theta)e_0$ with respect to the observer with four velocity $c \cdot e_0$ is given by $\vec{v} = v \cdot e_1$, $v = c \cdot \tanh\theta$. Show that the relativistic velocity addition law takes the form

$$v_3 = \frac{v_1 + v_2}{1 + v_1 v_2 / c}.$$

Example 3.3.7: (Appearance of the night sky)

We consider an inertial observer at a point $x \in \mathbb{M}^4$ and his past lightcone

$$I^-(x) = \{x + \boldsymbol{y} \in \mathbb{M}^4 \mid \boldsymbol{y}^2 = 0, e_0 \cdot \boldsymbol{y} > 0\},$$

which is the set of all lightrays sent out in Minkowski space which reach the observer at the event x. The map

$$f : I^-(x) \to S^2, \quad \boldsymbol{y} = (|\vec{y}|, \vec{y}) \mapsto f(\boldsymbol{y}) = (1, \vec{y}/|\vec{y}|)$$

maps the backward lightcone to the two-sphere S^2 and is invariant under $\boldsymbol{y} \to \alpha\boldsymbol{y}$, $\alpha \in \mathbb{R}^+$. It therefore induces a map from the set of past lightrays at $\boldsymbol{x}$ to S^2. Each point $p \in S^2$ corresponds to a point in the sky as seen by the observer.

By combining this map with the stereographic projection $P : S^2 \to \mathbb{C} \cup \{\infty\}$

$$P(\vec{k}) = \begin{cases} \frac{k^1 - ik^2}{1-k^3} & k^3 \neq 1 \\ \infty & k^3 = 1 \end{cases} \qquad \forall \vec{k} \in S^2$$

we obtain a map $\rho = P \circ f : \mathbb{R}^4 \to \mathbb{C} \cup \{\infty\}$

$$\rho(\boldsymbol{y}) = \begin{cases} \frac{y^1 - iy^2}{y^0 - y^3} & y^0 \neq y^3 \\ \infty & y^0 = y^3 \end{cases}$$

which is again invariant under rescalings $\boldsymbol{y} \mapsto \alpha\boldsymbol{y}$, $\alpha \in \mathbb{R}_0^+$ and therefore induces a map from the set of past lightrays to $\mathbb{C} \cup \infty$. Note that by definition, the stereographic projection maps circles on S^2 to circles or lines in $\mathbb{C}\cup\{\infty\}$. More precisely, any circle in S^2 which does not contain the point $(0,0,1)$ is mapped to a circle in $\mathbb{C}\cup\{\infty\}$ and any circle through $(0,0,1)$ to a line.

We will now use this map to see how the appearance of the night sky changes if a boost or rotation is applied to the observer. For this, it is convenient to use the identification of $\mathbb{R}^4$ with the set of hermitian matrices as in Exercise 28

$$\boldsymbol{x}\in\mathbb{R}^4\mapsto\phi(\boldsymbol{x})=x_0+\sum_{i=1}^3 x^i\sigma_i=\begin{pmatrix} x^0+x^3 & x^1-ix^2\\ x^1+ix^2 & x^0-x^3\end{pmatrix}, \tag{3.20}$$

in which the Minkowski metric is given by $\boldsymbol{x}\cdot\boldsymbol{y}=-\det(\phi(\boldsymbol{x})\cdot\phi(\boldsymbol{y}))$. With this identification, lightlike past oriented vectors are given by hermitian matrices $M\in M(2,\mathbb{C})$, $M^\dagger=M$ with vanishing determinant.

Applying this map to the past lightcone, we obtain

$$\phi(\boldsymbol{y})=\begin{pmatrix} y^0+y^3 & y^1-iy^2\\ y^1+iy^2 & y^0-y^3\end{pmatrix}=(y^0-y^3)\cdot\begin{pmatrix}|z|^2 & z\\ \bar z & 1\end{pmatrix}\quad\text{with}\quad z=\rho(\boldsymbol{y}). \tag{3.21}$$

As shown in Exercise 28, the proper orthochronous Lorentz group is isomorphic to $PSL(2,\mathbb{C})=SL(2,\mathbb{C})/\{\pm 1\}$, and the action of a Lorentz transformation L on $\boldsymbol{v}\in\mathbb{R}^4$ agrees with the action of the associated element $\tilde L\in SL(2,\mathbb{C})$ on the hermitian matrix $\phi(\boldsymbol{v})$

$$\tilde L\cdot\phi(\boldsymbol{v})\cdot\tilde L^\dagger=\phi(L\boldsymbol{v})\qquad\forall\boldsymbol{v}\in\mathbb{R}^4, L\in SO^{+,\uparrow}(3,1).$$

We now consider the transformation of the hermitian matrix (3.21) associated with a backward lightray under a Lorentz transformation given by a matrix $\tilde L\in SL(2,\mathbb{C})$

$$\tilde L=\begin{pmatrix} a & b\\ c & d\end{pmatrix}\in SL(2,\mathbb{C}).$$

This yields

$$L\cdot\phi(\boldsymbol{y})\cdot L^\dagger=(y^0-y^3)|cz+d|^2\cdot\begin{pmatrix}\left|\frac{az+b}{cz+d}\right|^2 & \frac{az+b}{cz+d}\\ \frac{\bar a\bar z+\bar b}{\bar c\bar z+\bar d} & 1\end{pmatrix}.$$

A Lorentz transformation acting on the past lightcone therefore corresponds to a *Möbius transformation* acting on $\mathbb{C}\cup\{\infty\}$

$$\tilde L=\begin{pmatrix} a & b\\ c & d\end{pmatrix}\in SL(2,\mathbb{C})\qquad z\mapsto\frac{az+b}{cz+d}.$$

The defining characteristic of Möbius transformations is that they map circles and lines in $\mathbb{C}\cup\{\infty\}$ to circles and lines in $\mathbb{C}\cup\{\infty\}$. Due to the properties of the stereographic projection, circles and lines in $\mathbb{C}\cup\{\infty\}$ correspond to circles on the night sky as perceived by the observer. Lorentz transformations acting on the backward lightcone therefor map circles on the night sky to circles in the night sky. This implies in particular that the Lorentz contraction is *invisible.*

Chapter 4

Topological Vector Spaces

4.1 Types of topological vector spaces

Infinite dimensional vector spaces play an important role in quantum mechanics . This follows already from the fact that the Heisenberg algebra, which encodes two of the most important observables in physics, namely the positions and momenta, has no non-trivial finite-dimensional representations. One can show this by taking the *trace*, which leads to a contradiction. In the infinite-dimensional case, this contradiction disappears, because there are operators of which one cannot take the trace, i.e. whose trace is ill-defined.

Quantum mechanics therefore forces one to consider representations on infinite-dimensional vector spaces. However, it turns out that the structure of a vector space alone is too weak. To have the desired properties, infinite dimensional vector spaces need to be equipped with additional structure, namely with a topology that is compatible with the structure of a vector space. This leads to the notion of topological vector spaces.

Infinite dimensional vector spaces arising in physics are topological vector spaces, often with further structures such as norms, semi-norms or hermitian products. In this chapter, we will study different examples of such vector spaces which play an important role in physics and and investigate their properties.

Throughout this chapter, we will make extensive use of the Lebesgue measure. All integrals considered in the following are Lebesgue integrals. We will assume familiarity with Lebesgue integration and elementary measure theory. A concise introduction is given in J. Dieudonné, Treatise on Analysis 1.

4.1.1 Topological vector spaces, metric spaces and normed spaces

We start by introducing the notion of *topological spaces* and topological vector spaces.

Definition 4.1.1: (Topology, continuity)

A *topology* on a set E is a family T of subsets $F \subset E$, called *open sets in* E such that

1. $\emptyset, X \in T$
2. *General unions* of open sets are open sets: If $F_i \in T\ \forall i \in I \Rightarrow \bigcup_{i\in I} F_i \in T$

3. *Finite intersections* of open sets are open sets: If $F_i \in T$, $\forall i \in \{1, ..., n\} \Rightarrow \bigcap_{i=1}^n F_i \in T$.

A set E with a topology is called *a topological space.*

A function $f : E \to G$ between topological spaces E and G is called *continuous* if for every open set $F \subset G$ the preimage $f^{-1}(F) = \{x \in E \mid f(x) \in F\}$ is open.

Example 4.1.2: If X, Y are topological spaces with topologies T_X, T_Y, then the direct product $X \times Y$ is a topological space with topology $T_{X\times Y} = \{U \times V \mid U \in T_X, V \in T_Y\}$.

A topological vector space is a vector space whose topology is compatible with the vector space structure:

Definition 4.1.3: (Topological vector space, isomorphisms of topological vector spaces)

A *topological vector space* E over $k = \mathbb{R}$ or $k = \mathbb{C}$ is a vector space E over k with a topology T such that the vector space addition $+ : E\times E \to E$ and the scalar multiplication $\cdot : k\times E \to E$ are *continuous* with respect to T. In other words: the preimage of each open set in E is open with respect to the induced topologies on $E \times E$ and $k \times E$ (where k is equipped with the standard topology).

When considering isomorphisms of topological vector spaces and their duals, one imposes that these structures are compatible with the topology on E. In other words: they are required to be continuous.

Definition 4.1.4: (Isomorphisms of topological vector spaces, continuous duals)

An *isomorphism* of topological vector spaces E, G is a bijective, linear map $L : E \to G$ that is a homeomorphism, i.e. continuous with respect to the topologies on E, G and invertible such that L^{-1} continuous with respect to the topologies on E, G.

The *continuous dual* E^* of a topological vector space E over k is the set of continuous linear functionals $E \to k$, i.e. linear maps $E \to k$ which are continuous with respect to the topology on E.

An important concept which allows one to classify topologies and hence topological vector spaces is the notion of *Hausdorff.* It also plays a crucial role in the notion of manifolds.

Definition 4.1.5: (Neighbourhood, Hausdorff)

A topological vector space is called *Hausdorff* if for any two elements $x, y \in E$ there exist open sets U_x, U_y with $x \in U_x$, $y \in U_y$ such that $U_x \cap U_y = \emptyset$.

A subset $V \subset E$ is called a *neighbourhood* of $x \in E$ if there exists an open set $U \subset V$ with $x \in U$.

It turns out that in the finite-dimensional case the requirement that the topology is Hausdorff is strong enough to determine the topology of a topological vector space completely. Note, however, that this is *not* true for infinite dimensional topological vector spaces.

Lemma 4.1.6: On any finite-dimensional vector space V there is a *unique* topology, the *canonical topology*, which is Hausdorff and turns V into a topological vector space.

If V is a finite-dimensional topological vector space over k and $B = \{e_1, ..., e_n\}$ a basis of V, then the map $\phi : k^n \to V$

$$(x^1, ..., x^n) \mapsto \sum_{i=1}^{n} x^i e_i$$

is an isomorphism of topological vector spaces.

An important class of topologies are those defined by *distance functions.* This leads to the concept of *metric spaces.*

Definition 4.1.7: (Metric space, distance function, metrisable space)

A *metric space* is a set E together with a *distance function* $d : E \times E \to \mathbb{R}$ that has the following properties:

1. *Positivity*: $d(x, y) \geq 0$ for all $x, y \in E$ and $d(x, y) = 0$ implies $x = y$.
2. *Symmetry*: $d(x, y) = d(y, x)$ for all $x, y \in E$
3. *Triangle inequality*: $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in E$

Any metric space is a topological space. The topology T induced by the distance function is the family of sets generated by r-balls $B_r(x) = \{y \in E \mid d(y, x) < r\}$, $r \in \mathbb{R}$, i.e. the family of sets which can be expressed as general union or finite intersection of r-balls.

A topological vector space is called *metrisable* if its topology agrees with a topology induced by a distance function on E.

Remark 4.1.8: Metric spaces and metrizable spaces are *Hausdorff.*

Example 4.1.9: Consider k^n ($k = \mathbb{R}$ or $k = \mathbb{C}$) with the standard basis $B = \{e_1, ..., e_n\}$. A distance function on k^n is given by

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} |x_i - y_i| \qquad \forall \boldsymbol{x} = \sum_{i=1}^{n} x^i e_i, \boldsymbol{y} = \sum_{i=1}^{n} y^i e_i \in k^n.$$

A special class of metric spaces are metric spaces whose distance function is induced by a norm

Definition 4.1.10: (Norm, semi-norm)

A *semi-norm* on a vector space E over k is a map $||\ \ || : E \to \mathbb{R}$ which

- is *positive semi-definite* $||x|| \geq 0$ for all $x \in E$
- satisfies $||\lambda x|| = |\lambda| ||x||$ for all $x \in E$, $\lambda \in k$.
- satisfies the *triangle inequality*: $||x + y|| \leq ||x|| + ||y||$ for al $x, y \in E$.

A *norm* on E is a semi-norm which is *positive definite* $||x|| > 0$ for all $x \neq 0$.

Lemma 4.1.11: (Normed vector spaces are metric spaces)

A norm $|| \; || : E \to \mathbb{R}$ on a vector space E over $k = \mathbb{R}$ or $k = \mathbb{C}$ defines a distance function $d : E \times E \to \mathbb{R}$ on E via

$$d(x, y) = ||x - y|| \qquad \forall x, y \in E.$$

The distance function is *translation invariant*, i.e. satisfies $d(x + z, y + z) = d(x, y)$ for all $x, y, z \in E$ and $d(\lambda x, \lambda y) = |\lambda| d(x, y)$ for all $x, y \in E$, $\lambda \in \mathbb{R}$.

Example 4.1.12: Consider k^n ($k = \mathbb{R}$ or $k = \mathbb{C}$) with the standard basis $B = \{e_1, ..., e_n\}$. A norm on k^n is given by

$$||\boldsymbol{x}|| = \sum_{i=1}^{n} |x_i| \qquad \forall \boldsymbol{x} = \sum_{i=1}^{n} x^i e_i, x^i \in k.$$

The distance function induced by this norm is the one from example 4.2.1.

Example 4.1.13: Let $f : V \to k$ be a linear form on a vector space V over k. Then

$$||\boldsymbol{x}|| = |f(\boldsymbol{x})| \qquad \forall \boldsymbol{x} \in V$$

defines a norm on V.

The fact that normed vector spaces are metric spaces implies in particular that they are equipped with a topology induced by the distance function. It turns out that this topology is always compatible with the vector space structure and turns the vector space E into a topological space.

Lemma 4.1.14: For any normed vector space E over $k = \mathbb{R}$ or $k = \mathbb{C}$, the vector space addition $+ : E \times E \to E$, $(\boldsymbol{x}, \boldsymbol{y}) \mapsto \boldsymbol{x} + \boldsymbol{y}$ and the scalar multiplication $\cdot_\lambda : E \to E$, $\boldsymbol{x} \mapsto \lambda \boldsymbol{x}$ with $\lambda \in k$ are uniformly continuous.

To classify norms on vector spaces, one needs a concept of equivalence of norms. It is natural to call two norms on a vector space *equivalent* if they induce the same topology. One can show that this is the case if and only if sequences that converge with respect to one norm also converge with respect to the other norm. This leads to the following definition.

Definition 4.1.15: (Equivalence of norms)

Two norms $|| \; ||_1, || \; ||_2 : E \to \mathbb{R}$ on a vector space E are *equivalent* if there exist $a, b > 0$ such that

$$a||x||_1 \leq ||x||_2 \leq b||x||_1 \qquad \forall x \in E.$$

Remark 4.1.16: One can show that on a *finite-dimensional* vector space, *all* norms are equivalent. This is not the case for infinite-dimensional topological vector spaces. A counter example is given in Example 4.1.29.

Exercise 32: Consider the vector space k^n, where $k = \mathbb{R}$ or $k = \mathbb{C}$, with standard basis $B = \{e_1, ..., e_n\}$ and express vectors $\boldsymbol{x} \in k^n$ with respect to this basis as $\boldsymbol{x} = \sum_{i=1}^n x^i e_i$. Show that the following norms are equivalent:

1. $||\boldsymbol{x}|| = \sqrt{\sum_{i=1}^n |x^i|^2}$
2. $||\boldsymbol{x}|| = \sum_{i=1}^n |x^i|$
3. $||\boldsymbol{x}|| = \max\{|x^i| \, | i \in \{1, ..., n\}\}$.

4.1.2 Topological vector spaces with semi-norms

The existence of a semi-norm is not enough to turn a vector space into a metric space. However, it is possible to construct normed vector spaces and hence metric spaces from topological vector spaces with semi-norms by taking a quotient. This is similar to of a situation in Example **??**, in which quotients of vector spaces are used to turn linear maps between vector spaces V and W into *injective* linear maps $V/\ker(\phi) \to W$. Note, however, that a norm is *not* a linear map, although the basic idea of the construction is similar.

Lemma 4.1.17: (Quotients of vector spaces with semi-norms)

If E is a topological vector space with a semi-norm $||\ ||$, the set $N = \{x \in E \mid ||x|| = 0\}$ is a linear subspace of E. The quotient space E/N is a normed vector space with norm $||[x]|| = ||x + N|| := ||x||$.

Proof: Exercise. □

Exercise 33: Consider a topological vector space E over $k = \mathbb{R}$ or $k = \mathbb{C}$ with a semi-norm $||\ \ ||$. Show that the set $N = \{x \in E \mid ||x|| = 0\}$ is a linear subspace of E. Consider the quotient space E/N. Show that the definition

$$||\,[x]\,|| = ||x|| \qquad \forall x \in E$$

where $[x] = x + N = \{y \in E \mid ||y - x|| = 0\}$ are the equivalence classes of elements $x \in E$ defines a norm on E/N.

Hint: Show first that this definition is consistent, i.e. that for $x, y \in E$ with $[x] = [y]$, we have $||\,[x]\,|| = ||\,[y]\,||$. Then you know that this defines a map $||\ \ || : E/N \to k$ and you can show that this map has the properties of a norm.

Another important situation in which semi-norms give rise to topological vector spaces are vector spaces which carry a *family of semi-norms*. These vector spaces play an important role in the theory of distributions.

Definition 4.1.18: (Topology defined by semi-norms, locally convex)

Let $(||\ \ ||_\alpha)_{\alpha \in I}$, $||\ \ ||_\alpha : E \to \mathbb{R}$ be a family of semi-norms on a vector space E. The family of semi-norms induces a topology on E. In this topology, the class T of open sets is the class of subsets $U \subset E$ such that for all $x \in U$ there exists a *finite* family of semi-norms $(||\ \ ||_{\alpha_i})_{i=1,...,n}$ and positive numbers $(r_j)_{j=1,...,n}$ such that

$$K(x; (\alpha_j), (r_j)) := \{y \in E \mid ||y - x||_{\alpha_j} < r_j \ \forall j \in \{1, ..., n\}\} \subset U.$$

This topology is compatible with the vector space structure and turns E into a topological vector space.

A vector space E whose topology is given by a family of semi-norms is called a *locally convex vector space*. A sequence $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$, $\boldsymbol{x}_n \in E$ in a locally convex vector space E is said to *converge towards an element* $\boldsymbol{x} \in E$ with respect to a family of semi-norms if $||\boldsymbol{x}_n - x||_\alpha \to 0$ for all $\alpha \in I$.

We will now show that the space of smooth functions with compact support (also often called *test functions* in the physics literature) is an example of a locally convex vector space.

Definition 4.1.19: (Smooth functions with compact support)

A *smooth function with compact support* on $\mathbb{R}^n$ is a smooth function $\varphi : \mathbb{R}^n \to \mathbb{C}$ whose support $\text{Supp}(\varphi) = \overline{\{\boldsymbol{x} \in \mathbb{R}^n : \varphi(\boldsymbol{x}) \neq 0\}}$ is compact. The set of functions with compact support is denoted $\mathcal{C}_c^\infty(\mathbb{R}^n)$. With pointwise addition and multiplication by $\mathbb{C}$ it becomes an infinite dimensional vector space.

A family of semi-norms on the space of smooth functions with compact support is obtained by considering the supremum of general derivatives. The motivation for defining these semi-norms is the wish to *control derivatives* of functions.

Lemma 4.1.20: (Semi-norms on $\mathcal{C}_c^\infty(\mathbb{R}^n)$)

We consider general multi-indices $\alpha = (j_1, ..., j_n) \in \mathbb{N}^n$, $n \in \mathbb{N}$ and set $D^\alpha f = \partial_{j_1} \cdots \partial_{j_n} f$. Then, a family of semi-norms $(|| \ ||_\alpha)_{\alpha \in \mathbb{N}^\mathbb{N}}$ is given by

$$||f||_\alpha = \sup_{\boldsymbol{x} \in \mathbb{R}^n} \{|D^\alpha f(\boldsymbol{x})|\} \qquad f \in \mathcal{C}_c^\infty(\mathbb{R}^n), \alpha \in \mathbb{N}^\mathbb{N}.$$

The space $\mathcal{C}_c^\infty(\mathbb{R}^n)$ with the topology induced by this family of semi-norms is a locally convex vector space.

Proof: Exercise. □

Remark 4.1.21: The locally convex vector space $\mathcal{C}_c^\infty(\mathbb{R}^n)$ is *not metrizable.*

We now define a notion of convergence on the vector space $\mathcal{C}_c^\infty(\mathbb{R}^n)$ equipped with this family of semi-norms.

Definition 4.1.22: (Convergence in $\mathcal{C}_c^\infty(\mathbb{R}^n)$)

A sequence $\{\varphi_i\}_{i \in \mathbb{N}}$, $\varphi_i \in \mathcal{C}_c^\infty(\mathbb{R}^n)$ converges towards a function $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}^n)$, $\varphi_i \xrightarrow{\mathcal{C}_c^\infty(\mathbb{R}^n)} \varphi$, with respect to this family of semi-norms if

1. There exists a compact subset $K \subset \mathbb{R}^n$ such that $\text{Supp}(\varphi) \subset K$, $\text{Supp}(\varphi_i) \subset K \ \forall i \in \mathbb{N}$
2. For all $\alpha = (j_1, ..., j_n) \in \mathbb{N}^n$, the sequence of derivatives $D^\alpha \varphi_i = \partial_{j_1} \cdots \partial_{j_n} \varphi_i$ converges uniformly to $D^\alpha \varphi = \partial_{j_1} \cdots \partial_{j_n} \varphi$. In other words: for any multi-index $\alpha \in \mathbb{N}^n$ and any $\epsilon > 0$ there exists an $i_\epsilon^\alpha \in \mathbb{N}$ such that for all $\boldsymbol{x} \in \mathbb{R}^n$, $i > i_\epsilon$ $|D^\alpha \varphi_i(\boldsymbol{x}) - D^\alpha \varphi(\boldsymbol{x})| < \epsilon$.

One can show that the requirement that a function $f : \mathcal{C}_c^\infty(\mathbb{R}^n) \to \mathbb{C}$ is continuous in the sense of Definition 4.1.3 if and only if $f(\varphi_n) \xrightarrow{n \to \infty} f(\varphi)$ for all sequences $\{\varphi_n\}_{n \in \mathbb{N}}$, $\varphi_n \in \mathcal{C}_c^\infty(\mathbb{R}^n)$ with $\varphi_n \xrightarrow{\mathcal{C}_c^\infty(\mathbb{R}^n)} \varphi$, $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}^n)$. Using this notion of convergence, one finds that the continuous dual of the topological vector space $\mathcal{C}_c^\infty(\mathbb{R}^n)$ is the vector space of distributions.

Definition 4.1.23: (Distribution)

A distribution is a linear map $T : \mathcal{C}_c^\infty(\mathbb{R}^n) \to \mathbb{C}$, $\varphi \mapsto T[\varphi]$ that is continuous with respect to convergence in $\mathcal{D}(\mathbb{R}^n)$:

$$\varphi_i \xrightarrow{\mathcal{C}_c^\infty(\mathbb{R}^n)} \varphi \qquad \Rightarrow \qquad T[\varphi_i] \to T[\varphi] \tag{4.1}$$

The space of distributions on $\mathbb{R}^n$, denoted $\mathcal{D}^*(\mathbb{R}^n)$ is an infinite dimensional vector space with respect to pointwise addition and multiplication by $\mathbb{C}$.

Lemma 4.1.24: (Topology on $\mathcal{D}^*(\mathbb{R}^n)$, convergence of distributions)

The space of distributions $\mathcal{D}^*(\mathbb{R}^n)$ becomes a locally convex topological vector space when equipped with the topology defined by the family of semi-norms $(||T||_\varphi)_{\varphi\in\mathcal{C}_c^\infty(\mathbb{R}^n)}$, $||T||_\varphi = |T[\phi]|$.

A sequence $\{T_i\}_{i\in\mathbb{N}}$, $T_i \in \mathcal{D}^*(\mathbb{R}^n)$ of distributions converges towards a distribution $T \in \mathcal{D}^*(\mathbb{R}^n)$, $T_i \xrightarrow{\mathcal{D}^*} T$ if $T_i[\varphi] \to T[\varphi]$ for all $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}^n)$.

4.1.3 Banach spaces and Hilbert spaces

We return to normed topological vector spaces. A special class of normed topological vector spaces are Banach spaces, which are characterised by the fact that they are complete with respect to their norm.

Definition 4.1.25: (Banach space)

A *Banach space* is a normed vector space E over $k = \mathbb{C}$ or $k = \mathbb{R}$ which is *complete* with respect to the norm $||\ \ || : E \to \mathbb{R}$, i.e. any Cauchy sequence $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$, $||\boldsymbol{x}_n - \boldsymbol{x}_m|| < \epsilon$ for all $n, m \geq n_0$ converges towards an element $\boldsymbol{x} \in E$.

A map $\Phi : E \to F$ between Banach spaces $(E, ||\ \ ||_E)$ and $(F, ||\ \ ||_F)$ is called *continuous* if for all convergent sequences $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$, $\boldsymbol{x}_n \in E$ with $||\boldsymbol{x}_n - \boldsymbol{x}||_E \to 0$ the sequence of images converges $||\Phi(\boldsymbol{x}_n) - \Phi(\boldsymbol{x})||_F \to 0$.

For any Banach space, the continuous dual is also a Banach space:

Lemma 4.1.26: If E is a Banach space with norm $||\ \ ||$ then the continuous dual of E is a Banach space with the *operator norm* $||\ \ ||_{E^*} : E^* \to \mathbb{R}$, $||f||_{E^*} = \sup\{|f(x)| \mid x \in E, ||x|| \leq 1\}$.

Example 4.1.27: The space $\mathcal{C}(I)$ of continuous functions with values in $\mathbb{R}$ on a closed interval $I = [0, 1]$ with norm

$$||f|| = \int_a^b |f(t)|dt \tag{4.2}$$

is a normed space, but not a Banach space. The sequence $\{f_n\}_{n\geq 3}$

$$f_n(t) = \begin{cases} 1 & 0 \leq t \leq \frac{1}{2} \\ 0 & \frac{1}{2} + \frac{1}{n} \leq t \leq 1 \\ 1 - n(t - \frac{1}{2}) & \frac{1}{2} < t < \frac{1}{2} + \frac{1}{n} \end{cases} \tag{4.3}$$

is a Cauchy sequence with respect to the norm (4.2), but does not converge towards a function in $\mathcal{C}(I)$.

Example 4.1.28: We consider the set l^p, $1 \leq p < \infty$ of absolutely p-summable series $\boldsymbol{x} = \{x_n\}_{n\in\mathbb{N}}$, $x_n \in k$, in $k = \mathbb{R}$ or $k = \mathbb{C}$, i.e. series $\boldsymbol{x} = \{x_n\}_{n\in\mathbb{N}}$, $x_n \in k$ such that $\sum_{n=1}^\infty |x_k|^p < \infty$. These sets are vector spaces with respect to the addition and scalar multiplication

$$\{x_n\}_{n\in\mathbb{N}} + \{y_n\}_{y\in\mathbb{N}} = \{x_n + y_n\}_{n\in\mathbb{N}} \qquad t\{x_n\}_{n\in\mathbb{N}} = \{tx_n\}_{n\in\mathbb{N}}\ \forall t \in \mathbb{R}.$$

These vector spaces are Banach spaces when equipped with the norm

$$||\boldsymbol{x}||_p = \left(\sum_{n=1}^{\infty} |x_k|^p\right)^{1/p}.$$

The space l^∞ of bounded series $\boldsymbol{x} = \{x_n\}_{n\in\mathbb{N}}$, $x_n \in k$, $\sup_{n\in\mathbb{N}}\{|x_n|\} < \infty$ is a Banach space with norm

$$||\boldsymbol{x}||_\infty = \sup_{n\in\mathbb{N}}\{|x_n|\}.$$

Example 4.1.29: We consider the vector space $l_{1\infty} = l^1 \cap l^\infty$ of sequences which are both absolutely summable and bounded. The norms $||\ ||_1$ and $||\ ||_\infty$ on $l_{1\infty}$ are *not equivalent.* This can be seen by considering the sequences $\boldsymbol{x} = \{x^n\}_{n\in\mathbb{N}}$ with $x \in]0,1[$. We have $||\boldsymbol{x}||_\infty = x$ and

$$||\boldsymbol{x}||_1 = \sum_{k=1}^{\infty} x^k = \frac{1}{1-x} - 1 = \frac{x}{1-x}.$$

If we set $\boldsymbol{x}_k = \{(1-\frac{1}{k})^n\}_{n\in\mathbb{N}}$ for $k \in \mathbb{N}, k \geq 2$ we have $||\boldsymbol{x}_k||_\infty = 1 - \frac{1}{k} \xrightarrow{k\to\infty} 1$ and

$$||\boldsymbol{x}_k||_1 = \frac{1-\frac{1}{k}}{\frac{1}{k}} = k - 1 \xrightarrow{k\to\infty} \infty.$$

This implies that the norms $||\ ||_\infty$ and $||\ ||_1$ are not equivalent.

Exercise 34: (Banach spaces of series)

1. We consider the set l^∞ of all *bounded sequences* in $k = \mathbb{R}$ or $k = \mathbb{C}$, i.e. sequences $\boldsymbol{x} = \{x_k\}_{k\in\mathbb{N}}$, $x_k \in k$ for which $\sup_{k\in\mathbb{N}}\{|x_k|\} < \infty$. Show that l^∞ is a vector space with respect to pointwise addition $\{x_k\}_{k\in\mathbb{N}} + \{y_k\}_{k\in\mathbb{N}} = \{x_k + y_k\}_{k\in\mathbb{N}}$ and scalar multiplication $\alpha\{x_k\}_{k\in\mathbb{N}} = \{\alpha x_k\}_{k\in\mathbb{N}}$ for $\alpha \in k$. Show that

$$||\boldsymbol{x}||_\infty = \sup_{n\in\mathbb{N}}\{|\{x_n\}|\} \qquad \boldsymbol{x} = \{x_k\}_{k\in\mathbb{N}} \in l^\infty$$

defines a norm on l^∞ and turns l^∞ into a Banach space.

2. For $1 \leq p < \infty$, we consider the set l^p of p-summable sequences in k, i.e. the set of sequences $\{x_k\}_{k\in\mathbb{N}}$, $x_k \in k$ with $\sum_{k=1}^{\infty} |x_k|^p < \infty$. For $\{x_k\}_{k\in\mathbb{N}} \in l^p$, we define

$$||\{x_k\}||_p = \left(\sum_{k=1}^{\infty} |x_k|^p\right)^{1/p}.$$

Sequences in l^1 are called *absolutely convergent sequences*, sequences in l^2 *square summable* sequences.

Show that l^p has the structure of a vector space with respect to pointwise addition $\{x_k\}_{k\in\mathbb{N}} + \{y_k\}_{k\in\mathbb{N}} = \{x_k + y_k\}_{k\in\mathbb{N}}$ and scalar multiplication $\alpha\{x_k\}_{k\in\mathbb{N}} = \{\alpha x_k\}_{k\in\mathbb{N}}$ for $\alpha \in k$. Show that $||\ ||_p$ defines a norm on l^p.

Hint: To prove the triangle inequality, proceed as follows

- Show that for all sequences $\{x_k\}_{k\in\mathbb{N}}$, $\{y_k\}_{k\in\mathbb{N}} \in l^p$:

$$||\{x_k + y_k\}||_p^p = \sum_{k=1}^{\infty} |x_k + y_k|^p \leq \sum_{k=1}^{\infty} |x_k||x_k + y_k|^{p-1} + \sum_{k=1}^{n} |y_k||x_k + y_k|^{p-1}$$

- Apply *Hölder's inequality*

$$\sum_{k=1}^{\infty} |x_k y_k| \leq \left(\sum_{k=1}^{\infty} |x_k|^p\right)^{1/p} \left(\sum_{k=1}^{\infty} |y_k|^q\right)^{1/q} \qquad \forall p, q \geq 1, \frac{1}{p} + \frac{1}{q} = 1$$

  to each of the summands to show

$$||\{x_k + y_k\}||_p^p = \sum_{k=1}^{\infty} |x_k + y_k|^p \leq (||\{x_k\}||_p + ||\{y_k\}||_p) \frac{||\{x_n + y_n\}||_p^p}{||\{x_n + y_n\}||_p}$$

- Derive the triangle inequality.

One can show that $(l^p, ||\ \ ||_p)$ is a Banach space, i.e. complete with respect to the norm $||\ \ ||_p$.

3. We now consider the *continuous dual* of the Banach spaces $(l^p, ||\ \ ||_p)$, $1 \leq p \leq \infty$. We start by considering the case $p = \infty$. For sequences $\boldsymbol{x} = \{x_k\}_{k\in\mathbb{N}} \in l^\infty$, $\boldsymbol{y} = \{y_k\}_{k\in\mathbb{N}} \in l^1$, we set

$$L_(\boldsymbol{x}, \boldsymbol{y}) = \sum_{k=1}^{\infty} x_k y_k.$$

Show that that the series $\sum_{k=1}^{\infty} x_k y_k$ is absolutely convergent for all absolutely convergent sequences $\boldsymbol{y} = \{y_k\}_{k\in\mathbb{N}} \in l^1$ and bounded sequences $\boldsymbol{x} = \{x_k\}_{k\in\mathbb{N}}$ hence defines linear maps $L_{\boldsymbol{x}} = L(\boldsymbol{x}, \cdot) : l^1 \to k$ and $L_{\boldsymbol{y}} = L(\cdot, \boldsymbol{y}) : l^\infty \to k$. Hint: We have $|x_k y_k| \leq |y_k| \sup_{k\in\mathbb{N}}\{|x_k|\}$. Show that $L_{\boldsymbol{x}} : l^1 \to k$ is continuous for all $\boldsymbol{x} = \{x_k\}_{k\in\mathbb{N}} \in l^\infty$ and $L_{\boldsymbol{y}} : l^\infty \to k$ is continuous for all $\boldsymbol{y} = \{y_k\}_{k\in\mathbb{N}} \in l^1$. Conclude that the assignment $\boldsymbol{x} \mapsto L_{\boldsymbol{x}}$ defines a linear map from l^∞ into the continuous dual of l^1 and that the assignment $\boldsymbol{y} \mapsto L_{\boldsymbol{y}} : l^\infty \to \mathbb{R}$ defines a linear map from l^1 into the continuous dual of l^∞.

4. We now consider the case $1 < p < \infty$. For sequences $\boldsymbol{x} = \{x_k\}_{k\in\mathbb{N}} \in l^p$ and $\boldsymbol{y} = \{y_k\}_{k\in\mathbb{N}} \in l^q$ with $\frac{1}{p} + \frac{1}{q} = 1$ we set

$$L(\boldsymbol{x}, \boldsymbol{y}) = \sum_{k=1}^{\infty} x_k y_k$$

Use Hölder's inequality to show that the series $L(\boldsymbol{x}, \boldsymbol{y})$ is absolutely convergent for all $\boldsymbol{x} = \{x_k\}_{k\in\mathbb{N}} \in l^p$ and $\boldsymbol{y} = \{y_k\}_{k\in\mathbb{N}} \in l^q$ with $\frac{1}{p} + \frac{1}{q} = 1$. Show that it defines continuous linear maps $L_{\boldsymbol{x}} = L(\boldsymbol{x}, \cdot) : l^q \to k$ and $L_{\boldsymbol{y}} = L(\cdot, \boldsymbol{y}) : l^p \to k$. Conclude that the assignment $\boldsymbol{x} \mapsto L_{\boldsymbol{x}}$ defines a linear map from l^p into the continuous dual of l^q and that the assignment $\boldsymbol{y} \mapsto L\boldsymbol{y} : l^\infty \to \mathbb{R}$ defines a linear map from l^q into the continuous dual of l^p.

Remark 4.1.30: One can show that the maps $\boldsymbol{x} \to L_{\boldsymbol{x}}$ define vector space isomorphisms from l^∞ into the continuous dual of l^1 and from l^p, $1 < p < \infty$ into the continuous dual of $l^{1+\frac{1}{p-1}}$. The continuous duals of l^1 and l^∞ can therefore be identified, respectively, with the vector spaces l^∞ and l^1, and the continuous dual of l^p, $1 < p < \infty$ with $l^{1+\frac{1}{p-1}}$

An important set of Banach spaces are the Lebesgue spaces or L^p-spaces, which are related to Lebesgue integrals. To construct these spaces, we need to introduce the concepts of functions defined almost everywhere and of identities that hold almost everywhere. These are functions that are defined or identities that old everywhere on $\mathbb{R}^n$ except in a set of points which is so small that it cannot affect Lebesgue integrals. More precisely, these are functions that are defined and identities that hold everywhere in $\mathbb{R}^n$ except on a set of Lebesgue measure zero.

Definition 4.1.31: (Almost everywhere)

An identity I is said to *hold almost everywhere* if the set of points for which it does not hold $\{p \in \mathbb{R}^n \mid I \text{ not true in } p\}$ is of Lebesgue measure zero.

A function is said to be defined *almost everywhere* if it is defined on a subset $U \subset \mathbb{R}^n$ such that $\mathbb{R}^n \setminus U$ is of Lebesgue measure zero. If f is a function defined almost everywhere on $\mathbb{R}^n$ which takes values in $\mathbb{C}$ we write $f : \mathbb{R}^n \to \mathbb{C} \cup \{\infty\}$.

Two functions $f, g : \mathbb{R}^n \to \mathbb{C} \cup \{\infty\}$ are called *equal almost everywhere* if the set $\{\boldsymbol{x} \in \mathbb{R}^n : f(\boldsymbol{x}) \neq g(\boldsymbol{x})\}$ is of Lebesgue measure zero.

Example 4.1.32: (Zero sets for the Lebesgue measure)

1. If $A \subset B$, $A, B \subset \mathbb{R}^n$ and B is a set of Lebesgue measure zero, A is a set of Lebesgue measure zero.
2. Any set $S \subset \mathbb{R}^k$ contained in a k-dimensional subspace of $\mathbb{R}^n$, $k < n$ is a zero set for the Lebesgue measure on $\mathbb{R}^n$.
3. If $\{A_i\}_{i \in \mathbb{N}}$, $A_i \in \mathbb{R}^n$ are subsets of measure zero, $\bigcup_{i=1}^{\infty} A_i$ is a subset of measure zero
4. If $f, g \in C^0(\mathbb{R}^n)$ and $f = g$ almost everywhere, then $f(\boldsymbol{x}) = g(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^n$.

Using the notion of functions defined almost everywhere, we can define p-integrable and bounded functions.

Definition 4.1.33: (p-integrable functions, locally integrable functions, bounded functions)

1. For $1 \leq p < \infty$, $\mathcal{L}^p(\mathbb{R}^n)$ is the set of functions $f : \mathbb{R}^n \to \mathbb{C} \cup \{\infty\}$, for which
$$||f||_p = \left(\int_{\mathbb{R}^n} d^n x |f(\boldsymbol{x})|^p \right)^{\frac{1}{p}} < \infty. \tag{4.4}$$
Functions in $\mathcal{L}^1(\mathbb{R}^n)$ are called *integrable*, functions in $\mathcal{L}^2(\mathbb{R}^n)$ *square integrable* and functions in $\mathcal{L}^p(\mathbb{R}^n)$ *p-integrable*.
2. A function $f : \mathbb{R}^n \to \mathbb{C} \cup \{\infty\}$ is called *locally integrable* if $\int_K |f(\boldsymbol{x})| d^n x$ exists for all compact $K \subset \mathbb{R}^n$. The set of locally integrable functions is denoted $\mathcal{L}^1_{loc}(\mathbb{R}^n)$.
3. $\mathcal{L}^\infty(\mathbb{R}^n)$ is the set of locally integrable functions $f : \mathbb{R}^n \to \mathbb{C} \cup \{\infty\}$ for which
$$||f||_\infty = \inf\{c \geq 0 \mid |f| \leq c \text{ almost everywhere on } \mathbb{R}^n\} < \infty. \tag{4.5}$$
Elements of $\mathcal{L}^\infty(\mathbb{R}^n)$ are called *bounded almost everywhere* or *bounded*.

Example 4.1.34: (Integrable, locally integrable, p-integrable)

1. The function $f : \mathbb{R}^n \to \mathbb{R}$, $f(\boldsymbol{x}) = 1$ for all $\boldsymbol{x} \in \mathbb{R}^n$ is bounded and locally integrable, but not p-integrable: $f \in \mathcal{L}^1_{loc}(\mathbb{R}^n)$, $f \in \mathcal{L}^\infty(\mathbb{R}^n)$, $f \notin \mathcal{L}^p(\mathbb{R}^n)$ for all $p \in [0, \infty[$.

2. The function $f:\mathbb{R}\to\mathbb{R}$

$$f(x)=\begin{cases} x^{-\alpha} & x\in]0,1[\\ 0 & x\leq 0 \text{ or } x\geq 1\end{cases} \qquad \alpha\in\mathbb{R}^+ \tag{4.6}$$

is not bounded: $f\notin\mathcal{L}^\infty(\mathbb{R})$. It is in $\mathcal{L}^p(\mathbb{R})$, $p\in[1,\infty[$ if and only if $p\alpha<1$. It is locally integrable and integrable if and only if $\alpha<1$.

3. The function $f:\mathbb{R}\to\mathbb{R}$

$$f(x)=\begin{cases} x^{-\alpha} & x\geq 1 \\ 0 & x<1\end{cases} \qquad \alpha\in\mathbb{R}^+ \tag{4.7}$$

is bounded and locally integrable: $f\in\mathcal{L}^\infty(\mathbb{R}^n)$, $f\in\mathcal{L}^1_{loc}(\mathbb{R}^n)$. It is in $L^p(\mathbb{R}^n)$, $p\in[1,\infty[$ if and only if $p\alpha>1$.

4. $f\in\mathcal{C}^\infty_c(\mathbb{R}^n)\Rightarrow f\in\mathcal{L}^p(\mathbb{R}^n)$ for all $p\in[1,\infty]$.

5. $f\in C^0(\mathbb{R}^n)$, f bounded $\Rightarrow f\in\mathcal{L}^1_{loc}(\mathbb{R}^n)$.

Clearly, the sets $\mathcal{L}^p(\mathbb{R}^n)$, $1\leq p\leq\infty$ are vector spaces with respect to pointwise addition of functions and multiplication by $\mathbb{C}$. Also, it is clear that $||\ \ ||:\mathcal{L}^p(\mathbb{R}^n)\to\mathbb{R}$ is positive semi-definite and satisfies $||\lambda f||_p=|\lambda|||f||_p$. It can also be shown that it satisfies the triangle inequality. However, it does not define a norm on $\mathcal{L}^p(\mathbb{R}^n)$ because it fails to be positive-definite.

Lemma 4.1.35: For $1\leq p\leq\infty$, the spaces $\mathcal{L}^p(\mathbb{R}^n)$ with pointwise addition of functions and multiplication by $\mathbb{C}$ are vector spaces. $||\cdot||_p$ defines a semi-norm on $\mathcal{L}^p(\mathbb{R}^n)$. It satisfies $||f||_p=0$ if and only if $f=0$ almost everywhere on $\mathbb{R}^n$.

Hence, in order to obtain a normed vector space, we have to take the quotient of $\mathcal{L}^p$ with respect to the linear subspace of functions which vanish almost everywhere on $\mathbb{R}^n$ as in lemma 4.1.17. This leads to the definition of L^p-spaces of Lebesgue spaces.

Definition 4.1.36: (L^p-spaces, Lebesgue spaces)

The L^p-space or Lebesgue space $L^p(\mathbb{R}^n)$, $1\leq p\leq\infty$, is the quotient vector space $\mathcal{L}^p(\mathbb{R}^n)/\mathcal{N}$, $\mathcal{N}=\{f:\mathbb{R}^n\to\mathbb{C}^\cup\{\infty\}, f=0 \text{ almost everywhere on } \mathbb{R}^n\}$. Its elements are equivalence classes $[f]=f+\mathcal{N}$ with respect to the equivalence relation $f\sim g$ if $f=g$ almost everywhere on $\mathbb{R}^n$ or, equivalently, $f-g\in\mathcal{N}$. The semi-norm $||\ \ ||_p$ on $\mathcal{L}^p(\mathbb{R}^n)$ induces a norm on $L^p(\mathbb{R}^n)$.

With this definition, one obtains a class of normed vector spaces $(L^p(\mathbb{R}^n),||\ \ ||_p)$ which additionally, are complete with respect to the norm $||\ \ ||_p$ and therefore Banach spaces.

Theorem 4.1.37: (Lebesgue spaces are Banach spaces)

For all $1\leq p\leq\infty$ the Lebesgue spaces $L^p(\mathbb{R}^n)$ with the L^p-norm $||\ \ ||_p:L^p(\mathbb{R}^n)\to\mathbb{R}$ are Banach spaces.

The vector space of equivalence classes $\mathcal{C}^\infty_c(\mathbb{R}^n)/\mathcal{N}$ is *dense* in $L^p(\mathbb{R}^n)$, i.e. for any $f\in L^p(\mathbb{R}^n)$, there exists a sequence $\{\varphi_n\}_{n\in\mathbb{N}}$, $\varphi\in\mathcal{C}^\infty_c(\mathbb{R}^n)/\mathcal{N}$ with $||\varphi_n-f||_p\xrightarrow{n\to\infty}0$

Theorem 4.1.38: The *continuous dual* of the Banach spaces $L^p(\mathbb{R}^n)$ with $1 \leq p \leq \infty$ is isomorphic to the Banach space $L^q(\mathbb{R}^n)$, where $\frac{1}{p} + \frac{1}{q} = 1$. (For $p = \infty$ interpret this equation as $q = 1$.). The isomorphism $\phi^p : L^q(\mathbb{R}^n) \to L^p(\mathbb{R}^n)^*$ is given by

$$f \in L^q(\mathbb{R}^n) \mapsto L_f : L^p(\mathbb{R}^n) \to \mathbb{C}, L_f(g) := \int_{\mathbb{R}^n} f(\boldsymbol{x})g(\boldsymbol{x})d^n x.$$

A special class of Banach spaces are Hilbert spaces, which are Banach spaces E whose norm is induced by a hermitian product on E.

Definition 4.1.39: (Scalar product, Hilbert space, Pre-Hilbert space)

A *scalar product* or *hermitian product* on a vector space E over $k = \mathbb{R}$ or $k = \mathbb{C}$ is a positive definite hermitian form on E, i.e. a map $\langle\, ,\, \rangle : E \times E \to \mathbb{C}$ that satisfies

1. $\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle$, $\langle \boldsymbol{z}, \boldsymbol{x} + \boldsymbol{y} \rangle = \langle \boldsymbol{z}, \boldsymbol{x} \rangle + \langle \boldsymbol{z}, \boldsymbol{y} \rangle$ for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in E$.
2. $\langle \lambda \boldsymbol{x}, \boldsymbol{y} \rangle = \lambda \langle \boldsymbol{x}, \boldsymbol{y} \rangle$, $\langle \boldsymbol{x}, \lambda \boldsymbol{y} \rangle = \bar{\lambda} \langle \boldsymbol{x}, \boldsymbol{y} \rangle$ for all $\boldsymbol{x}, \boldsymbol{y} \in E$ and $\lambda \in \mathbb{C}$
3. $\langle \boldsymbol{y}, \boldsymbol{x} \rangle = \overline{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}$ for all $\boldsymbol{x}, \boldsymbol{y} \in E$
4. $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$ for all $\boldsymbol{x} \in E$, $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$ if and only if $\boldsymbol{x} = 0$

A vector space E together with a *scalar product* is called a *Pre-Hilbert space*. It is called a *Hilbert space* if it is complete with respect to the norm $||\boldsymbol{x}|| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$, i.e. any Cauchy sequence with respect to this norm converges towards an element of $\mathcal{H}$.

One can show that all Hilbert spaces are Banach spaces, but not the other way around.

Lemma 4.1.40: If $(E, \langle\, ,\, \rangle)$ is a pre-Hilbert space, the map $||\cdot|| : E \to \mathbb{R}_0^+$, $||\boldsymbol{x}|| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$ defines a norm on E. If $(E, \langle\, ,\, \rangle)$ is a Hilbert space, $(E, ||\cdot||)$ is a Banach space.

Proof: Exercise. □

Exercise 35: Let E with scalar product $\langle\, ,\, \rangle : E \times E \to \mathbb{C}$ be a Pre-Hilbert space and denote by $||\boldsymbol{x}|| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$ the associated norm. Prove the following identities

1. *Cauchy Schwartz inequality* $|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leq ||\boldsymbol{x}|| \cdot ||\boldsymbol{y}|| \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in E$
2. *Triangle inequality*: $||\boldsymbol{x} + \boldsymbol{y}|| \leq ||\boldsymbol{x}|| + ||\boldsymbol{y}||$
3. $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \frac{1}{2}(||\boldsymbol{x} + \boldsymbol{y}||^2 - ||\boldsymbol{x}||^2 - ||\boldsymbol{y}||^2)$
4. $||\boldsymbol{x}||^2 + ||\boldsymbol{y}||^2 = \frac{1}{2}||\boldsymbol{x} + \boldsymbol{y}||^2 + \frac{1}{2}||\boldsymbol{x} - \boldsymbol{y}||^2$

Hints: To prove the Cauchy Schwartz inequality, consider the cases $||\boldsymbol{x}||, ||\boldsymbol{y}|| \neq 0$ and the case where $||\boldsymbol{x}|| = 0$ or $||\boldsymbol{y}|| = 0$ separately. For $||\boldsymbol{y}|| \neq 0$, consider the norm squared of the vector $\boldsymbol{x} - \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{||\boldsymbol{y}||^2}\boldsymbol{y}$. To prove the triangle inequality, square both sides of the equation, and use the Cauchy Schwartz inequality.

We will now consider some important examples of Hilbert spaces as well as an example of Pre-Hilbert spaces, which is not a Hilbert space.

Example 4.1.41: (Hilbert spaces)

1. The vector space $\mathbb{C}^n$ with scalar product $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \sum_{i=1}^n \bar{y}_i x_i$ is a Hilbert space
2. The vector space $L^2(\mathbb{R}^n)$ is a is Hilbert space with the scalar product
$$\langle f, g\rangle = \int_{\mathbb{R}^n} \bar{f}(\boldsymbol{x}) g(\boldsymbol{x}) d^n x. \tag{4.8}$$
3. The space $l^2 = \{\{x_n\}_{n\in\mathbb{N}} \mid \sum_{n=0}^\infty |x_n|^2 < \infty\}$ of *square summable* sequences is a Hilbert space with the scalar product
$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \sum_{n=1}^\infty \bar{x}_n y_n \qquad \boldsymbol{x} = (x_n)_{n\in\mathbb{N}}, \boldsymbol{y} = (y_n)_{n\in\mathbb{N}}. \tag{4.9}$$

Remark 4.1.42: The Hilbert space of absolutely convergent series can be viewed as the infinite-dimensional analogue of the vector spaces $\mathbb{C}^n$. We will see later that any *separable* Hilbert space can be identified with this space via the choice of a (countably infinite) orthonormal basis.

Example 4.1.43: (Pre-Hilbert space but not Hilbert space) We consider the space $\mathcal{C}(I)$ from example 4.1.27 with scalar product
$$\langle f, g\rangle = \int_a^b f(t) g(t) \tag{4.10}$$
and associated norm $|| \, || = \sqrt{\langle \, , \, \rangle}$. This is a pre-Hilbert space, but not a Hilbert space since the sequence (4.3) given in example 4.1.27 is a Cauchy sequence but does not converge.

We will now investigate the continuous duals of Hilbert spaces. We start with the observation that for the Hilbert spaces l^2 of square summable series and the Hilbert space $L^2(\mathbb{R}^n)$ of square integrable functions, the continuous duals were determined already in the previous sections. There, we found that these Hilbert spaces are *self-dual*, i.e. that their continuous dual is isomorphic to the spaces itself $l^2* = l^2$, $L^2(\mathbb{R}^n)^* = L^2(\mathbb{R}^n)$. Riesz's theorem asserts that that this is a general pattern for *all Hilbert spaces. Hilbert spaces are self-dual.*

Theorem 4.1.44: (Continuous dual of Hilbert spaces, Riesz's theorem)

Let E be a Hilbert space with hermitian product $\langle \, , \, \rangle : E \times E \to \mathbb{C}$. Then, the continuous dual E^* of E is isomorphic to E. More precisely, the linear map
$$R : E \to E^* \, , \, \boldsymbol{x} \mapsto \langle \boldsymbol{x}, \cdot \rangle$$
is a continuous and bijective and satisfies $||R\boldsymbol{x}||_{E^*} = ||\boldsymbol{x}||$, where the norm $|| \;\; ||_{E^*}$ on E^* is given by $||f|| = \sup\{|f(\boldsymbol{x})| \mid \boldsymbol{x} \in E, ||\boldsymbol{x}|| \leq 1\}$.

4.1.4 Summary

In this section, we considered examples of topological vector spaces, which included locally convex vector spaces associated with distributions, metric and metrizable spaces as well as normed spaces, which as special cases included Banach and Hilbert spaces.

The relation between these spaces is given by the following diagram:

$$\begin{array}{ll}
\text{Hilbert spaces} & \text{Examples: } l^2, L^2(\mathbb{R}^n), \mathbb{C}^n \\
\cap & \\
\text{Banach spaces} & \text{Examples: } l^p, L^p(\mathbb{R}^n), 1 \leq p \leq \infty \\
\cap & \\
\text{Normed vector spaces} & \text{Examples: } C([a,b]) \text{ with norm } ||f|| = \int_a^b |f(t)| \\
\cap & \\
\text{Metric vector spaces/metrizable vector spaces} & \\
\cap & \\
\text{Locally convex vector spaces} & \text{Examples: } \mathcal{C}_c^\infty(\mathbb{R}^n), \mathcal{D}^*(\mathbb{R}^n) \\
\cap & \\
\text{Topological vector spaces.} &
\end{array} \tag{4.11}$$

We also investigated the continuous duals for three examples of topological vector spaces which play an important role in mathematics and physics. The results are summarised in the following table.

Example	Structure	Continuous dual
$L^p(\mathbb{R}^n)$, $1 \leq p \leq \infty$	Banach space	$L^q(\mathbb{R}^n)$, $\frac{1}{p} + \frac{1}{q} = 1$
l^p, $1 \leq p \leq \infty$	Banach space	l^q, $\frac{1}{p} + \frac{1}{q} = 1$
$L^2(\mathbb{R}^n)$	Hilbert space	$L^2(\mathbb{R}^n)$
l^2	Hilbert space	l^2
$\mathcal{C}_c^\infty(\mathbb{R}^n)$	Locally convex vector space	Distributions $\mathcal{D}^*(\mathbb{R}^n)$

This should be contrasted with the situation for finite-dimensional vector spaces

- Every finite dimensional vector space has a *unique* topology, the *canonical topology*, that turns it into a topological vector space
- Any finite-dimensional vector space can be given the structure of a normed space, and any two norms on a finite-dimensional vector space are equivalent.
- Every finite-dimensional vector space over $\mathbb{C}$ can be made into a Hilbert space via the choice of a basis.
- The linear forms $\phi : V \to k$ on a finite-dimensional vector space V over $k = \mathbb{C}$ or $k = \mathbb{R}$ are continuous. The dual vector space V^* is isomorphic to V.

4.2 Distributions

In this section, we will give a more detailed investigation of smooth functions with compact support and distributions. We recall the definitions from the previous section and start by considering some examples of distributions.

Example 4.2.1: (Distributions)

1. A function $f : \mathbb{R}^n \to \mathbb{C} \cup \{\infty\}$ is called *locally integrable*, $f \in \mathcal{L}^1(\mathbb{R}^n)$ if $\int_K |f(\boldsymbol{x})| d^n x$ exists for all compact subsets $K \subset \mathbb{R}^n$. The space of locally integrable functions on $\mathbb{R}^n$ is denoted by $\mathcal{L}^1(\mathbb{R}^n)$. For any $f \in \mathcal{L}^1$, the the integral $\int_{\mathbb{R}^n} f(\boldsymbol{x})\varphi(\boldsymbol{x}) d^n x$ exists for all test functions $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}^n)$, since

$$\int_{\mathbb{R}^n} f(\boldsymbol{x})\varphi(\boldsymbol{x}) d^n x = \int_{\mathrm{Supp}(\varphi)} f(\boldsymbol{x})\varphi(\boldsymbol{x}) d^n x \leq \int_{\mathrm{Supp}(\varphi)} |f(\boldsymbol{x})||\varphi(\boldsymbol{x})| d^n x$$
$$\leq \mathrm{Sup}\{|\varphi(\boldsymbol{x}) \mid \boldsymbol{x} \in \mathrm{Supp}(\varphi)\} \cdot \int_{\mathrm{supp}(\varphi)} |f(\boldsymbol{x})| d^n x < \infty.$$

The linear map

$$T_f : \varphi \mapsto T_f[\varphi] := \int_{\mathbb{R}^n} f(x)\varphi(x) d^n x$$

is a distribution.

2. For any $\boldsymbol{x} \in \mathbb{R}^n$, the evaluation map $\delta_{\boldsymbol{x}} : \varphi \mapsto \delta_{\boldsymbol{x}}(\varphi) = \varphi(\boldsymbol{x})$ is a distribution

Definition 4.2.2: (Transformation of distributions)

The transformation $T \to \phi^* T$ of a distribution T under a diffeomorphism $\phi : \mathbb{R}^n \to \mathbb{R}^n$ is defined by

$$(\phi^* T)[\varphi \circ \phi \cdot |\det(d\phi)|] = T[\varphi] \qquad \forall \varphi \in \mathcal{C}_c^\infty(\mathbb{R}^n), T \in \mathcal{D}^*(\mathbb{R}^n).$$

Exercise 36: (Transformations of the δ-distribution)

We consider the delta distribution $\delta_{\boldsymbol{x}}$ on $\mathbb{R}^n$. Let $\phi = (\phi_1, ..., \phi_n) : \mathbb{R}^n \to \mathbb{R}^n$ be a diffeomorphism, i.e. a bijective map that is continuously differentiable and such that its inverse ϕ^{-1} is continuously differentiable. This implies that its *Jacobi matrix* $d\phi = (\partial_i \phi_j)_{i,j=1,\dots,n}$ is invertible everywhere and therefore $\det(d\phi) \neq 0$, $\det(d\phi) = \frac{1}{\det(d\phi^{-1})}$.

The *transformation of the delta distribution* $\delta_{\boldsymbol{x}} \to \phi^* \delta \boldsymbol{x}$ under the diffeomorphism ϕ is given by the equation

$$(\phi^* \delta_{\boldsymbol{x}})[\varphi \circ \phi \cdot |\det(d\phi)|] = \delta_{\boldsymbol{x}}[\varphi] \qquad \forall \varphi \in \mathcal{C}_c^\infty(\mathbb{R}^n). \tag{4.12}$$

1. In the physics literature, the evaluation of the delta distribution on a function $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}^n)$ is often expressed using an integral notation, in which $\delta_{\boldsymbol{x}}$ is treated as if it was a function

$$\delta_{\boldsymbol{x}}[\varphi] = \int_{\mathbb{R}^n} \delta_{\boldsymbol{x}}(\boldsymbol{y})\varphi(\boldsymbol{y}) d^n y.$$

Use the transformation formula for integrals to show that in the integral notation formula (4.12) takes the form

$$\int_{\mathbb{R}^n} \phi^* \delta_{\boldsymbol{x}}(\boldsymbol{y}) f(\phi(\boldsymbol{y})) |\det(d\phi)(\boldsymbol{y})| d^n y = \int_{\mathbb{R}^n} \delta_{\boldsymbol{x}}(\boldsymbol{v}) f(\boldsymbol{v}) d^n v.$$

2. We consider the diffeomorphism $\phi : \mathbb{R}^n \to \mathbb{R}^n$, $\phi(\boldsymbol{y}) = \alpha \boldsymbol{y}$ for all $\boldsymbol{x} \in \mathbb{R}^n$ with $\alpha \in \mathbb{R} \setminus \{0\}$. Show that the transformation of the delta distribution under ϕ is given by

$$\phi^* \delta_0 = \frac{1}{|\alpha|^n} \delta_0$$

or, equivalently, in physics notation, by

$$\delta_0(\alpha\boldsymbol{x}) = \frac{1}{|\alpha|^n}\delta_0(\boldsymbol{x}).$$

3. Show that for a general diffeomorphism $\phi : \mathbb{R}^n \to \mathbb{R}^n$ as above, one has

$$\phi^*\delta_0[\varphi] = \frac{1}{|\det(d\phi)(\phi^{-1}(0))|}\varphi(\phi^{-1}(0)) \quad \Leftrightarrow \quad \delta_0(\phi(\boldsymbol{x})) = \frac{1}{|\det(d\phi)|}\delta_{\phi^{-1}(0)}(\boldsymbol{x}).$$

As distributions are continuous linear forms on the vector space $\mathcal{C}_c^\infty(\mathbb{R}^n)$, i.e. form the *continuous dual* of the space of smooth functions with compact support, many operations on on the space $\mathcal{C}_c^\infty(\mathbb{R}^n)$ can be transported to the space $\mathcal{D}^*(\mathbb{R}^n)$ of distributions via the evaluation map ev : $\mathcal{D}^*(\mathbb{R}^n) \times \mathcal{C}_c^\infty(\mathbb{R}^n) \to \mathbb{C}$, $ev(T, \varphi) = T[\varphi]$.

An example we encountered in the previous section was the notion of convergence of distributions. We will now show that the evaluation map can also be used to obtain a notion of *differentiation* for distributions. For this, we first consider distributions associated with differentiable functions $f \in \mathcal{L}^1(\mathbb{R}) \cap C^1(\mathbb{R})$ as in example 4.2.1. It is natural to impose that the differential of such a distribution should be the distribution associated with the function $\dot{f}$, if $\dot{f}$ is in $L^1(\mathbb{R})$. For $f \in \mathcal{L}^1(\mathbb{R}) \cap C^1(\mathbb{R})$ such that $\dot{f} \in \mathcal{L}^1(\mathbb{R})$ and $\varphi \in \mathcal{C}_c^\infty(\mathbb{R})$, we obtain by partial integration

$$T_{\dot{f}}[\varphi] = \int_{\mathbb{R}} \dot{f}(x) \cdot \varphi(x)\, dx = -\int_{\mathbb{R}} f(x) \cdot \dot{\varphi}(x)\, dx \tag{4.13}$$

To generalise the differentiation of distributions to all distributions on $\mathbb{R}$, we can therefore define the derivative $\dot{T}$ of a distribution T as the distribution $\dot{T} : \varphi \mapsto -T[\dot{\varphi}]$ for all $\varphi \in \mathcal{C}_c^\infty(\mathbb{R})$. Clearly, this is a continuous function of the test functions.

Note that the minus sign is essential in this definition to obtain agreement with (4.13). To generalise this definition to distributions on $\mathbb{R}^n$ and to more general derivatives and linear differential operators, we therefore need to find a suitable generalisation of this minus sign. This generalisation is given by the concept of the adjoint of a linear differential operator.

Definition 4.2.3: (Adjoint of a linear differential operator)

Let L be a linear differential operator of order k on $\mathbb{R}^n$, $L = \sum_{|\alpha|\leq k} c_\alpha D^\alpha$, where $\alpha = (j_1, ..., j_k) \in \mathbb{N}^k$ is a multi-index of length $|\alpha| = k$ and $c_\alpha \in \mathcal{C}^\infty(\mathbb{R}^n)$ with values in $\mathbb{R}$. Then a linear differential operator L^* of order k is called *adjoint* to L if for all $f \in \mathcal{C}^k(\mathbb{R}^n)$, $g \in \mathcal{C}_c^k(\mathbb{R}^n)$

$$\int_{\mathbb{R}^n} (L^*f)(\boldsymbol{x}) \cdot g(\boldsymbol{x}) d^n x = \int_{\mathbb{R}^n} f(\boldsymbol{x}) \cdot (Lg)(\boldsymbol{x}) d^n x. \tag{4.14}$$

Clearly, the definition of the adjoint generalises identity (4.13). Moreover, it has the following important properties.

Lemma 4.2.4: (Properties of the adjoint)

Every linear differential operator of order k on $\mathbb{R}^n$ $L = \sum_{|\alpha|\leq k} c_\alpha D^\alpha$, $c_\alpha \in \mathcal{C}^\infty(\mathbb{R}^n)$ with values in $\mathbb{R}$, has a unique adjoint L^*. It satisfies

$$(\lambda L)^* = \lambda L^* \quad \forall \lambda \in \mathbb{R} \qquad (L_1 + L_2)^* = L_1^* + L_2^* \qquad (L_1 \circ L_2)^* = L_2^* \circ L_1^* \tag{4.15}$$

Using the concept of the adjoint, we can now define differentiation of distribution for general differential operators on $\mathbb{R}^n$.

Definition 4.2.5: (Differentiation of distributions, weak differentiation)

Let L be a linear differential operator of order k on $\mathbb{R}^n$, $T \in \mathcal{D}^*(\mathbb{R}^n)$. Then, the action of L on T is defined by $LT[\varphi] = T[L^*\varphi]$ for all $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}^n)$.

Remark 4.2.6: Note that this definition is similar to the definition of the dual of a linear map in in Def. 1.1.7, but that there is one important difference. In Def. 1.1.7, the dual ϕ^* of a linear map $\phi \in \mathrm{End}(V)$ was defined by

$$\phi^* : \alpha \mapsto \phi^*(\alpha) \qquad \phi^*(\alpha)(\boldsymbol{x}) = \alpha(\phi(\boldsymbol{x})) \qquad \forall \alpha \in V^*, \boldsymbol{x} \in V. \tag{4.16}$$

If we set $V = \mathcal{C}_c^\infty(\mathbb{R}^n)$, $V^* = \mathcal{D}^*(\mathbb{R}^n)$ and identified the linear map ϕ with a linear operator L, Def. 1.1.7 would lead to the definition

$$LT[\varphi] = T[L\varphi] \qquad \forall \varphi \in \mathcal{C}_c^\infty(\mathbb{R}^n). \tag{4.17}$$

However, this would imply

$$(L \circ M)T[\varphi] = T[L \circ M\varphi] = LT[M\varphi] = M \circ (LT)[\varphi] \qquad \forall \varphi \in \mathcal{C}_c^\infty(\mathbb{R}^n). \tag{4.18}$$

In order to obtain the familiar composition law $(M \circ L)T = M \circ (LT)$, we therefore have to work with the adjoint instead.

Example 4.2.7: We consider the delta distribution $\delta_0 : \varphi \mapsto \varphi(0)$ for all $\varphi \in \mathcal{D}(\mathbb{R})$. Its derivative is the distribution $\dot{\delta}_0 : \varphi \mapsto -\dot{\varphi}(0)$.

Example 4.2.8: We consider the Heavyside step function on $\mathbb{R}$

$$\theta(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}. \tag{4.19}$$

As it is discontinuous in $x = 0$, it is not differentiable *as a function*. However, as $\theta \in \mathcal{L}^1_{loc}(\mathbb{R})$, it has an associated distribution $T_\theta : \varphi \mapsto \int_{-\infty}^\infty \theta(x)\varphi(x)dx$, which is differentiable as a distribution with $\dot{T}_\theta = \delta_0$

$$\dot{T}_\theta[\varphi] = -\int_{-\infty}^\infty \theta(x)\dot{\varphi}(x)dx = -\int_0^\infty \dot{\varphi}(x)dx = \varphi(0). \tag{4.20}$$

Exercise 37: We consider the function $f : \mathbb{R}^3 \to \mathbb{R}$ defined by

$$f(\boldsymbol{x}) = \begin{cases} \frac{1}{|\boldsymbol{x}|} & \boldsymbol{x} \in \mathbb{R}^n \setminus \{0\} \\ 0 & \boldsymbol{x} = 0 \end{cases} \qquad \forall \boldsymbol{x} \in \mathbb{R}^3.$$

This function is not differentiable in 0 *as a function*, but we can consider the associated distribution $T : \varphi \mapsto \int_{\mathbb{R}^3} \varphi(\boldsymbol{x})f(\boldsymbol{x})d^3x$. Determine the derivative of this distribution.

The action of a differential operator of order 0, for which $\alpha = \emptyset$ and $c_\alpha = f \in \mathcal{C}^\infty(\mathbb{R}^n)$, on a distribution corresponds to the product of a distribution with a function $f \in \mathcal{C}^\infty(\mathbb{R}^n)$.

Definition 4.2.9: (Product of a distribution and a function)

For any distribution $T : \mathcal{C}_c^\infty(\mathbb{R}^n) \to \mathbb{R}$ and any function $\psi \in \mathcal{C}^\infty(\mathbb{R}^n)$ the map $T\psi : \varphi \mapsto T[\psi \cdot \varphi]$ is a distribution: $T\psi \in \mathcal{D}^*(\mathbb{R}^n)$. It is called the *product of T and ψ*.

Remark 4.2.10: Although distributions can be multiplied by functions, the product of distributions is *a priori ill-defined*. Giving a meaning to the product of distributions and defining it consistently is a difficult task. It is related to the problem of renormalisation in quantum field theory. One should therefore never *naively* multiply two distributions.

Lemma 4.2.11: The differential of distributions is continuous with respect to the convergence of distributions. For a linear differential operator of order k on $\mathbb{R}^n$ $L = \sum_{|\alpha|\leq k} c_\alpha D^\alpha$, $c_\alpha \in \mathcal{C}^\infty(\mathbb{R}^n)$ with values in $\mathbb{R}$ and any sequence of distributions $\{T_n\}_{n\in\mathbb{N}}$, $T_n \in \mathcal{D}^*(\mathbb{R}^n)$ we have

$$T_n \xrightarrow{\mathcal{D}^*(\mathbb{R}^n)} T \quad \Rightarrow \quad LT_n \xrightarrow{\mathcal{D}^*(\mathbb{R}^n)} LT. \tag{4.21}$$

The next concept we want to extend to distributions is the notion of translation in the argument of a function. Given a function $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}^n)$, the *translate* of φ by $\boldsymbol{x} \in \mathbb{R}^n$ is the function $t_{\boldsymbol{x}}\varphi(\boldsymbol{y}) = \varphi(\boldsymbol{x} - \boldsymbol{y})$ which is again an element of $\mathcal{C}_c^\infty(\mathbb{R}^n)$. The translation operator $t_{\boldsymbol{x}} : \mathcal{C}_c^\infty(\mathbb{R}^n) \to \mathcal{C}_c^\infty(\mathbb{R}^n)$, $\varphi \mapsto t_{\boldsymbol{x}}\varphi$ defines a linear map on $\mathcal{C}_c^\infty(\mathbb{R}^n)$. As the Lebesgue measure on $\mathbb{R}^n$ is characterised by the property that it is invariant under translations, the translation and the associated translation operator play an important role in integration theory. Extending the concept of translation to distributions leads to the concept of convolution.

Definition 4.2.12: (Convolution)

For $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}^n)$, $T \in \mathcal{D}^*(\mathbb{R}^n)$ the *convolution* of T and φ turns out to be a the function $T * \varphi : \mathbb{R}^n \to \mathbb{C}$ given by

$$T * \varphi(\boldsymbol{x}) = T[t_{\boldsymbol{x}}\varphi] \; \forall \boldsymbol{x} \in \mathbb{R}^n \quad \text{where} \quad t_{\boldsymbol{x}}\varphi(\boldsymbol{y}) = \varphi(\boldsymbol{x} - \boldsymbol{y}). \tag{4.22}$$

Example 4.2.13: For $f \in \mathcal{L}^1_{loc}(\mathbb{R}^n)$, $T_f * \varphi(\boldsymbol{x}) = \int_{\mathbb{R}^n} f(\boldsymbol{y})\varphi(\boldsymbol{x} - \boldsymbol{y}) d^n y$.

Example 4.2.14: We consider the delta distribution $\delta_0 \in \mathcal{D}^*(\mathbb{R})$. We have $\delta_0 * \varphi(x) = \delta_0[t_x\varphi] = \varphi(x)$ and therefore $\delta_0 * \varphi = \varphi$ for all $\varphi \in \mathcal{C}_c^\infty(\mathbb{R})$. The delta distribution therefore plays the role of the *unit* for the convolution.

The delta distribution and convolutions play an essential role in the solution of inhomogeneous differential equations. This is due to the fact that the convolution commutes with differentiation in the following sense:

Theorem 4.2.15: (Properties of the convolution)

1. The convolution is linear in both arguments

$$(\lambda T_1 + \mu T_2) * \varphi = \lambda T_1 * \varphi + \mu T_2 * \varphi \qquad \forall \lambda, \mu \in \mathbb{R}, T_1, T_2 \in \mathcal{D}^*(\mathbb{R}^n), \varphi \in \mathcal{C}_c^\infty(\mathbb{R}^n)$$
$$T * (\lambda\varphi_1 + \mu\varphi_2) = \lambda T * \varphi_1 + \mu T * \varphi_2 \qquad \forall \lambda, \mu \in \mathbb{R}, T \in \mathcal{D}^*(\mathbb{R}^n), \varphi_1, \varphi_2 \in \mathcal{C}_c^\infty(\mathbb{R}^n).$$

2. For all $T \in \mathcal{D}^*(\mathbb{R}^n)$, $\varphi \in \mathcal{C}_c^\infty(\mathbb{R}^n)$: $T * \varphi \in \mathcal{C}^\infty(\mathbb{R}^n)$ and $\partial_i(T * \varphi) = (\partial_i T) * \varphi = T * (\partial_i \varphi)$

Theorem 4.2.16: Let $L = \sum_{|\alpha|\leq k} c_\alpha D^\alpha$ be a differential operator of order k with constant coefficients $c_\alpha \in \mathbb{C}$. Let $E \in \mathcal{D}^*(\mathbb{R}^n)$ be an *fundamental solution of* L, i.e. a distribution such that

$$LE = \delta_0. \tag{4.23}$$

Then for any $\rho \in \mathcal{C}_c^\infty(\mathbb{R}^n)$, the function $u = E * \rho : \mathbb{R}^n \to \mathbb{C}$ is a solution of the inhomogeneous differential equation

$$Lu = \rho.$$

Proof: This is a direct consequence of the second property in Theorem 4.2.15 and the fact that the delta function acts as a unit for the convolution. Together, they imply

$$Lu = L(E * \rho) = (LE) * \rho = \delta_0 * \rho = \rho$$

. □

Remark 4.2.17: In physics, fundamental solutions of differential operators are often called *Green's functions*. The act of determining a fundamental solution for a linear differential operator with constant coefficients is sometimes called "solving the equations of motions for elementary excitations". The picture behind this is that the the differential operator determines the equation of motion of a physical system without external forces. The delta distribution has the interpretation of an external force or force field which acts only at a given point in space and/or at a given time.

The solution for more general forces is then found using the *superposition* principle for linear differential equations, namely the fact that their solutions form an infinite dimensional vector space and the fact that the delta distribution acts as a unit for the convolution. This corresponds to taking the convolution of an elementary solution with the function ρ.

Example 4.2.18: (Elementary solutions for the elliptic Laplace operator)

We consider the elliptic Laplace operator $\Delta = \sum_{i=1}^n (\partial_i)^2$ on $\mathbb{R}^n$. The elementary solutions of the *Laplace equation* $\Delta E = \delta_0$ in $\mathbb{R}^n$ are the *Newton potentials* $N_{\boldsymbol{a}} : \mathbb{R}^n\{\boldsymbol{a}\} \to \mathbb{R}$

$$N_{\boldsymbol{a}}(\boldsymbol{x}) = \begin{cases} \alpha_n \frac{1}{||\boldsymbol{x}-\boldsymbol{a}||^{n-2}} & n \neq 2 \\ \frac{1}{2\pi} \ln ||\boldsymbol{x} - \boldsymbol{a}|| & n = 2 \end{cases},$$

where $\alpha_n \in \mathbb{R}^+$ is a normalisation constant related to the volume of the $n-1$-sphere S^{n-1}.

Example 4.2.19: (Elementary solutions for the Helmholtz equation)

Using the ansatz $f(\boldsymbol{x}, t) = u(\boldsymbol{x})e^{i\omega t}$, $u : \mathbb{R}^3 \to \mathbb{C}$ the hyperbolic *wave equation* on $\mathbb{R}^3 \times \mathbb{R}$

$$\left(\Delta - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right) f(\boldsymbol{x}, t) = 0$$

leads to the elliptic Helmholtz operator $\Delta + k^2$ and to the Helmholtz equation

$$(\Delta + k^2)u = 0 \qquad k = \tfrac{\omega}{c}.$$

The elementary solutions of the Helmholtz operator are given by

$$E(\boldsymbol{x}) = -\frac{1}{4\pi}\frac{\cos(k||\boldsymbol{x}||)}{||\boldsymbol{x}||}.$$

4.3 Fourier transforms

In this section, we will investigate the Fourier transform as a continuous linear map on different vector spaces. We start by considering the space $L^1(\mathbb{R}^n)$.

Definition 4.3.1: (Fourier transform)

For $f \in \mathcal{L}^1(\mathbb{R}^n)$, the integral

$$\hat{f}(\boldsymbol{y}) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} f(x) e^{-i\langle \boldsymbol{x}, \boldsymbol{y} \rangle} d^n x \qquad \langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^n x^i y^i \tag{4.24}$$

exists. The function $\hat{f} : \mathbb{R}^n \to \mathbb{C}$, $\boldsymbol{y} \mapsto \hat{f}(\boldsymbol{y})$ is called the *Fourier transform* of f.

Example 4.3.2:

We determine the Fourier transform of the function $f : \mathbb{R}^n \to \mathbb{R}$, $f(\boldsymbol{x}) = e^{-\boldsymbol{x}^2/2}$.

$$\begin{aligned} \hat{f}(\boldsymbol{y}) &= \tfrac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{-\boldsymbol{x}^2/2} e^{-i\boldsymbol{x}\cdot\boldsymbol{y}} d^n x = \prod_{i=1}^n \int_{-\infty}^{\infty} \tfrac{1}{\sqrt{2\pi}} e^{-x_i^2/2 - i x_i y_i} dx_i \\ &= \prod_{i=1}^n \int_{-\infty}^{\infty} \tfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i + i y_i)^2 - \frac{1}{2} y_i^2} dx_i = e^{-\frac{1}{2}\boldsymbol{y}^2} \left(\tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx \right)^n = e^{-\frac{1}{2}\boldsymbol{y}^2}. \end{aligned} \tag{4.25}$$

Example 4.3.3:

We determine the Fourier transform of the function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = e^{-|x|}$.

$$\begin{aligned} \hat{f}(y) =& \tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-|x|} e^{-ixy} dx = \tfrac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-|x|} (e^{-ixy} + e^{ixy}) dx \\ =& \tfrac{1}{\sqrt{2\pi}} \lim_{R\to\infty} \left[-\frac{e^{-x(1+iy)}}{1+iy} - \frac{e^{-x(1-iy)}}{1-iy} \right]_{x=0}^{x=R} = \sqrt{\tfrac{2}{\pi}} \frac{1}{1+y^2} \end{aligned} \tag{4.26}$$

Exercise 38: Show that for the *characteristic function* $\chi_{[-1,1]} : \mathbb{R} \to \mathbb{R}$ the Fourier transform is given by

$$\chi_{[-1,1]}(x) = \begin{cases} 1 & |x| \leq 1 \\ 0 & |x| > 1 \end{cases} \qquad \Rightarrow \qquad \hat{\chi}_{[-1,1]}(x) = \sqrt{\frac{2}{\pi}} \frac{\sin x}{x} \tag{4.27}$$

Note that $\hat{\chi}$ is *not* in $\mathcal{L}^1(\mathbb{R})$ although $\chi \in \mathcal{L}^1(\mathbb{R})$.

Lemma 4.3.4: (Properties of the Fourier transform)

The Fourier transform $\hat{f}$ has the following properties:

1. It is *continuous*: $\hat{f} \in C^0(\mathbb{R}^n)$ for all $f \in \mathcal{L}^1(\mathbb{R}^n)$.
2. It is *linear*: $\widehat{\alpha f + \beta g} = \alpha \hat{f} + \beta \hat{g}$ for all $\alpha, \beta \in \mathbb{C}$, $f, g \in \mathcal{L}^1(\mathbb{R}^n)$.
3. It is *bounded*: $||\hat{f}||_\infty \leq \frac{1}{(2\pi)^{n/2}} ||f||_1$ for all $f \in \mathcal{L}^1(\mathbb{R}^n)$.
4. If $f \in \mathcal{L}^1$ vanishes almost everywhere, $\hat{f} = 0$.

Corollary 4.3.5:

The Fourier transformation defines a linear, continuous map $F : L^1(\mathbb{R}^n) \to C^0(\mathbb{R}^n)$.

Lemma 4.3.6: (Identities for the Fourier transform)

The Fourier transform satisfies the following identities

1. For all $f \in L^1(\mathbb{R}^n)$ and $g(\boldsymbol{x}) = f(\lambda \boldsymbol{x})$ almost everywhere, $\lambda \geq 0 \Rightarrow \hat{g}(\boldsymbol{y}) = \frac{1}{\lambda^n}\hat{f}(\boldsymbol{y}/\lambda)$.
2. For $f \in L^1(\mathbb{R}^n)$ and $g(\boldsymbol{x}) = \overline{f(-\boldsymbol{x})}$ almost everywhere $\Rightarrow \hat{g}(\boldsymbol{y}) = \overline{\hat{f}(\boldsymbol{y})}$.
3. For $f \in L^1(\mathbb{R}^n)$, $\hat{f}_{\boldsymbol{x}}(\boldsymbol{y}) = t_{\boldsymbol{x}} f(\boldsymbol{y}) = \hat{f}(\boldsymbol{y}) e^{-i\langle \boldsymbol{x}, \boldsymbol{y}\rangle}$
4. For $f \in \mathcal{C}^1_c(\mathbb{R}^n) = \{f \in C^1(\mathbb{R}^n) \mid \overline{\{\boldsymbol{x} \in \mathbb{R}^n \mid f(\boldsymbol{x}) \neq 0\}} \text{ compact}\}$: $\widehat{(\partial_j f)}(\boldsymbol{y}) = i y_j \hat{f}(\boldsymbol{y})$.
5. If the function $\boldsymbol{x} \mapsto x_j f$ is integrable, $\hat{f}$ is continuously differentiable with respect to y_j and $\widehat{(x_j f)} = i\partial_j \hat{f}$.
6. For $f, g \in L^1(\mathbb{R}^n)$, the functions $\hat{f} g$ and $g \hat{f}$ are integrable and
$$\int_{\mathbb{R}^n} \hat{f}(x) g(x) d^n x = \int_{\mathbb{R}^n} f(y) \hat{g}(y) d^n y. \tag{4.28}$$
7. For any $f \in \mathcal{C}^\infty_c(\mathbb{R}^n)$, the Fourier transform is integrable: $\hat{f} \in \mathcal{L}^1(\mathbb{R}^n)$.

Proof: Identities 1.), 2.), 3.), 4.) can be proved by substitution of integration variables and are left as an exercise. 7.) follows from integration theory.

To prove 5.), we need to use integration theory to show that $\hat{f}$ is differentiable and differentiation commutes with the integral in the definition of the Fourier transformation. We then obtain

$$\partial_j \hat{f}(\boldsymbol{y}) = \frac{\partial}{\partial y_j}\left(\boldsymbol{y} \mapsto \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} f(\boldsymbol{x}) e^{-i\langle \boldsymbol{x}, \boldsymbol{y}\rangle} d^n x\right) = \frac{-i}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} x^j f(\boldsymbol{x}) e^{-i\langle \boldsymbol{x}, \boldsymbol{y}\rangle} d^n x = -i\widehat{x_j f}(\boldsymbol{y}).$$

6.): As $\hat{f}, \hat{g}$ are continuous and bounded, $\hat{f} g$ and $\hat{g} f$ are integrable. Using Fubini's theorem one obtains

$$\begin{aligned}\int_{\mathbb{R}^n} f(\boldsymbol{x}) \hat{g}(\boldsymbol{x}) d^n x &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} f(\boldsymbol{x}) g(\boldsymbol{y}) e^{-i\langle \boldsymbol{x}, \boldsymbol{y}\rangle} d^n x d^n y \\ &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(\boldsymbol{x}) e^{-i\langle \boldsymbol{x}, \boldsymbol{y}\rangle} d^n x\right) g(\boldsymbol{y}) d^n y = \int_{\mathbb{R}^n} \hat{f}(\boldsymbol{y}) g(\boldsymbol{y}) d^n y\end{aligned}$$

□

As in the case of distributions, one can define a convolution on the space $L^1(\mathbb{R}^n)$. It turns out that the Fourier transform has special properties with respect to the convolution, namely that it transforms the convolution of functions to the product of convolutions. To see this, we start by defining the convolution of functions in $L^1(\mathbb{R}^n)$.

Definition 4.3.7: (Convolution)

For $f, g \in \mathcal{L}^1(\mathbb{R}^n)$, the *convolution* of f and g is the function $f * g : \mathbb{R}^n \to \mathbb{C}$ defined by

$$f * g(\boldsymbol{y}) = \int_{\mathbb{R}^n} f(\boldsymbol{x}) g(\boldsymbol{y} - \boldsymbol{x}) d^n x. \tag{4.29}$$

If g vanishes almost everywhere, $f * g = 0 = g * f$ for all $f \in \mathcal{L}^1(\mathbb{R}^n)$. The convolution therefore induces a bilinear map $* : L^1(\mathbb{R}^n) \times L^1(\mathbb{R}^n) \to L^1(\mathbb{R}^n)$.

Remark 4.3.8: Note that this definition of the convolution coincides with the one for distributions $T_f : \mathcal{C}_c^\infty(\mathbb{R}^n) \to \mathbb{C}$ associated with a functions $f \in \mathcal{L}^1(\mathbb{R}^n)$. If $T_f : \mathcal{C}_c^\infty(\mathbb{R}^n) \to \mathbb{C}$ is the distribution associated with a function $f \in \mathcal{L}^1(\mathbb{R}^n)$ and g is a function $g \in \mathcal{C}_c^\infty(\mathbb{R}^n)$, we have $T_f * g = f * g$.

Lemma 4.3.9: (Properties of the convolution)

The convolution induces a *continuous* bilinear map $* : L^1(\mathbb{R}^n) \times L^1(\mathbb{R}^n) \to L^1(\mathbb{R}^n)$. It satisfies

1. $||f * g||_1 \leq ||f||_1 ||g||_1$ for all $f, g \in L^1(\mathbb{R}^n)$.
2. $\widehat{(f * g)} = (2\pi)^{n/2} \hat{f} \cdot \hat{g}$ for all $f, g \in L^1(\mathbb{R}^n)$.
3. $f * g(\boldsymbol{x}) = g * f(\boldsymbol{x})\ \forall \boldsymbol{x} \in \mathbb{R}^n,\ \ f, g \in L^1(\mathbb{R}^n)$
4. $(f * g) * h = f * (g * h)\ \forall \boldsymbol{x} \in \mathbb{R}^n,\ f, g, h \in L^1(\mathbb{R}^n)$.

Proof: Exercise. □

To work with a notion of Fourier transform that is defined for locally integrable functions or as a continuous linear map from $L^1(\mathbb{R}^n)$ to $C^0(\mathbb{R}^n)$ will not be sufficient in the following. Firstly, we would like to construct the *inverse* of a Fourier transform, which is complicated if we view the Fourier transform as a continuous linear map $F : L^1(\mathbb{R}^n) \to C^0(\mathbb{R}^n)$. Secondly, many applications of the Fourier transform such as its role in quantum mechanics involve functions in $L^2(\mathbb{R}^n)$, not $L^1(\mathbb{R}^n)$. We will see in the following that it is possible to extend the Fourier transform to $L^2(\mathbb{R}^n)$ and that this is closely related to constructing its inverse.

As a first step, we note that the Fourier transform is *injective* in the following sense:

Theorem 4.3.10: (Uniqueness and Inversion of Fourier transform)

1. If $f \in L^1(\mathbb{R}^n)$ such that $\hat{f} \in L^1(\mathbb{R}^n)$, we have

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \hat{f}(\boldsymbol{y}) e^{i\langle \boldsymbol{y}, \boldsymbol{x} \rangle} \qquad \text{almost everywhere.} \tag{4.30}$$

2. For all $f, g \in L^1(\mathbb{R}^n)$ with $\hat{f} = \hat{g}$, we have $f(\boldsymbol{x}) = g(\boldsymbol{x})$ for almost all $\boldsymbol{x} \in \mathbb{R}^n$.

Hence, the Fourier transformation is injective in the sense that two functions in $L^1(\mathbb{R}^n)$ whose Fourier transformations agree must be equal almost everywhere. However, the inversion formula given in the preceding theorem is only valid for functions that are such that their Fourier transform is also in $L^1(\mathbb{R}^n)$. As the Fourier transform does *not* preserve the L^1- norm $||\ \ ||_1$, there is in general no guarantee that the Fourier transformation of a function in $L^1(\mathbb{R}^n)$ is in $L^1(\mathbb{R}^n)$. However, it turns out that the Fourier transform preserves the L^2-norm $||\ \ ||_2$.

Theorem 4.3.11: For all $f \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$, $\hat{f} \in L^2(\mathbb{R}^n)$ and $||f||_2 = ||\hat{f}||_2$.

Hence, the Fourier transformation is *defined* for functions in $L^1(\mathbb{R}^n)$ and is injective for those functions, but preserves the L^2-norm. This suggests that in order to obtain a Fourier transformation which is an isomorphism of Banach or Hilbert spaces, we should extend it to $L^2(\mathbb{R}^n)$. That this is possible is shown by Plancherel's theorem, whose proof makes use of the fact that the smooth functions with compact support and hence the intersection $L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ are dense in $L^2(\mathbb{R}^n)$.

Corollary 4.3.12: (Plancherel's theorem)

There exists a unique isomorphism $\mathcal{F} : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ such that

1. $||\mathcal{F}(f)||_{L_2} = ||f||_{L_2}$ for all $f \in L_2(\mathbb{R}^n)$
2. $\mathcal{F}(f) = \hat{f}$ for all $f \in L_1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$
3. $\mathcal{F}^{-1}(f)(\boldsymbol{y}) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} f(\boldsymbol{x}) e^{i\langle \boldsymbol{x}, \boldsymbol{y} \rangle} d^n x$ for all $f \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$.

4.4 Hilbert spaces

We will now focus on Hilbert spaces, which play an important role in quantum mechanics. Throughout this section, we denote by $\mathcal{H}$ a Hilbert space with scalar product or hermitian product $\langle\, ,\, \rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$.

Many of the constructions that are possible for finite-dimensional vector spaces such as projections on linear subspaces and decomposition of vectors with respect to a basis are also possible for (separable) Hilbert spaces. We start by considering the projection on a closed linear subspace.

Theorem 4.4.1: (Orthogonal projection on a closed subspace)

Let $\mathcal{H}$ be a Hilbert space and $F \subset \mathcal{H}$ a closed subspace of $\mathcal{H}$, i.e. a linear subspace $F \subset \mathcal{H}$ such that any Cauchy sequence with elements in F converges towards an element of F. Then for each $\boldsymbol{x} \in \mathcal{H}$ there exists a unique $\boldsymbol{y} \in F$, the *projection of* $\boldsymbol{x}$ *on* F denoted $\boldsymbol{y} = P_F(\boldsymbol{x})$, such that $||\boldsymbol{x} - \boldsymbol{y}|| = d(\boldsymbol{x}, F) = \inf\{||\boldsymbol{x} - \boldsymbol{z}|| \mid \boldsymbol{z} \in F\}$. The projection map $P_F : \mathcal{H} \to F$, $\boldsymbol{x} \mapsto P_F(\boldsymbol{x})$ has the following properties:

1. It is linear and continuous.
2. It satisfies: $P_F \circ P_F = P_F$
3. $\langle \boldsymbol{x} - P_F(\boldsymbol{x}), \boldsymbol{z} \rangle = 0$ for all $\boldsymbol{z} \in F$.
4. Any element $\boldsymbol{x} \in H$ can be expressed uniquely as a sum of its projection on F and an element of the *orthogonal complement* $F^\perp = \{\boldsymbol{y} \in H \mid \langle \boldsymbol{z}, \boldsymbol{y} \rangle = 0\ \forall \boldsymbol{z} \in F\}$.

$$\boldsymbol{x} = P_F(\boldsymbol{x}) + \boldsymbol{z} \qquad \boldsymbol{z} \in F^\perp. \tag{4.31}$$

Proof:

1. Uniqueness: To prove that $\boldsymbol{y}$ is unique, we select a sequence $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ with $||\boldsymbol{x}_n - \boldsymbol{x}|| \to d(\boldsymbol{x}, F) = \inf\{||\boldsymbol{x} - \boldsymbol{z}|| \mid \boldsymbol{z} \in F\}$. The last identity in exercise (35) implies

$$||\boldsymbol{x}_n - \boldsymbol{x}_k||^2 = 2(||\boldsymbol{x} - \boldsymbol{x}_n||^2 + ||\boldsymbol{x} - \boldsymbol{x}_k||^2) - 4\underbrace{||\boldsymbol{x}\tfrac{1}{2}(\boldsymbol{x}_n + \boldsymbol{x}_m)||^2}_{\geq d(\boldsymbol{x},F)^2} \tag{4.32}$$

and therefore $||\boldsymbol{x}_n - \boldsymbol{x}_k||^2 < \epsilon^2$ if $||\boldsymbol{x} - \boldsymbol{x}_k||^2, ||\boldsymbol{x} - \boldsymbol{x}_n||^2 < d(\boldsymbol{x}, F)^2 + \frac{\epsilon^2}{4}$. This shows that $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence. As F is closed, this sequence converges to an element $\boldsymbol{y} \in F$ with $||\boldsymbol{x} - \boldsymbol{y}|| = d(\boldsymbol{x}, F)$. For any element $\boldsymbol{z} \in F$ with $||\boldsymbol{x} - \boldsymbol{z}|| = d(\boldsymbol{x}, F)$, we obtain using again the last identity in exercise (35)

$$||\boldsymbol{y} - \boldsymbol{z}||^2 = 4d(\boldsymbol{x}, F)^2 - 4\underbrace{||\boldsymbol{x} - \tfrac{1}{2}(\boldsymbol{y} + \boldsymbol{z})||}_{\geq d(\boldsymbol{x},F)^2} \leq 0 \tag{4.33}$$

and therefore $\boldsymbol{z} = \boldsymbol{y}$. This shows that $\boldsymbol{y} = P_F(\boldsymbol{x})$ is unique.

2. Let $\boldsymbol{y} = P_F(\boldsymbol{x})$ for a general $\boldsymbol{x} \in \mathcal{H}$. Then, $\boldsymbol{y} \in F$ and $d(\boldsymbol{y}, F) = 0 = ||\boldsymbol{y} - \boldsymbol{y}||$. The uniqueness of the projection then implies $P_F(\boldsymbol{y}) = \boldsymbol{y}$ and therefore $P_F \circ P_F = P_F$.

3. To prove that $\langle \boldsymbol{x} - P_F(\boldsymbol{x}), \boldsymbol{z}\rangle = 0$ for all $\boldsymbol{z} \in F$, we use the third identity in (35) and find for all $t \in \mathbb{R}^+$

$$||\boldsymbol{x} - (P_F(\boldsymbol{x}) - t\boldsymbol{z})||^2 = t^2||\boldsymbol{z}||^2 + d(\boldsymbol{x}, F)^2 - 2t\langle \boldsymbol{x} - P_F(\boldsymbol{x}), \boldsymbol{z}\rangle \geq d(\boldsymbol{x}, F)^2, \tag{4.34}$$

which implies $t(\langle \boldsymbol{x} - P_F(\boldsymbol{x}), \boldsymbol{z}\rangle - \frac{t}{2}||\boldsymbol{z}||^2) \leq 0$ for all $t \in \mathbb{R}^+$. Hence, we must have $\langle \boldsymbol{x} - P_F(\boldsymbol{x}), \boldsymbol{z}\rangle - \frac{t}{2}||\boldsymbol{z}||^2 \leq 0$ for all $t \in \mathbb{R}^+$. By letting t tend to zero, we obtain $\langle \boldsymbol{x} - P_F(\boldsymbol{x}), \boldsymbol{z}\rangle = 0$. Setting $\boldsymbol{x} = P_F(\boldsymbol{x}) + (x - P_F(\boldsymbol{x}))$ for any $\boldsymbol{x} \in \mathcal{H}$ therefore defines a unique decomposition $\boldsymbol{x} = P_F(\boldsymbol{x}) + \boldsymbol{z}$ with $\boldsymbol{z} \in F^\perp$.

4. To show linearity of $P_F : \mathcal{H} \to F$, we use

$$\langle \boldsymbol{x} + \boldsymbol{y} - P_F(\boldsymbol{x} + \boldsymbol{y}), \boldsymbol{z}\rangle = \langle \boldsymbol{x} + \boldsymbol{y} - P_F(\boldsymbol{x}) - P_F(\boldsymbol{y}), \boldsymbol{z}\rangle = 0 \qquad \forall \boldsymbol{z} \in F. \tag{4.35}$$

This implies

$$\langle P_F(\boldsymbol{x} + \boldsymbol{y}) - P_F(\boldsymbol{x}) - P_F(\boldsymbol{y}), P_F(\boldsymbol{x} + \boldsymbol{y}) - P_F(\boldsymbol{x}) - P_F(\boldsymbol{y})\rangle = ||P_F(\boldsymbol{x} + \boldsymbol{y}) - P_F(\boldsymbol{x}) - P_F(\boldsymbol{y})||^2 = 0$$

and therefore $P_F(\boldsymbol{x}+\boldsymbol{y}) = P_F(\boldsymbol{x}) + P_F(\boldsymbol{y})$. To show that P_F is continuous, we take a sequence $(\boldsymbol{y}_n)_{n\in\mathbb{N}}$ with $||\boldsymbol{y}_n - \boldsymbol{y}|| \to 0$. The decomposition $\boldsymbol{y} = P_F(\boldsymbol{y}) + \boldsymbol{y} - P_F(\boldsymbol{y})$ with $\boldsymbol{y} - P_F(\boldsymbol{y}) \in F^\perp$ implies

$$||\boldsymbol{y} - \boldsymbol{y}_n||^2 = ||P_F(\boldsymbol{y}) - P_F(\boldsymbol{y}_n)||^2 + ||\boldsymbol{y} - P_F(\boldsymbol{y}) - \boldsymbol{y}_n + P_F(\boldsymbol{y}_n)||^2 \geq ||P_F(\boldsymbol{y}) - P_F(\boldsymbol{y}_n)||^2$$

$||\boldsymbol{y} - \boldsymbol{y}_n|| \to 0$ therefore implies $||P_F(\boldsymbol{y}) - P_F(\boldsymbol{y}_n)|| \to 0$ and hence continuity of P_F. □.

Corollary 4.4.2:

For any $A \subset \mathcal{H}$ $\mathrm{Span}(A) = \{\sum_{i=1}^n \lambda_i \boldsymbol{a}_i \mid \lambda_i \in k, \boldsymbol{a}_i \in A\}$ is dense in $\mathcal{H}$ if and only if the *orthogonal complement* $A^\perp$ is trivial $A^\perp = \{\boldsymbol{x} \in H \mid \langle \boldsymbol{x}, \boldsymbol{a}\rangle = 0\ \forall \boldsymbol{a} \in A\} = \{0\}$.

Proof: Clearly, $\overline{\mathrm{Span}(A)}$ is a closed subspace of $\mathcal{H}$ and $A^\perp = \overline{\mathrm{Span}(A)}^\perp$. By theorem 4.4.1, any element $\boldsymbol{x} \in \mathcal{H}$ can be decomposed uniquely as $\boldsymbol{x} = P_{\overline{\mathrm{Span}(A)}}(\boldsymbol{x}) + \boldsymbol{z}$ with $\boldsymbol{z} \in A^\perp$. If $A^\perp$ is trivial, this implies $\boldsymbol{x} \in \overline{\mathrm{Span}(A)}$ and hence density of $\mathrm{Span}(A)$ in $\mathcal{H}$.

Conversely, if $\mathrm{Span}(A)$ is dense in $\mathcal{H}$, then for every element $\boldsymbol{x} \in A^\perp$ there exists a sequence $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$, $\boldsymbol{x}_n \in \mathrm{Span}(A)$ with $\boldsymbol{x}_n \to \boldsymbol{x}$. As we have $\langle \boldsymbol{x}_n, \boldsymbol{x}\rangle = 0$ for all $n \in \mathbb{N}$, this implies $\langle \boldsymbol{x}, \boldsymbol{x}\rangle = 0$. We have $\boldsymbol{x} = 0$ and therefore $A^\perp = \{0\}$. □

Definition 4.4.3: (Separable Hilbert space)

A Hilbert space $\mathcal{H}$ is called *separable* if there exists a countably infinite set $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ with $\boldsymbol{x}_n \in H$ for all $n \in \mathbb{N}$ and whose Span is dense in $\mathcal{H}$: $\overline{\text{Span}(\{\boldsymbol{x}_n \mid n \in \mathbb{N}\})} = \mathcal{H}$. In other words, for every $\boldsymbol{y} \in H$ and every $\epsilon > 0$ there exists a finite set of indices $n_1, ..., n_k \in \mathbb{N}$ and coefficients $\alpha_1, ..., \alpha_k \in \mathbb{C}$ such that

$$||\sum_{i=1}^{k} \alpha_i \boldsymbol{x}_{n_i} - \boldsymbol{y}|| < \epsilon. \tag{4.36}$$

Example 4.4.4: (Separable Hilbert spaces)

1. The Hilbert space l^2 of square summable series is separable.
2. $L^2(\mathbb{R}^n)$ with the scalar product (4.8) is separable.

Example 4.4.5: (Non-separable Hilbert space)

The vector space

$$H = \{f : \mathbb{R} \to \mathbb{R} \mid \exists c \in \mathbb{R}^+ : \sum_{t\in I} |f(t)|^2 < c \text{ for all finite } I \subset \mathbb{R}\} \qquad \langle f, g\rangle = \sum_{t\in\mathbb{R}} f(t)g(t)$$

is a Hilbert space but not separable.

Proof:

1. For all $f \in H$, the set $I(f) = \{t \in \mathbb{R} \mid f(t) \neq 0\}$ is countable. To demonstrate this, we consider the sets $I_n(f) = \{t \in \mathbb{R} : \mid |f(t)| \geq \frac{1}{n}\}$, $n \in \mathbb{N}$. I_n must be finite, because otherwise one could construct a sequence $\{t_i\}_{i\in\mathbb{N}}$, $t_i \in \mathbb{N}$ with $\sum_{i=1}^{k} |f(t_i)|^2 \geq \frac{k^2}{n} \to \infty$ with $k \to \infty$. Clearly, $I(f) = \bigcup_{n\in\mathbb{N}} I_n(f)$ and therefore $I(f)$ is countable.

2. Suppose $\text{Span}(A) \subset H$ is dense in H. As the functions $g_s : \mathbb{R} \to \mathbb{R}$

$$g_s(t) = \begin{cases} 1 & t = s \\ 0 & t \neq s \end{cases}$$

are elements of H for all $s \in \mathbb{R}$, for all $s \in \mathbb{R}$ there must exist a sequence $\{f_n^s\}_{n\in\mathbb{N}}$, $f_n \in \text{Span}(A)$ with $||f_n^s - g^s|| \to 0$. This implies $|f_n^s(t) - g_s(t)| \to 0$ for all $t \in \mathbb{N}$ and hence $f_n^s(s) \neq 0$ for $n \geq n_0$ for some $n_0 \in \mathbb{N}$.

Hence, for any $s \in \mathbb{R}$, A must contain a function $f_s \in H$ with $f_s(s) \neq 0$. This implies that A cannot be a countable infinite set, since for every $f \in H$, $I(f)$ is countable and a countable unit of countable sets must be countable, whereas $\mathbb{R}$ is not countable. □

Definition 4.4.6: (Hilbert basis)

A *Hilbert basis* of a Hilbert space $\mathcal{H}$ is a finite or countably infinite sequence $(e_i)_{i\in\mathbb{N}}$, $e_i \in \mathcal{H}$, of elements in $\mathcal{H}$ that are

1. *orthonormal*: $\langle e_i, e_j\rangle = \delta_{ij}$ for all $i, j \in \mathbb{N}$
2. *dense in* $\mathcal{H}$: $\overline{\text{Span}(\{e_i \mid i \in \mathbb{N}\})} = \mathcal{H}$.

Lemma 4.4.7: A Hilbert space admits a Hilbert basis if and only if it is separable.

Proof: Clearly, if $\mathcal{H}$ admits a Hilbert basis, it is separable, and if $H = \{0\}$ the statement is trivial. We need to prove that if $H \neq \{0\}$ is separable, it admits a Hilbert basis. This is the infinite-dimensional counterpart of Gram Schmidt orthogonalisation.

If $\mathcal{H}$ is separable, there exists a dense set of elements $B = \{\boldsymbol{x}_n \mid n \in \mathbb{N}\}$. Let i_1 be $i_1 = \min\{n \in \mathbb{N} \mid \boldsymbol{x}_n \neq 0\; \boldsymbol{x}_k = 0 \forall k < n\}$ and set $e_1 = \boldsymbol{x}_{i_1}/||\boldsymbol{x}_{i_1}||$. If $\boldsymbol{x}_k \in \text{Span}(\{e_1\})$ for all $k \geq i_1$, we have $H = \text{Span}(\{e_1\})$ and the proof is complete. Otherwise, there exists an index $i_2 \in \mathbb{N}$, $i_2 > i_1$ such that $\boldsymbol{x}_{i_2} \notin \text{Span}(\{e_1\})$ and $\boldsymbol{x}_k \in \text{Span}\{\{e_1\}\}$ for all $k \in [i_1, i_2[$. Set

$$e_2 = \frac{\boldsymbol{x}_{i_2} - \langle \boldsymbol{x}_{i_2}, e_1\rangle e_1}{||\boldsymbol{x}_{i_2} - \langle \boldsymbol{x}_{i_2}, e_1\rangle e_1||} \tag{4.37}$$

If $\boldsymbol{x}_k \in \text{Span}(e_1, e_2)$ for all $k \geq i_2$, we have $H = \text{Span}(\{e_1, e_2\})$ and the proof is complete. Otherwise, there exists an index $i_3 \in \mathbb{N}$, $i_3 > i_2$ such that $\boldsymbol{x}_{i_3} \notin \text{Span}(\{e_1, e_2\})$ and $\boldsymbol{x}_k \in \text{Span}\{e_1, e_2\}$ for all $k \in [i_2, i_3[$. Set

$$e_3 = \frac{\boldsymbol{x}_{i_3} - \langle \boldsymbol{x}_{i_3}, e_1\rangle e_1 - \langle \boldsymbol{x}_{i_3}, e_2\rangle e_2}{||\boldsymbol{x}_{i_3} - \langle \boldsymbol{x}_{i_2}, e_1\rangle e_1 - \langle \boldsymbol{x}_{i_3}, e_2\rangle e_2||} \tag{4.38}$$

and continue in this way.

If $\mathcal{H}$ is finite dimensional, the procedure stops after a finite number of steps. Otherwise, we obtain a sequence $(e_i)_{i\in\mathbb{N}}$ with $\langle e_i, e_j\rangle = \delta_{ij}$. In this case, we need to show that $\overline{\text{Span}(\{e_i \mid i \in \mathbb{N}\})} = H$. This follows with Corollary 4.4.2 if we can show that $\text{Span}(\{e_i \mid i \in \mathbb{N}\})^\perp = \{0\}$. Consider an element $\boldsymbol{x}$ with $\langle \boldsymbol{x}, e\rangle_i = 0$ for all $i \in \mathbb{N}$. By construction of the elements e_i, this implies $\langle \boldsymbol{x}, \boldsymbol{x}\rangle_n = 0$ for all $n \in \mathbb{N}$. As $\text{Span}(\{\boldsymbol{x}_n \mid n \in \mathbb{N}\})$ is dense in $\mathcal{H}$, this implies $\boldsymbol{x} = 0$. □

Theorem 4.4.8: (Decomposition with respect to a Hilbert basis)

Let $\mathcal{H}$ be a separable Hilbert space with a Hilbert basis $B = \{e_i \mid i \in \mathbb{N}\}$. For any vector $\boldsymbol{x} \in \mathcal{H}$ we have the identities

$$\boldsymbol{x} = \sum_{i=1}^{\infty} \langle e_i, \boldsymbol{x}\rangle e_i \tag{4.39}$$

$$||\boldsymbol{x}||^2 = \langle \boldsymbol{x}, \boldsymbol{x}\rangle = \sum_{i=1}^{\infty} |\langle e_i, \boldsymbol{x}\rangle|^2 \qquad \text{(Parseval's equation)}. \tag{4.40}$$

In other words: The series $S_n(\boldsymbol{x}) = \sum_{i=1}^n \langle e_i, \boldsymbol{x}\rangle e_i$ and $R_n(\boldsymbol{x}) = \sum_{i=1}^n |\langle e_i, \boldsymbol{x}\rangle|^2$ converge absolutely towards $\boldsymbol{x}$ and $||\boldsymbol{x}||^2$. If $(a_i)_{i\in\mathbb{N}}$ is a sequence such that $\sum_{i=1}^\infty |a_i|^2 < \infty$, then $\boldsymbol{a} = \sum_{i=1}^\infty a_i e_i$ exists (i.e. the series converges) and $a_i = \langle e_i, \boldsymbol{a}\rangle$.

Proof:

As we can write $\boldsymbol{x} = \boldsymbol{x} - S_n(\boldsymbol{x}) + S_n(\boldsymbol{x})$ with $\langle \boldsymbol{x} - S_n(\boldsymbol{x}), e_k\rangle = 0$ for all $k \leq n$, we have

$$||\boldsymbol{x}||^2 = ||\boldsymbol{x} - S_n(\boldsymbol{x})||^2 + ||S_n(\boldsymbol{x})||^2 \quad \Rightarrow \quad ||S_n(\boldsymbol{x})||^2 = \sum_{k=1}^n |\langle \boldsymbol{x}, e_k\rangle| \leq ||\boldsymbol{x}||^2. \tag{4.41}$$

This implies that for all $\epsilon > 0$ there exists an $n_\epsilon \in \mathbb{N}$ such that

$$R_m(\boldsymbol{x}) - R_n(\boldsymbol{x}) = ||S_n(\boldsymbol{x}) - S_m(\boldsymbol{x})||^2 = ||\sum_{k=n+1}^m \langle e_k, \boldsymbol{x}\rangle e_k||^2 = \sum_{k=n+1}^m |\langle e_k, \boldsymbol{x}\rangle|^2 < \epsilon \;\forall n, m \geq n_\epsilon.$$

Hence, the sequence $(S_n(\boldsymbol{x}))_{n\in\mathbb{N}}$ is a Cauchy sequence and converges towards a unique element $\boldsymbol{y} \in \mathcal{H}$. For any $\epsilon > 0$ there exists $n \in \mathbb{N}$ such that $||\boldsymbol{y} - S_n(\boldsymbol{x})|| < \epsilon$. We then have

$$|\langle \boldsymbol{y} - \boldsymbol{x}, e_k\rangle| \leq |\langle \boldsymbol{y} - S_n(\boldsymbol{x}), e_k\rangle| + |\langle \boldsymbol{x} - S_n(\boldsymbol{x}), e_k\rangle| \leq ||\boldsymbol{y} - S_n(\boldsymbol{x})|| < \epsilon \qquad \forall k \leq n. \quad (4.42)$$

With $\epsilon \to 0$, $n \to \infty$, we therefore obtain $|\langle \boldsymbol{y} - \boldsymbol{x}, e_k\rangle| = 0$ for all $k \in \mathbb{N}$. This implies $\boldsymbol{y} - \boldsymbol{x} \in \text{Span}(B)^\perp = \overline{\text{Span}(B)}^\perp$. As B is dense in $\mathcal{H}$, Corollary 4.4.2 implies $\boldsymbol{x} = \boldsymbol{y}$ and

$$||\boldsymbol{x}||^2 = \lim_{n\to\infty} ||S_n(\boldsymbol{x})||^2 = \sum_{k=1}^{\infty} |\langle \boldsymbol{x}, e_i\rangle|^2. \quad (4.43)$$

Conversely, if $(a_i)_{i\in\mathbb{N}}$ is a sequence such that $\sum_{i=1}^{\infty} |a_i|^2 < \infty$, then for each $\epsilon > 0$ there exists n_ϵ such that

$$|| \sum_{k=n+1}^{m} a_k e_k||^2 = \sum_{k=n+1}^{m} |a_k|^2 < \epsilon \qquad \forall n, m > n_\epsilon. \quad (4.44)$$

This implies that $A_n = \sum_{k=0}^{n} a_k e_k$ is a Cauchy sequence and converges towards a unique element $\boldsymbol{a} \in \mathcal{H}$. We have

$$|\langle \boldsymbol{a}, e_i\rangle - a_i| \leq |\langle \boldsymbol{a} - A_n, e_i\rangle| + |\langle A_n, e_i\rangle - a_i| \leq ||\boldsymbol{a} - A_n|| \xrightarrow{n\to\infty} 0 \quad (4.45)$$

with $n \to \infty$, we therefore obtain $\langle \boldsymbol{a}, e_i\rangle = a_i$. □

Corollary 4.4.9: For any separable Hilbert space $\mathcal{H}$ with scalar product $\langle\ ,\ \rangle$, there is a unitary isomorphism $\phi : H \to l^2$, i.e. an isomorphism $\phi : H \to l^2$ which preserves the scalar product $\langle \phi(\boldsymbol{x}), \phi(\boldsymbol{y})\rangle_{l^2} = \langle \boldsymbol{x}, \boldsymbol{y}\rangle$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{H}$.

Proof: This is simply a rephrasing of Theorem 4.4.8. To obtain such a unitary isomorphism $\phi : H \to l^2$, one chooses a Hilbert basis $B = \{e_i \mid i \in \mathbb{N}\}$ of $\mathcal{H}$ and sets $\phi(\boldsymbol{x}) = \{\langle \boldsymbol{x}, e_i\rangle\}_{i\in\mathbb{N}}$.□

Remark 4.4.10: A Hilbert basis is the infinite-dimensional analogue of an orthonormal basis for finite-dimensional vector spaces. The Hilbert space of absolutely convergent series in Example 4.1.28 is the infinite-dimensional analogue of the vector spaces $\mathbb{R}^n$ and $\mathbb{C}^n$. Theorem 4.4.8 can therefore be viewed as the infinite-dimensional analogue of the identification of finite-dimensional vector spaces with $\mathbb{C}^n$ or $\mathbb{R}^n$.

As in the finite dimensional case, this identification is *not canonical* because it requires the choice of a basis. Moreover, in the infinite dimensional case, the vector space structure alone is not enough for the identification. Instead, one needs the stricter requirement that the vector space is a *separable Hilbert space*.

Chapter 5

Quantum mechanics

5.1 Operators on Hilbert spaces

Definition 5.1.1: (Bounded operator, compact operator)

A linear operator $O : B \to B$ on a Banach space $\mathcal{B}$ is called *bounded* if there exists a $c \in \mathbb{R}^+$ such that $||O\boldsymbol{v}|| \leq c||\boldsymbol{v}|| \; \forall \boldsymbol{v} \in \mathcal{B}$. It is called *compact* if $\overline{O(\mathcal{U})}$ is compact for every bounded $\mathcal{U} \subset \mathcal{B}$.

Lemma 5.1.2: A linear operator O on a Banach space $\mathcal{B}$ is bounded if and only if it is *continuous*, i. e. for all sequences $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$, $\boldsymbol{x}_n \in \mathcal{B}$ with $||\boldsymbol{x}_n - \boldsymbol{x}|| \to 0$ we have $||O\boldsymbol{x}_n - O\boldsymbol{x}|| \to 0$. The vector space of bounded linear operators on a Hilbert space $\mathcal{B}$ with norm $||O|| = \sup\{||O\boldsymbol{x}|| \mid \boldsymbol{x} \in E, ||\boldsymbol{x}|| = 1\}$ is a Banach space. It will be denoted $B(\mathcal{B})$ in the following.

Proof: Let $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$, $\boldsymbol{x}_n \in \mathcal{B}$ be a sequence with $||\boldsymbol{x}_n - \boldsymbol{x}|| \xrightarrow{n\to\infty} 0$, $\boldsymbol{x} \in \mathcal{H}$. If O is bounded, we have $||O\boldsymbol{x}_n - O\boldsymbol{x}|| \leq c||\boldsymbol{x}_n - \boldsymbol{x}|| \xrightarrow{n\to\infty} 0$, hence O is continuous. If O is continuous, it is continuous in 0 and there exists a $\delta > 0$ such that $||O\boldsymbol{x}|| < 1$ for all $\boldsymbol{x} \in \mathcal{B}$ with $||\boldsymbol{x}|| < \delta$. This implies

$$||O\boldsymbol{y}|| = \left|\left|\frac{||\boldsymbol{y}||}{\delta}\, O\left(\frac{\delta \boldsymbol{y}}{||\boldsymbol{y}||}\right)\right|\right| = \frac{||\boldsymbol{y}||}{\delta}\left|\left|O\left(\frac{\delta \boldsymbol{y}}{||\boldsymbol{y}||}\right)\right|\right| \leq \frac{||\boldsymbol{y}||}{\delta} \qquad \forall \boldsymbol{y} \in \mathcal{H}.$$

□

Example 5.1.3: (Bounded operators)

1. The identity operator on a Banach space $\mathcal{B}$ is bounded. It is compact if and only if $\mathcal{B}$ is finite-dimensional.
2. If $\mathcal{H}$ is a Hilbert space, projectors on a closed subspace $F \subset \mathcal{H}$ are bounded.

Many of the linear operators which we will consider in the following will be bounded operators defined on a Banach space $\mathcal{B}$ or a Hilbert space $\mathcal{H}$. However, there are operators which are important in physics and that are *unbounded* or not even defined on everywhere on $\mathcal{B}$ or $\mathcal{H}$. Examples are the momentum and position operators and the Hamiltonian operators of systems whose energy is not bounded.

Example 5.1.4: (Unbounded operators)

1. For $\mathcal{H} = L^2(\mathbb{R}^n)$, the *position operator* $X_i(\psi)(\boldsymbol{x}) = x_i \cdot \psi(\boldsymbol{x})$ and the momentum operator $P_i\psi(\boldsymbol{x}) = i\frac{\partial}{\partial x_i}\psi(\boldsymbol{x})$ are *not* bounded, and there exist elements $\psi \in L^2(\mathbb{R}^n)$ for which $X\psi$, $P\psi \notin L^2(\mathbb{R}^n)$.

2. Let $\alpha = (j_1, ..., j_n) \in \mathbb{N}^n$ a multi-index, $D^\alpha = \partial_{j_1} \cdots \partial_{j_n}$ and $D = \sum_{|\alpha|\leq k} c_\alpha D^\alpha$, $c_\alpha \in \mathcal{C}^\infty(\mathbb{R}^n)$ a linear differential operator on the Hilbert space $\mathcal{H} = L^2(\mathbb{R}^n)$. Then D is in general not bounded and there exist $\psi \in L^2(\mathbb{R}^n)$ for which $D\psi \notin L^2(\mathbb{R}^n)$.

In both of the examples above, one finds that the unbounded operators become bounded and map elements of the Hilbert space $L^2(\mathbb{R}^n)$ to elements of $L^2(\mathbb{R}^n)$ if one restricts attention to the linear subspace $\mathcal{C}_c^\infty(\mathbb{R}^n)$ of smooth functions with compact support. However, note that this space is *not* a Banach space. It is dense in $L^2(\mathbb{R}^n)$ but *not* normed, since it is not metrisable. One therefore has to give up the restriction to bounded linear operators defined on the whole Hilbert or Banach space and to consider also operators which are only defined on linear subspaces of a Banach space $\mathcal{B}$ or Hilbert space $\mathcal{H}$. This leads to the notion of domain and range.

Definition 5.1.5: (Domain, range, densely defined)

Let A be a linear operator acting on a Banach space $\mathcal{B}$. The *domain* $\mathfrak{d}_A \subset \mathcal{B}$ of A is the linear subspace spanned by elements $\boldsymbol{x} \in \mathcal{B}$ on which A is defined as a continuous linear map $A : \mathfrak{d}_A \to \mathcal{B}$. The *range* $\mathfrak{r}_A \subset \mathcal{B}$ is the set $\mathfrak{r}_A = \{\boldsymbol{x} \in \mathcal{B} \mid \exists \boldsymbol{y} \in \mathfrak{d}_A : \boldsymbol{x} = A\boldsymbol{y}\}$. An operator A is called *densely defined* if its domain $\mathfrak{d}_A$ is dense in $\mathcal{B}$: $\overline{\mathfrak{d}_A} = \mathcal{B}$.

Example 5.1.6: (Position and momentum operator)

We consider the Hilbert space $\mathcal{H} = L^2(\mathbb{R})$. Let $X : \psi \mapsto X\psi$, $X\psi(x) = x\psi(x)$ the position operator and $P : \psi \mapsto P\psi$, $P\psi(x) = i\frac{\partial}{\partial x}\psi(x)$ the momentum operator on $L^2(\mathbb{R})$. Then the domains $\mathfrak{d}_X$, $\mathfrak{d}_P$ are the linear subspaces of $L^2(\mathbb{R}^n)$ defined by the conditions

$$\mathfrak{d}_X = \{\psi \in L^2(\mathbb{R}) \mid \int_{-\infty}^{\infty} x^2|\psi(x)|^2 dx < \infty\} \tag{5.1}$$
$$\mathfrak{d}_P = \{\psi \in L^2(\mathbb{R}) \mid \int_{-\infty}^{\infty} \left|\frac{\partial}{\partial x}\psi(x)\right|^2 dx < \infty\}.$$

Clearly, $\mathfrak{d}_X, \mathfrak{d}_P \neq \mathcal{H}$ since for $\psi, \phi \in L^2(\mathbb{R})$

$$\psi(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{\sqrt{x}} & x \geq 1 \end{cases} \qquad \phi(x) = \begin{cases} 0 & x \in]-\infty, 0] \cup [1, \infty[\\ \frac{1}{x^{1/4}} & x \in]0, 1[\end{cases}$$

$$\int_{-\infty}^{\infty} x^2|\psi(x)|^2 = \int_1^\infty x dx = \infty \qquad \int_{-\infty}^{\infty} \left|\frac{\partial}{\partial x}\phi(x)\right|^2 dx = \int_0^1 \frac{1}{16x^{5/2}} dx = \infty.$$

However, the operators X, P are densely defined. We have

$$\int_{-\infty}^{\infty} x^2|\varphi(x)|^2 dx < \infty \qquad \int_{-\infty}^{\infty} \left|\frac{\partial}{\partial x}\varphi(x)\right|^2 dx < \infty \qquad \forall \varphi \in \mathcal{C}_c^\infty(\mathbb{R}).$$

As $\mathcal{C}_c^\infty(\mathbb{R})$ is dense in $L^2(\mathbb{R})$, this implies $\mathfrak{d}_X, \mathfrak{d}_P$ dense in $\mathcal{H} = L^2(\mathbb{R})$.

Remark 5.1.7: The domain of an operator is not a mathematical subtlety but an *essential part* of the definition of a linear operator and has a physical meaning. One can show - for instance by considering the momentum operator on the interval $[0, 1]$ - that there are several possibilities of *extending* a given operator to larger domains and these different extensions differ in their spectrum. The domain is also important when one discusses the question if two operators on a Hilbert space are equal: Two operators $A : \mathfrak{d}_A \to \mathfrak{r}_A$, $B : \mathfrak{d}_B \to \mathfrak{r}_B$ are *equal* $A = B$ if they satisfy $\mathfrak{d}_A = \mathfrak{d}_B$ and $A = B$ on $\mathfrak{d}_A = \mathfrak{d}_B$.

The role of domain of an operator gives rise to another important distinction, namely that of a self-adjoint operator on a Hilbert space from an hermitian operator.

Definition 5.1.8: (Adjoint, hermitian, self-adjoint operators)

Let $A : \mathfrak{d}_A \to \mathfrak{r}_A$ be a linear operator on a Hilbert space $\mathcal{H}$ whose *domain* $\mathfrak{d}_A \subset \mathcal{H}$ is dense in $\mathcal{H}$. The *adjoint* of the operator A is the operator $A^\dagger : \mathfrak{d}_{A^\dagger} \to \mathfrak{r}_{A^\dagger}$ defined by

$$\langle \boldsymbol{x}, A\boldsymbol{y}\rangle = \langle A^\dagger \boldsymbol{x}, \boldsymbol{y}\rangle \qquad \forall \boldsymbol{y} \in \mathfrak{d}_A, \boldsymbol{x} \in \mathfrak{d}_{A^\dagger} \tag{5.2}$$

$$\text{where} \quad \mathfrak{d}_{A^\dagger} = \{\boldsymbol{x} \in \mathcal{H} \mid \exists \boldsymbol{z} \in \mathcal{H} : \langle \boldsymbol{x}, A\boldsymbol{y}\rangle = \langle \boldsymbol{z}, \boldsymbol{y}\rangle \ \forall \boldsymbol{y} \in \mathfrak{d}_A\}.$$

A linear operator A on $\mathcal{H}$ is called *hermitian* if $\mathfrak{d}_A \subset \mathfrak{d}_{A^\dagger}$ and $A\boldsymbol{x} = A^\dagger \boldsymbol{x}$ for all $\boldsymbol{x} \in \mathfrak{d}_A$. It is called *self-adjoint* if it is hermitian and $\mathfrak{d}_A = \mathfrak{d}_{A^\dagger}$.

Example 5.1.9:

1. For $\mathcal{H} = L^2(\mathbb{R})$, the position and momentum operator are *self-adjoint*, but $\mathfrak{d}_X, \mathfrak{d}_P \neq \mathcal{H}$.
2. An orthogonal projector P on a *closed* linear subspace $F \subset \mathcal{H}$ of a Hilbert space $\mathcal{H}$ has $\mathfrak{d}_P = \mathcal{H}$ and satisfies $P^\dagger = P$. Projection operators are self-adjoint. Conversely, any linear continuous operator P on $\mathcal{H}$ with $\mathfrak{d}_P = \mathcal{H}$, $P^\dagger = P$ and $P \circ P = P$ is a projection operator on a closed subspace $F \subset \mathcal{H}$.

Remark 5.1.10:

1. The condition that $\mathfrak{d}_A$ is dense in $\mathcal{H}$ is needed for the uniqueness of the adjoint. If there exists a $\boldsymbol{z} \in \mathcal{H}$ such that $\langle \boldsymbol{x}, A\boldsymbol{y}\rangle = \langle \boldsymbol{z}, \boldsymbol{y}\rangle$ for all $\boldsymbol{y} \in \mathfrak{d}_A$, the fact that $\mathfrak{d}_A$ is dense in $\mathcal{H}$ guarantees that $\boldsymbol{z}$ is unique.
2. The distinction between hermitian and self-adjoint is essential and physically meaningful. The spectral theorem exists only for self-adjoint operators, and only self-adjoint Hamiltonians give rise to a *unitary time evolution*.
3. In order to be interpreted as a physical observable, operators on a Hilbert space need to be *self-adjoint*. Hermitian is *not* enough.

Theorem 5.1.11: (Hellinger Toeplitz)

If a linear operator A is defined everywhere on a Hilbert space $\mathcal{H}$: $\mathfrak{d}_A = \mathcal{H}$ and $\langle A\boldsymbol{x}, \boldsymbol{y}\rangle = \langle \boldsymbol{x}, A\boldsymbol{y}\rangle$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{H}$, then A is bounded.

The Hellinger Toeplitz theorem implies that *no* unbounded operator that is a physical observable (and therefore self-adjoint) can be defined on the whole Hilbert space. This confirms the importance of considering operators whose domain is not the whole Hilbert space.

We now return to the bounded operators that are defined on the whole Hilbert space. We distinguish the following important cases.

Definition 5.1.12: (Normal, unitary, positive operator)

A bounded linear operator $A \in B(\mathcal{H})$ on a Hilbert space $\mathcal{H}$ is called *normal* if $A^\dagger A = AA^\dagger$. It is called *unitary*, if it is invertible and $A^\dagger A = AA^\dagger = 1$. It is called *positive*, if it is self-adjoint and $\langle \boldsymbol{x}, A\boldsymbol{x}\rangle \geq 0$ for all $\boldsymbol{x} \in \mathcal{H}$.

Example 5.1.13:

1. If A is self-adjoint, it is normal.
2. For all operators $A \in B(\mathcal{H})$, the operators $AA^\dagger$, $A^\dagger A$ are positive.
3. If A is self-adjoint and U unitary, $UAU^\dagger$ is self-adjoint.
4. If A is positive and U unitary $UAU^\dagger$ is positive.

Exercise 39:

Let $\mathcal{H}$ be a Hilbert space. Show that if $A \in B(\mathcal{H})$ is self-adjoint, then the exponential

$$\exp(iA) = \sum_{k=0}^{\infty} \frac{(iA)^k}{k!}$$

exists and is unitary.

Hint:

1. Show that if $A \in B(\mathcal{H})$ there exists a constant $c > 0$ such that

$$\sum_{k=0}^{n} \frac{||i^k A^k||}{k!} \leq \sum_{k=0}^{n} \frac{c^k}{k!}$$

where $||A|| = \sup\{||A\boldsymbol{x}|| \mid \boldsymbol{x} \in \mathcal{H}, ||\boldsymbol{x}|| = 1\}$. The series $\sum_{k=0}^{n} \frac{i^k A^k}{k!}$ therefore converges absolutely towards on element $\exp(iA) = \sum_{k=0}^{\infty} \frac{i^k A^k}{k!} \in B(\mathcal{H})$. Show that $||(iA)^\dagger|| = ||A||$ and therefore $\exp((iA)^\dagger)$ exists.

2. Show by reordering the series that

$$\exp(iA) \cdot \exp((iA)^\dagger) = \exp(iA)\exp(-iA) = e.$$

Definition 5.1.14: (Hilbert Schmidt operator)

A *Hilbert Schmidt operator* on an infinite-dimensional separable Hilbert space $\mathcal{H}$ is an operator O for which the series $\sum_{i=1}^{\infty} ||O\, e_i||^2$ converges for a Hilbert basis $\{e_i\}_{i\in\mathbb{N}}$. For a Hilbert Schmidt operator O, the limit

$$||O||_{HS} = \sqrt{\sum_{i=1}^{\infty} ||Oe_i||^2} \tag{5.3}$$

is called the *Hilbert Schmidt norm* of O (with respect to the Hilbert basis $\{e_i\}_{i\in\mathbb{N}}$).

Lemma 5.1.15:

1. For a Hilbert Schmidt operator A, the Hilbert Schmidt norm does not depend on the choice of the Hilbert basis. Its adjoint $A^\dagger$ is also a Hilbert Schmidt operator and $||A||_{HS} = ||A^\dagger||_{HS}$.

2. Hilbert Schmidt operators are compact.

3. Any sequence $\{A_{ij}\}_{i,j\in\mathbb{N}}$, $A_{ij} \in \mathbb{C}$ such that $\sum_{i,j=0}^\infty |A_{ij}|^2 < \infty$ defines a Hilbert Schmidt operator. The adjoint of a Hilbert Schmidt operator defined this way is given by the sequence $\{B_{ij}\}_{i,j\in\mathbb{N}}$, $B_{ij} = \bar{A}_{ji}$.

4. If A is a Hilbert Schmidt operator and $B \in B(\mathcal{H})$, BA and AB are Hilbert Schmidt operators.

Proof: 1. Let $\{f_j \mid j \in \mathbb{N}\}$ be another Hilbert basis. Parseval's equation implies

$$||A\, e_i||^2 = \sum_{j=1}^\infty |\langle A\, e_i, f_j\rangle|^2 = \sum_{j=1}^\infty |\langle e_i, A^\dagger\, f_j\rangle|^2 \tag{5.4}$$

$$||A^\dagger f_j||^2 = \sum_{i=1}^\infty |\langle e_i, A^\dagger\, f_j\rangle|^2. \tag{5.5}$$

Convergence of $\sum_{i=0}^m ||A\, e_i||^2$ is therefore equivalent to convergence of $\sum_{i=1}^n \sum_{j=1}^m |\langle A\, e_i, f_j\rangle|^2$ and of the sum $\sum_{j=1}^\infty ||A^\dagger f_j||^2$. If either sum converges , the other converges, and we have $\sum_{i=1}^\infty ||Ae_i|| = \sum_{j=0}^\infty ||A^\dagger f_j||$. Hence, A is a Hilbert Schmidt operator if and only if $\sum_{i=1}^\infty ||A\, f_i||^2 < \infty$ for *all* Hilbert bases. Moreover, A is a Hilbert Schmidt operator if and only if $A^\dagger$ is a Hilbert Schmidt operator. The square of the norm $||A||_{HS}^2 = \sum_{i=1}^\infty ||Ae_i||^2 = ||A^\dagger||_{HS}^2$ is independent of the basis.

2. It is sufficient to prove that the closure $\overline{A(B_1)}$ of the image of the unit ball $B_1 = \{\boldsymbol{x} \in \mathcal{H} \mid ||\boldsymbol{x}|| = 1\}$ is compact for any Hilbert Schmidt operator A. We consider a sequence $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$, $\boldsymbol{x}_n \in B_1$. Using Gram Schmidt orthogonalisation, we can construct a sequence $\{\boldsymbol{y}_n\}_{n\in\mathbb{N}}$ as in the proof of Lemma 4.4.7 which satisfies $\langle \boldsymbol{y}_n, \boldsymbol{y}_m\rangle = \delta_{nm}$. We can then complete $\{\boldsymbol{y}_n\}_{n\in\mathbb{N}}$ to a Hilbert basis $\{f_i\}_{i\in\mathbb{N}}$. The convergence of the Hilbert Schmidt norm implies $||Af_n|| \to 0$ for $n \to \infty$ and therefore $||A\boldsymbol{y}_n|| \to 0$ for $n \to \infty$. As the sequence $\boldsymbol{x}_n$ is given by taking finite linear combinations of the elements $\boldsymbol{y}_n$, this implies $||O\boldsymbol{x}_n|| \to 0$ for $n \to \infty$. The sequence $||A\boldsymbol{x}_n||$ therefore converges towards 0 for $n \to \infty$ and $\overline{A(B_1)}$ is compact.

3. We define a linear operator A on $\mathcal{H}$ by setting

$$Ae_i = \sum_{j=1}^\infty A_{ij} e_j. \tag{5.6}$$

This is well-defined since Parseval's equation implies

$$||Ae_i||^2 = \sum_{j=1}^\infty |\langle e_j, Ae_i\rangle|^2 = \sum_{j=1}^\infty |A_{ij}|^2 < \infty. \tag{5.7}$$

For general $\boldsymbol{x} \in \mathcal{H}$, we define

$$A\boldsymbol{x} = \sum_{i,j=0}^\infty A_{ij} x^i e_j \quad \text{where} \quad \boldsymbol{x} = \sum_{i=0}^\infty x^i e_i, \sum_{i=0}^\infty |x^i|^2 < \infty. \tag{5.8}$$

which converges since

$$||A\boldsymbol{x}||^2 = \sum_{i,j=0}^{\infty} |A_{ij}|^2 |x_i|^2 \leq \left(\sum_{i,j=1}^{\infty} |A_{ij}|^2 \right) \left(\sum_{k=1}^{\infty} |x_k|^2 \right) < \infty. \tag{5.9}$$

Hence, A is well defined and

$$\sum_{i=1}^{\infty} ||Ae_i||^2 \leq \sum_{i,j=1}^{\infty} |A_{ij}|^2 < \infty.$$

Let the adjoint of A be given by a sequence $\{B_{ij}\}_{i,j\in\mathbb{N}}$. By Parseval's equation, we have $\langle e_j, Ae_i\rangle = A_{ij}$, which implies $\langle A^\dagger e_j, e_i\rangle = \overline{\langle e_i, A^\dagger e_j\rangle} = \langle e_j, Ae_i\rangle = A_{ij} = \bar{B}_{ji}$.

4. If A is a Hilbert Schmidt operator, $B \in B(\mathcal{H})$, then we have $||BA\boldsymbol{x}|| \leq ||B|| \cdot ||A\boldsymbol{x}||$ for all $\boldsymbol{x} \in \mathcal{H}$, where $||B|| = \sup\{||B\boldsymbol{x}|| \mid \boldsymbol{x} \in \mathcal{H}, ||\boldsymbol{x}|| = 1\} < \infty$. This implies

$$\sum_{i=1}^{\infty} ||BAe_i||^2 \leq ||B||^2 \cdot \sum_{i=1}^{n} ||Ae_i||^2 < \infty,$$

and therefore BA is a Hilbert Schmidt operator. To show that AB is a Hilbert Schmidt operator, we recall that $A^\dagger$ is a Hilbert Schmidt operator if and only if A is a Hilbert Schmidt operator and $B^\dagger$ is bounded. Hence, we have AB is a Hilbert Schmidt operator if and only if $B^\dagger A^\dagger$ is a Hilbert Schmidt operator, and the latter follows from the first statement.

□

Lemma 5.1.16:

1. The vector space $l^{2,2}$ of doubly square summable sequences $\{A_{ij}\}_{i,j\in\mathbb{N}}$ becomes a Hilbert space when equipped with the hermitian form

$$\langle \{A_{ij}\}, \{B_{ij}\}\rangle_{l^{2,2}} = \sum_{i,j=1}^{\infty} \bar{B}_{ij} A_{ij} \qquad \forall \{A_{ij}\}, \{B_{ij}\} \in l^{2,2}.$$

2. The vector space $\mathcal{H}_{HS}$ of Hilbert Schmidt operators becomes a Hilbert space when equipped with the hermitian product

$$\langle A, B\rangle_{HS} = \sum_{i=1}^{\infty} \langle e_i, B^\dagger A e_i\rangle \qquad \forall A, B \in \mathcal{H}_{HS}.$$

3. A unitary isomorphism between $l^{2,2}$ and $\mathcal{H}_{HS}$ is given by the linear map $\{A_{ij}\}_{i,j\in\mathbb{N}} \mapsto A$, where A is defined by

$$Ae_i = \sum_{j=1}^{\infty} A_{ij} e_j.$$

Remark 5.1.17: The relation between Hilbert Schmidt operators and doubly square summable sequences $\{A_{ij}\}_{i,j\in\mathbb{N}}$ mimics the relation between Hilbert spaces and the Hilbert space of absolutely square summable sequences $\{x_i\}_{i\in\mathbb{N}}$. Previously, we found that a choice of a Hilbert basis provides an isomorphism between the Hilbert space and the Hilbert space l^2 of square summable sequences. In the case of a Hilbert Schmidt operator, the choice of a Hilbert basis gives rise to an isomorphism between the Hilbert space of Hilbert Schmidt operators and the Hilbert space of doubly square summable sequences $\{A_{ij}\}_{i,j\in\mathbb{N}}$.

Note, however, that Hilbert Schmidt operators can capture only a very small part of the operators relevant to quantum physics. This is due to the fact that they are compact, which is a very strong requirement. Most operators arising in quantum mechanics or the study of infinite-dimensional Hilbert spaces are non-compact.

We will now generalise the concept of a *trace* to the infinite-dimensional case. However, we note that unlike in the finite-dimensional case, the infinite dimensional version of the trace does not exists for all operators. The convergence of the trace is guaranteed for a special class of operators, which are called *trace class.*

Definition 5.1.18: (Trace class operator)

An operator $A \in B(\mathcal{H})$ is called *trace class* or *nuclear* if it is a Hilbert Schmidt operator and

$$||A||_1 = \sup\{|\langle A, B\rangle_{HS} \mid B \text{ Hilbert Schmidt },\ ||B|| \leq 1\} < \infty.$$

Some important properties of trace class operators are given by the following lemma.

Lemma 5.1.19: The set of trace class operators is a linear subspace of $B(\mathcal{H})$. If $A \in B(\mathcal{H})$ is trace class, we have

1. $||A^\dagger||_1 = ||A||_1$ and $A^\dagger$ is trace class.
2. $||A||_{HS} \leq ||A||_1$.
3. For all $C \in B(\mathcal{H})$: AC and CA are trace class.

Proof:

1. We have $\langle A, B\rangle_{HS} = \langle A^\dagger, B^\dagger\rangle_{HS}$ for all Hilbert Schmidt operators A, B and $||B^\dagger|| = ||B||$. This implies

$$\begin{aligned}||A^\dagger||_1 =&\sup\{|\langle A^\dagger, B\rangle_{HS} \mid B \text{ Hilbert Schmidt } ||B|| \leq 1\}\\ =&\sup\{|\langle A, B\rangle_{HS} \mid B \text{ Hilbert Schmidt } ||B|| \leq 1\} = ||A||_1.\end{aligned}$$

2. We have

$$||A||^2_{HS} = \langle A, A\rangle_{HS} \leq ||A|| \cdot \sup\{|\langle A, B\rangle_{HS} \mid B \text{ Hilbert Schmidt } ||B|| \leq 1\} \leq ||A||_{HS}||A||_1.$$

3. Let A be trace class and $C \in B(\mathcal{H})$. We have $\langle CA, B\rangle_{HS} = \langle A, c^\dagger B\rangle_{HS}$ and $C^\dagger B$ bounded if C, B bounded. This implies

$$\begin{aligned}||CA||_1 =&\sup\{|\langle CA, B\rangle_{HS} \mid B \text{ Hilbert Schmidt } ||B|| \leq 1\}\\ =&\sup\{|\langle A, C^\dagger B\rangle_{HS} \mid B \text{ Hilbert Schmidt } ||B|| \leq 1\} \leq ||C^\dagger|| \cdot ||A||_1 < \infty.\end{aligned}$$

To show that AC trace class, we note that $|\langle A, B\rangle_{HS}| = |\langle B^\dagger, A^\dagger\rangle_{HS}|$ and $(AC)^\dagger = C^\dagger A^\dagger$. As $||A^\dagger||_1 = ||A||_1$, the statement for CA proves the claim. □

We are now ready to define the trace of a trace class operator.

Definition 5.1.20: (Trace)

Let A be a trace class operator on a Hilbert space $\mathcal{H}$ and $\{e_i\}_{i\in\mathbb{N}}$ a Hilbert basis. The *trace* of A is defined as

$$\mathrm{Tr}\,(A) = \sum_{i=1}^{\infty} \langle e_i, Ae_i\rangle. \tag{5.10}$$

The series converges absolutely and is independent of the choice of basis.

Proof: Let P_n be the orthogonal projector on the closed subspace $\mathrm{Span}(\{e_1, .., e_n\})$. And set $\tilde{P}_n(e_i) = \lambda_i(\langle e_i, Ae_i\rangle)$, where $|\langle e_i, Ae_i\rangle| = \lambda_i\langle e_i, Ae_i\rangle$, $|\lambda_i| = 1$. Then $\tilde{P}_n$ is a Hilbert Schmidt operator and $||\tilde{P}_n|| \leq 1$. This implies

$$\sum_{i=1}^{n} |\langle e_i, Ae_i\rangle| = \langle \tilde{P}_n, A\rangle_{HS} \leq \sup\{|\langle A, B\rangle_{HS} \mid B \text{ Hilbert Schmidt } ||B|| \leq 1\} = ||A||_1$$

for all $n \in \mathbb{N}$. The series therefore converges absolutely. The proof of the independence of the choice of basis is left as an exercise. □

Lemma 5.1.21:

1. The trace is *linear* $\mathrm{Tr}\,(tA + sB) = t\mathrm{Tr}\,(A) + s\mathrm{Tr}\,(B)$ for all $t, s \in \mathbb{C}$, A, B trace class.
2. The trace satisfies $\mathrm{Tr}\,\left(AA^\dagger\right) = \mathrm{Tr}\,\left(A^\dagger A\right)$ for all A trace class.
3. The trace satisfies $\mathrm{Tr}\,\left(UBU^\dagger\right) = \mathrm{Tr}\,(B)$ for all B trace class and U unitary.
4. For all A trace class and $B \in B(\mathcal{H})$, we have $\mathrm{Tr}\,(AB) = \mathrm{Tr}\,(BA)$.

Proof: The first three properties are left as an exercise. To prove the last property, we first consider a self-adjoint trace class operator $A = A^\dagger$ and choose a Hilbert basis $\{e_i\}_{i\in\mathbb{N}}$. Then we have $A_{ij} = \langle e_j, Ae_i\rangle = A_{ji} = \langle e_i, Ae_j\rangle$. Using Parseval's equation, we find

$$\mathrm{Tr}\,(AB) = \sum_{i=1}^{\infty} \langle e_i, ABe_i\rangle = \sum_{i=1}^{\infty} \langle Ae_i, Be_i\rangle = \sum_{i,j=1}^{\infty} A_{ij}\langle e_j, Be_i\rangle = \sum_{j=1}^{\infty} \langle e_j, BAe_i\rangle = \mathrm{Tr}\,(BA)\,.$$

To prove this identity for general trace class operators A, we note that we can decompose the Hilbert Schmidt operator A as $A = \frac{1}{2}(A + A^\dagger) + \frac{1}{2}(A - A^\dagger)$ with $\frac{1}{2}(A + A^\dagger)$, $\frac{i}{2}(A - A^\dagger)$ self-adjoint. □

Exercise 40:

1. Show that the trace is linear.

2. Let $\mathcal{H}$ be a Hilbert space with Hilbert basis $B = \{e_i\}_{i\in\mathbb{N}}$ and $O \in B(\mathcal{H})$ a trace class operator. Show that

$$\mathrm{Tr}\,(O) = \sum_{i=1}^{\infty} \langle e_i, Oe_i\rangle$$

is independent of the choice of the Hilbert basis.

3. Show that for all trace class $A \in B(\mathcal{H})$ $\mathrm{Tr}\,\left(A^\dagger A\right) = \mathrm{Tr}\,\left(AA^\dagger\right)$ and for all trace class $B \in B(\mathcal{H})$ and unitary $U \in B(\mathcal{H})$, $\mathrm{Tr}\,\left(UBU^\dagger\right) = \mathrm{Tr}\,(B)$.

Lemma 5.1.22: An operator $A \in B(\mathcal{H})$ on a Hilbert space $\mathcal{H}$ is a Hilbert-Schmidt operator if and only if $\mathrm{Tr}\left(A^\dagger A\right) < \infty$.

Proof: Let $\{e_i\}_{i\in\mathbb{N}}$ be a Hilbert basis. We have

$$\mathrm{Tr}\left(A^\dagger A\right) = \sum_{i=1}^{\infty} \langle e_i, A^\dagger A e_i\rangle = \sum_{i=1}^{\infty} ||Ae_i||^2 = ||A||_{HS}^2$$

Hence $\mathrm{Tr}\left(A^\dagger A\right) < \infty$ if and only if A is a Hilbert Schmidt operator. □

We will now investigate the infinite dimensional analogue of eigenvalues, eigenvectors and of the diagonalisation of matrices. The first important difference between the infinite-dimensional and the finite-dimensional case is that in the infinite-dimensional one needs to distinguish spectral values and eigenvalues. The former can be viewed as a generalisation of the latter but are *not* associated with eigenvectors.

Definition 5.1.23: (Spectrum of a linear operator, eigenvalues)

Let O be a linear operator on a Hilbert space $\mathcal{H}$. A number $z \in \mathbb{C}$ is said to be *in the resolvent set* of O if $O - z \cdot 1$ is invertible with a *bounded inverse*. The *spectrum* $\sigma(O)$ of O is the set of all $z \in \mathbb{C}$ which are not in the resolvent set of O.

An *eigenvalue of* O is a number $z \in \mathbb{C}$ such that there exists a vector $\boldsymbol{x} \in \mathcal{H} \setminus \{0\}$ with $O\boldsymbol{x} = z\boldsymbol{x}$. The vector $\boldsymbol{x}$ is called an eigenvector of $\mathcal{H}$ with eigenvalue z. The set of all eigenvectors of O for a given eigenvalue $z \in \mathbb{C}$ is a closed subspace of $\mathcal{H}$, the *eigenspace for eigenvalue* z.

Remark 5.1.24:

1. If O is a bounded linear operator on $\mathcal{H}$, $z \in \mathbb{C}$ is in the resolvent set of O if and only if $O - z \cdot 1$ is invertible.
2. If $\dim(\mathcal{H}) < \infty$, $z \in \sigma(O)$ if and only if z is an eigenvalue of O.
3. If $\mathcal{H}$ is infinite-dimensional, there exist elements $z \in \sigma(O)$ that are not eigenvalues. An eigenvalue is often called a *discrete* eigenvalue in the physics literature. Elements of the spectrum are called *spectral values* or *continuous eigenvalues* in the physics literature.

Remark 5.1.25: Spectral values or continuous eigenvalues usually correspond to situations, where a (generalised) eigenvector can be found but is not an element of the Hilbert space. An example is the position operator on $L^2(\mathbb{R})$ $X : \psi \mapsto X\psi$ with $X\psi(x) = x\psi(x)$. A (generalised) eigenvector of this operator is a distribution δ_y, since we have $X\delta_y(x) = y\delta_y(x)$. However, this distribution is *not* an element of the Hilbert space $L^2(\mathbb{R})$ and therefore not an eigenvector. The position operator has no eigenvalues and eigenvectors. Its spectrum consists of the entire real line, since $X - \lambda 1$ is not invertible with a bounded inverse for any $\lambda \in \mathbb{R}$.

As we have seen previously, for hermitian matrices, all eigenvalues are *real* numbers. A similar statement holds for the spectra of hermitian operators.

Example 5.1.26: (Spectra of hermitian and positive operators)

1. A is hermitian if and only if $A - z \cdot 1$ is invertible with bounded inverse for all $z \in \mathbb{C}$ with $\mathrm{Im}(z) \neq 0$. Hermitian operators have a *real spectrum* $\sigma(A) \subset \mathbb{R}$.

2. A hermitian operator A is positive if and only if $\sigma(A) \subset \mathbb{R}_0^+$.

Exercise 41:

Show that the only possible eigenvalues of a projection operator are 0 and 1.

We are now ready to investigate the decomposition of Hilbert spaces into linear subspaces of eigenvectors associated with eigenvalues. As it can be expected that spectral values which are not eigenvalues are associated with complications, we start with the simplest situation, namely the one of *compact*, normal operators. In this case, it turns out that there is at most one spectral value which is not an eigenvalue.

Theorem 5.1.27: (Spectral decomposition for compact, normal operators)

Let $A \in B(\mathcal{H})$ compact and normal. Then $0 \in \sigma(A)$ and $\sigma(A)$ is either finite or of the form $\sigma(A) = \{0, \lambda_1, \lambda_2, ...\}$ where $\{\lambda_n\}_{n\in\mathbb{N}}$, $\lambda_i \in \mathbb{C} \setminus \{0\}$ is a sequence that converges to 0 with $n \to \infty$. For each λ_i, $i \in \mathbb{N}$ the eigenspace $E_i = \{\boldsymbol{x} \in \mathcal{H} \mid A\boldsymbol{x} = \lambda_i \boldsymbol{x}\}$ is of finite dimension $0 < \dim(E_i) < \infty$. The eigenspaces E_i are orthogonal: $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$ for all $\boldsymbol{x} \in E_i$, $\boldsymbol{y} \in E_j$ with $i \neq j$. The operator A can be expressed in terms of the projection operators P_{E_i} on E_i as

$$A = \sum_{i=1}^{\infty} \lambda_i P_{E_i}. \tag{5.11}$$

Remark 5.1.28: In the physics literature identity (5.11) is often expressed in bra-ket notation as

$$A = \sum_{i=1}^{\infty} \lambda_i |e_i\rangle\langle e_i|.$$

Corollary 5.1.29: (Spectrum of compact, positive operators)

If A is a compact, positive operator, it is in particular self-adjoint, which implies normal and all eigenvalues are positive or zero. Hence, the spectrum of a compact, positive operator is either finite with $0 \in \sigma(A)$ or takes the form $\sigma(A) = \{0, \lambda_1, \lambda_2, ...\}$ where $(\lambda_n)_{n\in\mathbb{N}}$, $\lambda_n \in \mathbb{R}^+$ is a sequence that converges to 0. The eigenspaces E_i, $i \in \mathbb{N}$ are orthogonal, and the operator A is given by

$$A = \sum_{i=1}^{\infty} \lambda_i P_{E_i}. \tag{5.12}$$

Theorem 5.1.27 provides us with a rather straightforward generalisation of the diagonalisation of complex matrices. However, its disadvantage is that it is only valid for a very specific class of of operators- most operators relevant to physics are not compact. In particular, identity operators, projectors on infinite-dimensional subspaces and Hamilton operators of systems whose energy is not bounded are non-compact operators. We therefore need a generalisation of this theorem. This forces us to deal with spectral values which are not eigenvalues and to introduce the notion of a spectral measure.

Definition 5.1.30: (Spectral measure)

A *Borel set* in $k = \mathbb{C}$ or $k = \mathbb{R}$ is a subset $E \subset k$ that can be expressed as a countable union $\bigcup_{i=1}^{\infty} E_i$ or countable intersection $\bigcap_{i=1}^{\infty} E_i$ of open subsets of k, their complements and countable unions and intersections thereof. We denote by $\mathrm{Bor}(k)$ the Borel sets in k.

A spectral measure on k is a function $P : E \in \text{Bor}(k) \to P(E) \in B(\mathcal{H})$ which assigns to each Borel set in k a projection operator $P(E) \in B(\mathcal{H})$ such that

1. $P(\emptyset) = 0$
2. $P(k) = 1$
3. $P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$ for mutually disjoint Borel sets: $E_i \subset k$, $E_i \cap E_j = \emptyset \; \forall i \neq j$.

Lemma 5.1.31: (Properties of spectral measures)

A spectral measure $P : \text{Bor}(k) \to B(\mathcal{H})$ has the following properties

1. *Monotonicity*: $E_1 \subset E_2 \Rightarrow P(E_1) \leq P(E_2)$.
2. *(Orthogonality)*: $E_1 \cap E_2 = \emptyset \Rightarrow P(E_1) \perp P(E_2)$, i.e. $P(E_2) \circ P(E_1) = P(E_1) \circ P(E_2) = 0$.
3. For all $E_1, E_2 \in \text{Bor}(k)$: $P(E_1 \cap E_2) = P(E_1)P(E_2)$.

Proof: Exercise.

We will now use the concept of spectral measure to define *integrals of operators*. For this we need the following lemma.

Lemma 5.1.32: Let $P :$ Bor$(k) \to B(\mathcal{H})$ be a spectral measure. Then for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{H}$, the function $\mu_{\boldsymbol{x},\boldsymbol{y}} :$ Bor$(k) \to \mathbb{C}$ defined by $\mu_{\boldsymbol{x},\boldsymbol{y}}(E) = \langle P(E)\boldsymbol{x}, \boldsymbol{y}\rangle$ is a measure on k, i.e. it satisfies

1. $\mu_{\boldsymbol{x},\boldsymbol{y}}(\emptyset) = 0$
2. $A \subset \bigcup_{j=1}^{\infty} A_j$, $A_j \in \text{Bor}(k)$ for all $j \in \mathbb{N} \Rightarrow \mu_{\boldsymbol{x},\boldsymbol{y}}(A) \leq \sum_{i=1}^{\infty} \mu_{\boldsymbol{x},\boldsymbol{y}}(A_j)$.

Lemma 5.1.33: The measure $\mu_{\boldsymbol{x},\boldsymbol{y}}$ is of *bounded variation*. For all $E \in \text{Bor}(k)$ we have

$$||\mu_{\boldsymbol{x},\boldsymbol{y}}(E)|| = \sup\{\sum_{n=1}^{\infty} |\mu(E_i)| \mid E = \bigcup_{i=1}^{\infty} E_i \ , \ E_i \in \text{Bor}(k), E_i \cap E_j = \emptyset \text{ for } i \neq j\} \leq ||\boldsymbol{x}||||\boldsymbol{y}||$$

In order to define the integral of operators associated with measures $\mu_{\boldsymbol{x},\boldsymbol{y}}$, $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{H}$ we need the notion of the *generalised Lebesgue Stieltjes integral.*

Definition 5.1.34: (Lebesgue Stieltjes integral)

Let μ be a measure of bounded variation on k, i.e. a measure μ on k for which there exists a constant $c \in \mathbb{R}^+$ such that

$$||\mu(E)|| = \sup\{\sum_{n=1}^{\infty} |\mu(E_i)| \mid E = \bigcup_{i=1}^{\infty} E_i \ , \ E_i \in \text{Bor}(k), E_i \cap E_j = \emptyset \text{ for } i \neq j\} \leq c \quad \forall E \in \text{Bor}(k).$$

Let $f : k \to \mathbb{R}$ be a *Borel function* or *measurable function*, i.e. a function such that the preimage $f^{-1}(T)$ is a Borel set for all Borel sets $T \subset \mathbb{R}$. Suppose that f is non-negative: $f \geq 0$ almost everywhere, and bounded, $||f||_\infty = \sup_{x \in k}\{|f(x)|\} < \infty$. Then the *generalised Lebesgue Stieltjes integral* of f over a Borel set E with respect to μ is defined as

$$\int_E f d\mu := \sup\left\{\sum_{i=1}^{n} \inf\{f(x) \mid x \in E_i\}\mu(E_i) \mid E = \bigcup_{i=1}^{\infty} E_i \ , \ E_i \in \text{Bor}(k), E_i \cap E_j = \emptyset \text{ for } i \neq j\right\}.$$

If $f : k \to \mathbb{C}$ is a complex valued, bounded Borel function, then the functions $\mathrm{Re}(f)^{\pm} : k \to \mathbb{R}_0^+$, $\mathrm{Im}(f)^{\pm} : k \to \mathbb{R}_0^+$ defined by

$$\begin{aligned} &\mathrm{Re}(f)^+(z) = \max\{\mathrm{Re}(f)(z), 0\} \qquad \mathrm{Re}(f)^-(z) = -\min\{\mathrm{Re}(f)(z), 0\} \\ &\mathrm{Im}(f)^+(z) = \max\{\mathrm{Im}(f)(z), 0\} \qquad \mathrm{Im}(f)^-(z) = -\min\{\mathrm{Im}(f)(z), 0\} \qquad \forall z \in k \end{aligned} \tag{5.13}$$

are non-negative, bounded Borel functions. The *generalised Lebesgue Stieltjes integral* of f over a Borel set E is defined as

$$\int_E f d\mu := \int_E \mathrm{Re}(f)^+ d\mu - \int_E \mathrm{Re}(f)^- d\mu + i \int_E \mathrm{Im}(f)^+ d\mu - i \int_E \mathrm{Im}(f)^- d\mu. \tag{5.14}$$

Remark 5.1.35: The integral $\int_E f d\mu$ is well defined. If $f : k \to \mathbb{R}$ is non-negative, bounded Borel function with $||f||_\infty = c$ and μ a measure of bounded variation, we have

$$0 \leq \inf\{f(x) \mid x \in E_i\}\mu(E_i) \leq c \cdot \mu(E_i) \qquad \forall E_i \in \mathrm{Bor}(k)$$

and therefore

$$\begin{aligned} \int_E f d\mu =&\sup\left\{\sum_{i=1}^n \inf\{f(x) \mid x \in E_i\}\mu(E_i) \mid E = \bigcup_{i=1}^\infty E_i \;,\; E_i \in \mathrm{Bor}(k), E_i \cap E_j = \emptyset \text{ for } i \neq j\right\} \\ \leq& c \cdot \sup\left\{\sum_{n=1}^\infty |\mu(E_i)| \mid E = \bigcup_{i=1}^\infty E_i \;,\; E_i \in \mathrm{Bor}(k), E_i \cap E_j = \emptyset \text{ for } i \neq j\right\} = c||\mu|| < \infty. \end{aligned}$$

We can now use the Lebesgue Stieltjes integral for the measure $\mu_{\boldsymbol{x},\boldsymbol{y}}$ to define a bounded hermitian form on $\mathcal{H}$ and to obtain an associated operator via Riesz's theorem.

Lemma 5.1.36: Let $f : k \to \mathbb{C}$ be a bounded *Borel function* Then its integral with respect to the measure $\mu_{\boldsymbol{x},\boldsymbol{y}}$ exists and the map $g : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$ $g(\boldsymbol{x}, \boldsymbol{y}) = \int_k f d\mu_{\boldsymbol{x},\boldsymbol{y}}$ defines a bounded hermitian form on $\mathcal{H}$. We have

$$|g(\boldsymbol{x}, \boldsymbol{y})| \leq ||\boldsymbol{x}||||\boldsymbol{y}||||f||_\infty \;\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{H}.$$

By Riesz's theorem there exists a unique operator $F \in B(\mathcal{H})$ with $\langle F\boldsymbol{x}, \boldsymbol{y}\rangle = \int_{\mathbb{C}} f d\mu_{\boldsymbol{x}\boldsymbol{y}}$. we denote this operator by $F =: \int_k f(z) dP(z)$.

This allows us to formulate the spectral theorem for self-adjoint operators.

Theorem 5.1.37: (Spectral decomposition for self-adjoint operators)

For a densely defined self-adjoint operator A on $\mathcal{H}$ there exists a unique spectral measure $P_A :\ \mathrm{Bor}(\mathbb{R}) \to B(\mathcal{H})$ such that

$$A = \int_{\mathbb{R}} x \, dP_A(x).$$

Conversely, any operator defined by a spectral measure in this way is self-adjoint.

With the definition $f(A) := \int_{\mathbb{R}} f(x) dP_A(x)$ for bounded Borel functions $f : \mathbb{R} \to \mathbb{C}$, we have $P_A(E) = \chi_E(A)$, where χ_E is the characteristic function of a Borel set E

$$\chi_E(x) = \begin{cases} 1 & x \in E \\ 0 & x \notin E \end{cases}. \tag{5.15}$$

In many books on quantum mechanics, this theorem is presented in terms of *spectral sequences* or *resolutions of the identity*. One can show that spectral sequences are equivalent to spectral measures on $\mathbb{R}$.

Definition 5.1.38: (Resolution of the identity)

A resolution of the identity is a function $\lambda \in \mathbb{R} \to P_\lambda \in B(\mathcal{H})$ into the set of orthogonal projection operators on $\mathcal{H}$ such that

1. $\lambda \le \mu \Rightarrow P_\lambda \le P_\mu$
2. $\lim_{\lambda\to-\infty} P_\lambda$, $\lim_{\lambda\to\infty} P_\lambda$ exist and satisfy
$$P_{-\infty} = \lim_{\lambda\to-\infty} P_\lambda = 0 \qquad P_\infty := \lim_{\lambda\to\infty} P_\lambda = \mathrm{id}_{\mathcal{H}}.$$
3. $\lim_{\epsilon\to 0,\epsilon>0} P_{\lambda+\epsilon} = P_\lambda$.

Here, all limits are taken with respect to the norm $||O|| = \sup\{||O\boldsymbol{x}|| \mid \boldsymbol{x} \in \mathcal{H}, ||\boldsymbol{x}|| = 1\}$ for bounded operators on $\mathcal{H}$. The resolution of the identity is called *constant* in a point $x \in \mathbb{R}$ if there exists an interval $]a, b[\subset \mathbb{R})$, $a < x < b$, such that $P_a\boldsymbol{y} = P_b\boldsymbol{y}$ for all $\boldsymbol{y} \in \mathcal{H}$. Otherwise it is called *discontinuous* in $x \in \mathbb{R}$.

Exercise 42: Let $P : \mathrm{Bor}(\mathbb{R}) \to B(\mathcal{H})$ a spectral measure on the Borel sets of $\mathbb{R}$. Show that the function $P_\lambda = P(]-\infty, \lambda])$, $\lambda \in \mathbb{R}$ defines a resolution of the identity. Show that two spectral measures $P, Q : \mathrm{Bor}(\mathbb{R}) \to B(\mathcal{H})$ give rise to the same resolution of the identity satisfy $P = Q$.

5.2 Axioms of quantum mechanics

Axiom 1: The observables, i.e. the measurable quantities of a physical system, are in one-to-one correspondence with self-adjoint operators on a separable Hilbert space. The possible results of measurements of an observable are given by the spectrum of the associated self-adjoint operator.

To formulate the second axiom of quantum mechanics, we have to introduce the notion of a *density operator* and of a *pure state*.

Definition 5.2.1: (Density operator)

A *density operator* on a Hilbert space $\mathcal{H}$ is a trace class, positive operator $\rho \in B(\mathcal{H})$ such that $\rho \circ O$ is *trace class* for *all observables* O.

Remark 5.2.2: Note that the fact that ρ is trace class implies directly that $\rho \circ O$ is trace class for all *bounded* operators $O \in B(\mathcal{H})$ by Lemma 5.1.19.

Lemma 5.2.3: For any density operator ρ, the associated *normalised density operator* $\hat{\rho} = \rho/\mathrm{Tr}\,(\rho)$ can be expressed as

$$\hat{\rho} := \frac{\rho}{\mathrm{Tr}\,(\rho)} = \sum_{i=1}^{\infty} \lambda_i P_{e_i} \qquad 0 \le \lambda_i \le 1, \sum_i \lambda_i = 1, \tag{5.16}$$

where $\{e_i\}_{i\in\mathbb{N}}$ is a Hilbert basis and P_{e_i} the projector on $\mathrm{Span}(e_i)$,

Proof: As $\rho = \rho \circ \mathrm{id}_{\mathcal{H}}$ is trace class, which implies $\mathrm{Tr}\,(\rho) < \infty$, the normalised density operator $\hat{\rho} = \rho/\mathrm{Tr}\,(\rho)$ is well-defined. Moreover, ρ is compact, normal and a Hilbert-Schmidt operator. Lemma 5.1.27 and Corollary 5.1.29 then imply that ρ can be expressed as $\rho = \sum_{i=1}^{\infty} \lambda_j P_{E_j}$ where P_{E_j} is the projector on the finite-dimensional eigenspace $E_j = \{\boldsymbol{x} \in \mathcal{H} \mid \rho(\boldsymbol{x}) = \lambda_j \boldsymbol{x}\}$ and $\lambda_j \in \mathbb{C}$ for all $j \in \mathbb{N}$. As ρ is positive, we have $\lambda_j \in \mathbb{R}^+$ for all $j \in \mathbb{N}$.

We have $\mathcal{H} = \bigoplus_{i=1}^{\infty} E_i \oplus E^{\perp}$, where $E^{\perp} = \{\boldsymbol{x} \in \mathcal{H} \mid \langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0 \text{ for all } \boldsymbol{y} \in \bigoplus_{i=1}^{\infty} E_i\}$. With the choice of a Hilbert basis B_0 of $E^{\perp}$ and Hilbert bases B_i of E_i, we obtain a Hilbert basis $B = B_0 \cup B_1 \cup B_2 \cup ... = \{e_i\}_{i \in \mathbb{N}}$ of $\mathcal{H}$. With respect to this Hilbert basis, ρ takes the form

$$\rho = \sum_{i=1}^{\infty} \rho_{ii} P_{e_i} \qquad \rho_{ii} \in \mathbb{R}_0^+ \tag{5.17}$$

where P_{e_i} is the projector on $\mathrm{Span}(\{e_i\})$. We have $\mathrm{Tr}\,(\rho) = \sum_{i=1}^{\infty} \langle e_i, \rho e_i \rangle = \sum_{i=1}^{\infty} \rho_{ii} < \infty$. This implies that the normalised density operator is given by

$$\hat{\rho} = \rho/\mathrm{Tr}\,(\rho) = \sum_{i=1}^{\infty} \hat{\rho}_{ii} P_{e_i} \qquad \text{with } \hat{\rho}_{ii} = \rho_{ii}/\mathrm{Tr}\,(\rho)\ ,\ 0 \leq \hat{\rho}_{ii} \leq 1. \tag{5.18}$$

□

Example 5.2.4:

If $\mathcal{H} = \mathbb{C}^n$, any linear operator is bounded and automatically trace class. A density operator is a hermitian matrix $M \in M(n, \mathbb{C})$. Such a matrix can be diagonalised with eigenvalues $\lambda_i \in \mathbb{R}_0^+$. The normalised density operator is given as a matrix $\hat{M} = M/\mathrm{Tr}\,(M)$. It is a hermitian matrix with eigenvalues $\lambda_i \in [0, 1]$. Diagonalising this operator yields the finite-dimensional version of (5.16).

Definition 5.2.5: (Pure state)

A quantum mechanical system is said to be in a *pure state* if its normalised density operator is of the form $\hat{\rho} = P_{\boldsymbol{x}}$, where $P_{\boldsymbol{x}}$ is the projector on a closed, one-dimensional subspace $\mathrm{Span}(\{\boldsymbol{x}\}) \subset \mathcal{H}$.

Using the notions of a density matrix and a pure state, we can formulate the second axiom of quantum mechanics.

Axiom 2: Quantum mechanics generally only predicts *the probabilities for outcomes of measurements* on quantum systems. The physical state of a quantum mechanical system at a given time is given by a *density operator* ρ. The *expectation value* for the measurement of an observable A is the average outcome of the measurements of the observable A in the limit where infinitely many measurements of A are taken. It is given by

$$\langle A \rangle = \mathrm{Tr}\,(\hat{\rho} A) = \frac{\mathrm{Tr}\,(\rho A)}{\mathrm{Tr}\,(\rho)} = \sum_i \lambda_i \langle e_i, A e_i \rangle. \tag{5.19}$$

If α is an eigenvalue of the observable A, the probability that a measurement of A yields the eigenvalue α is

$$p(A, \alpha) = \mathrm{Tr}\,(\hat{\rho} P_{A,\alpha}) \tag{5.20}$$

where $P_{A,\alpha}$ is the projector on the eigenspace of A for eigenvalue α. If $\alpha \in \sigma(A)$ is a spectral value but not an eigenvalue, and $[\alpha-\epsilon, \alpha+\epsilon] \subset \sigma(A)$ does not contain any eigenvalues, the probability that a measurement of A yields a result in $[\alpha-\epsilon, \alpha+\epsilon]$ is given by

$$p_{A,[\alpha-\epsilon,\alpha+\epsilon]} = \mathrm{Tr}\left(\hat{\rho} \cdot \int_{\alpha-\epsilon}^{\alpha+\epsilon} dP_A\right) = \mathrm{Tr}\left(\hat{\rho} \cdot P_A([\alpha-\epsilon, \alpha+\epsilon])\right) \tag{5.21}$$

where P_A is the spectral measure associated with A.

A definite prediction for the outcome of a measurement of an observable A is possible if and only if the density operator is of the form $\hat{\rho} = P_{A,\alpha}$, where $P_{A,\alpha}$ is the projector on a subspace of the eigenspace of eigenvalue $\alpha \in \sigma(A)$. In this case, the expectation value is $\langle A \rangle = \alpha$, the probability of measuring α is $p(A, \alpha) = 1$ and the *variance* vanishes $\langle (A - \langle A \rangle)^2 \rangle = 0$. More generally, if f is a polynomial with coefficients in $\mathbb{C}$, we have

$$\langle f(A) \rangle - f(\langle A \rangle) = 0$$

Exercise 43: (Expectation values)

Suppose that a quantum system is in a pure state with a normalised density operator $\hat{\rho} = P_{\boldsymbol{x}}$ that is a projector on a one-dimensional subspace $\mathrm{Span}(\{\boldsymbol{x}\})$ with $\boldsymbol{x} \in \mathcal{H}$. Show that for all observables A the expectation value $\langle A \rangle = \mathrm{Tr}\,(\hat{\rho} A)$ can be expressed in terms of the hermitian product $\langle\, ,\, \rangle$ on $\mathcal{H}$ as

$$\langle A \rangle = \langle \boldsymbol{x}, A\boldsymbol{x} \rangle.$$

Axiom 3: (Preparation and measurements)

The density operator ρ for the description of a quantum mechanical system is given by the experimental conditions under which the system was prepared. If the preparation of the system involved the measurement of an observable A with outcome $\alpha \in \sigma(A)$, where α is an eigenvalue of A, the normalised density operator after preparation satisfies

$$\mathrm{Tr}\,(\hat{\rho} P_{A,\alpha}) = 1, \tag{5.22}$$

where $P_{A,\alpha}$ is the projector on the eigenspace associated with eigenvalue α of A. If the system before preparation is described by a density operator ρ_1, the density operator after preparation is given by

$$\rho_2 = P_{A,\alpha} \rho_1 P_{A,\alpha} \qquad \hat{\rho}_2 = \frac{P_{A,\alpha} \hat{\rho}_1 P_{A,\alpha}}{\mathrm{Tr}\,(P_{A,\alpha} \hat{\rho}_1 P_{A,\alpha})}. \tag{5.23}$$

This postulate is also known under the name *collapse of the wave function*. After a measurement of an observable A with outcome $\alpha \in \sigma(A)$ is performed, the wave function is said to collapse to a linear combination of eigenvectors of O with eigenvalue α.

Note that the eigenspace of an observable A with eigenvalue α is not necessarily one-dimensional and therefore does not specify a unique element of $\mathcal{H}$. This corresponds to the fact that usually measurements of several observables are needed to determine the state of a physical system. A set of observables that can be measured simultaneously and whose measurements give all the information that can be gained about the state of the system is called a *complete set of commuting observables.*

Definition 5.2.6: (Complete set of commuting observables)

A complete set of commuting observables is a finite set of observables $\{O_i\}_{i=1,...n}$ which commute $[O_i, O_j] = O_i \circ O_j - O_j \circ O_i = 0 \; \forall i, j \in \{1, ..., n\}$ and are such that the intersection $\cap_{i=1}^n E_{\alpha_i}$ of the eigenspaces $E^i_{\alpha_i}$ of observable O_i with eigenvalue α_i is a one-dimensional subspace of $\mathcal{H}$. In other words, there exists an $\boldsymbol{x} \in \mathcal{H}$ such that

$$\bigcap_{i=1}^{n} E^i_{\alpha_i} = \text{Span}(\boldsymbol{x}).$$

Example 5.2.7: For the hydrogen atom (when the spins of electron and proton are neglected), a complete set of commuting observables is given by the Hamiltonian H, the total angular momentum J^2 and the angular momentum L_z in the z-direction.

Remark 5.2.8: The requirement that the set is finite is motivated by the fact that otherwise infinitely many measurements would be needed to determine the physical state of the system, which would be impossible in practice. However, it is clear that this requirement of finiteness must be dropped in quantum field theory.

There is no mathematical proof that a complete set of observables exists for all quantum mechanical system satisfying the axioms of quantum mechanics. However, a quantum mechanical system without a complete set of commuting observables would be difficult to interpret. In particular, it would not be possible to prepare such a system in a way that it is in a pure state by performing a finite number of measurements.

Axiom 4: (Time development)

The time development of a quantum mechanical system is described by a unitary operator U_t with continuous time dependence, the *time evolution operator*, which is determined by the differential equation

$$\dot{U}_t = \tfrac{1}{i\hbar} H_t U_t \qquad U_0 = \text{id}_{\mathcal{H}}. \tag{5.24}$$

where H_t is the Hamiltonian at time t, which gives the energy of the system. The time evolution of the physical state is given by

$$\langle A \rangle_t = \text{Tr}\left(U_t \hat{\rho} U_t^\dagger A\right). \tag{5.25}$$

Note that this time development of the system can be interpreted in two ways. The first interpretation is that of a time development of the density operator, while the observables remain constant and undergo no time evolution.

$$\langle A \rangle_t = \text{Tr}\left(\hat{\rho}_t A\right) \qquad \rho_t = U_t \rho U_t^\dagger. \tag{5.26}$$

Note that as U is unitary, $\text{Tr}\left(U_t \rho U_t^\dagger\right) = \text{Tr}\left(\rho\right)$ and therefore $\hat{\rho}_t = U_t \hat{\rho} U_t^\dagger$. Unitary operators preserve probability. In view of expression (5.16) for the density operator, this can be interpreted as a time development of each projector $P_i \to U_t P_i U_t^\dagger$ and hence as a time development of the vectors in the Hilbert space $\boldsymbol{x} \mapsto U_t \boldsymbol{x}$ for all $\boldsymbol{x} \in \mathcal{H}$. This is the *Schrödinger picture* of time development.

The second interpretation of formula (5.26) is due to the fact that the cyclic invariance of the trace, one can rewrite (5.26) as

$$\langle A\rangle_t = \mathrm{Tr}\left(PU_t^\dagger A U_t\right) = \mathrm{Tr}\,(\rho A_t) \qquad A_t = U_t^\dagger A U_t. \tag{5.27}$$

In this viewpoint, the density operators (and therefore the elements of $\mathcal{H}$) remain constant and undergo no time development. Instead, it is the observables of the system which develop with time. This is called the *Heisenberg picture* of time development.

Exercise 44: (Time development)

Consider a quantum system with normalised density operator $\hat{\rho}$, time evolution operator U_t and Hamiltonian H_t. Let A be an observable. Show that the time development of the expectation value $\langle A\rangle_t$ is given by

$$\frac{d}{dt}\langle A\rangle_t = \tfrac{1}{i\hbar}\langle [A, H_t]\rangle_t.$$

Consider now a quantum system that is given by the Heisenberg algebra and equipped with a Hamilton operator of the form

$$H_t = \frac{P^2}{2m} + V_t(X),$$

where P is the momentum operator, X the position operator and $V_t(X)$ a tie-dependent function of X. Suppose that X and P satisfy the commutation relations of the Heisenberg algebra $[X, P] = i\hbar\, 1$. Show that the time development of the expectation values $\langle X\rangle_t$, $\langle P\rangle_t$ is given by

$$\frac{d}{dt}\langle X\rangle_t = \tfrac{1}{m}\langle P\rangle_t$$
$$\frac{d}{dt}\langle P\rangle_t = -\langle V_t'(X)\rangle_t \quad \text{where} \quad V_t'(x) = \frac{d}{dx}V_t(x).$$

Show that if we interpret the derivative $F_t(x) := -V_t'(x) = -\frac{d}{dx}V_t(x)$ as the force arising from the time-dependent potential V_t, the expectation value of X satisfies

$$m\frac{d^2}{dt^2}\langle X\rangle_t = \langle F(X)\rangle_t.$$

The "acceleration" of the expectation value of the position is therefore given by the expectation value of the force. This statement is known under the name of *Ehrenfest's theorem.*

5.3 Quantum mechanics - the C^*-algebra viewpoint

The postulates of quantum mechanics introduced in the last section give a clear framework for the formulation and interpretation of quantum theories. However, it has the disadvantage that it is based on the notion of a *Hilbert space.* Hilbert spaces have no counterparts in the classical theory and cannot be directly accessed by measurements. Measurements can only be associated with *observables* of the theory. Only the observables have counterparts in the classical theory and a clear physical interpretation. Moreover, the notion of states in a Hilbert space becomes problematic in quantum field theory in curved spacetimes.

This raises the question if it is possible to give a formulation of quantum mechanics based on the *algebra of observables* alone, in which Hilbert spaces either do not appear at all or are *derived* from the algebra of observables. The answer to this question is positive and given by the theory of C^*-algebras.

Definition 5.3.1: (Banach algebra, C^*-algebra)

A (unital) *Banach algebra* is a (unital) associative algebra A that as a vector space has the structure of a Banach space with norm $|| \; ||$ and satisfies

1. $||\boldsymbol{x}\cdot\boldsymbol{y}|| \leq ||\boldsymbol{x}||\cdot||\boldsymbol{y}||$ for all $\boldsymbol{x},\boldsymbol{y}\in A$
2. $||e|| = 1$.

A (unital) C^*-*algebra* is a (unital) Banach algebra with an involution $*: A\to A$, $*^2 = 1$ that

1. is an anti-linear anti-algebra homomorphism
$$(\boldsymbol{x}\boldsymbol{y})^* = \boldsymbol{y}^*\boldsymbol{x}^* \qquad (t\boldsymbol{x}+s\boldsymbol{y})^* = \bar{t}\boldsymbol{x}^* + \bar{s}\boldsymbol{y}^* \qquad \forall \boldsymbol{x},\boldsymbol{y}\in A, t,s\in\mathbb{C}. \tag{5.28}$$
2. satisfies $||\boldsymbol{x}\boldsymbol{x}^*|| = ||\boldsymbol{x}||^2$ for all $\boldsymbol{x}\in A$.

A $*$-subalgebra of a C^*-algebra A is a subalgebra $B\subset A$ which is closed with respect to the norm $|| \; ||$ and satisfies $\boldsymbol{x}^*\in B$ for all $\boldsymbol{x}\in B$.

Example 5.3.2: (C^*-algebras)

1. The space $\mathcal{B}(\mathcal{H})$ of bounded linear operators on a Hilbert space $\mathcal{H}$ with norm $||A|| = \sup\{||A\boldsymbol{x}|| \mid \boldsymbol{x}\in\mathcal{H}, ||\boldsymbol{x}|| = 1\}$ is a Banach algebra. With the involution $*: O\mapsto O^\dagger$ it becomes a C^*-algebra.
2. Hilbert Schmidt operators form a $C^* -$ *algebra without unit*, where the norm is the Hilbert Schmidt norm and the involution is the adjoint $O^* = O^\dagger$
3. The space $L^1(\mathbb{R}^n)$ with pointwise addition, multiplication by $\mathbb{C}$, norm $||\cdot||_1$ and the convolution as product is a Banach algebra.

Exercise 45: Show that the properties of the involution in a C^*-algebra A imply
$$||e|| = 1 \qquad ||\boldsymbol{x}^*|| = ||\boldsymbol{x}|| \qquad ||\boldsymbol{x}\cdot\boldsymbol{y}|| \leq ||\boldsymbol{x}||||\boldsymbol{y}|| \qquad \forall \boldsymbol{x},\boldsymbol{y}\in A. \tag{5.29}$$

Definition 5.3.3: (Self adjoint, normal and unitary elements of a C^*-algebra)

An element $\boldsymbol{x}\in A$ of a C^*-algebra A is called *hermitian* or *self adjoint* if $\boldsymbol{x}^* = \boldsymbol{x}$. It is called *normal* if $\boldsymbol{x}^*\boldsymbol{x} = \boldsymbol{x}\boldsymbol{x}^*$. It is called *unitary* if $\boldsymbol{x}\boldsymbol{x}^* = e$.

Exercise 46: Show that for all elements $\boldsymbol{x}\in A$ of a C^*-algebra A, the elements $\frac{1}{2}(\boldsymbol{x}+\boldsymbol{x}^*)$ and $\frac{1}{2i}(\boldsymbol{x}-\boldsymbol{x}^*)$ are self-adjoint. Show that the unitary elements of A form a group. Show that an element $\boldsymbol{x}\in A$ is normal if and only if $\frac{1}{2}(\boldsymbol{x}+\boldsymbol{x}^*)$, $\frac{1}{2i}(\boldsymbol{x}-\boldsymbol{x}^*)$ commute.

Definition 5.3.4: (Spectrum, regular value, spectral values)

Let A be a Banach algebra and $\boldsymbol{x}\in A$. A complex number $z\in\mathbb{C}$ is a *regular value* of a if $\boldsymbol{x} - z\cdot e$ has an inverse. Otherwise it is called a *spectral value*. The *spectrum* $\sigma(\boldsymbol{x})$ is the set of spectral values of elements $\boldsymbol{x}\in A$.

Example 5.3.5: If $A = B(\mathcal{H})$ is the set of bounded linear operators on a Hilbert space $\mathcal{H}$, this definition coincides with Definition 5.1.23.

Definition 5.3.6: (State)

A *positive linear form* of a unital C^*-algebra A is a linear form $\psi \in A^*$ such that $\psi(\boldsymbol{x}\boldsymbol{x}^*) \geq 0$ for all $\boldsymbol{x} \in A$. A *state* on A is a positive linear form on A that satisfies $\psi(e) = e$.

Exercise 47: Show that the set of states of a unital C^*-algebra A is *convex*, i.e. if ψ_1, ψ_2 are states of A, $\psi_t = t\psi_1 + (1-t)\psi_2$ is a state of A for all $t \in [0,1]$.

Definition 5.3.7: (Pure state)

A state ψ of a unital C^*-algebra A is *pure* if $\psi = t\phi + (1-t)\eta$ with $t \in [0,1]$ and states ϕ, η implies $t = 1$ and $\psi = \phi$ or $t01$ and $\psi = \eta$.

Theorem 5.3.8: (Krein Mil'man)

Pure states exist for any unital C^*-algebra A and the set of states of A is the *closed convex hull* of its pure states. In other words: Any state ψ of A can be expressed as

$$\psi = \sum_{i=1}^{n} \lambda_i \psi_i \qquad \lambda_i \in [0,1],\ \sum_{i=1}^{n} \lambda_i = 1, \tag{5.30}$$

where ψ_i, $i \in \{1,..,n\}$ are pure states. Pure states are characterised by $\lambda_j = 1$ for a single $j \in \{1,...,n\}$ and $\lambda_i = 0$ for all $i \neq j$.

Definition 5.3.9: (Complete set of commuting observables)

A set of commuting observables of a a unital C^*-algebra A is a finite set of commuting self-adjoint elements $\boldsymbol{x}_1, ..., \boldsymbol{x}_n \in A$, $\boldsymbol{x}_i^* = \boldsymbol{x}_i$, $\boldsymbol{x}_i\boldsymbol{x}_j = \boldsymbol{x}_j\boldsymbol{x}_i$ for all $i, j \in \{1, ..., n\}$. A *complete set of commuting observables* of A is a finite set of commuting observables of A such that there exists a set of pure states $\psi_1, ..., \psi_n$ whose convex hull is the set of states of A and which satisfy

$$\psi_i(\boldsymbol{x}_j^2 - \psi_i(\boldsymbol{x}_j)^2 e) = 0 \qquad \forall i, j \in \{1,..,n\}. \tag{5.31}$$

Remark 5.3.10: As in the standard description, the requirement that the complete set of commuting observables is finite is a condition that arises from physics. An infinite set of commuting observables would imply that an infinite number of measurements is needed to determine the physical state of the system. There is a priori no guarantee that such a set of observables exists. However, it would be difficult to give a physical interpretation to a system without a complete set of commuting observables.

Example 5.3.11: We consider the C^*-algebra $B(\mathcal{H})$ of bounded linear operators on $\mathcal{H}$. If ρ is a density operator for $B(\mathcal{H})$, then

$$\psi_\rho : A \mapsto \psi_\rho(A) = \frac{\mathrm{Tr}\,(\rho A)}{\mathrm{Tr}\,(\rho)} \tag{5.32}$$

is a state. If ρ is a projector on a one-dimensional subspace of $\mathcal{H}$, it is a *pure state*.

Proof: Exercise

Example 5.3.12: We consider the C^*-algebra of matrices $A = M(n, \mathbb{C})$ with norm $||M|| = \text{Tr}\left(M^\dagger M\right) = \sqrt{\sum_{i,j=1}^n |m_{ij}|^2}$ and involution $M^* = M^\dagger$. Then, $\psi_M : A \mapsto \text{Tr}\,(AM)\,/\text{Tr}\,(M)$ is a state for any hermitian matrix $M \in M(n, \mathbb{C})$. It is a pure state if and only if the eigenvalues $(\lambda_1, ..., \lambda_n)$ of M are such that $\lambda_i = 0$ for all $i \neq j$, $\lambda_j \in \mathbb{R} \setminus \{0\}$ for exactly one $j \in \{1, ..., n\}$.

Exercise 48: Show that for a state $\psi : A \to \mathbb{C}$, the map $\phi : A \times A \to \mathbb{C}$, $\phi : (\boldsymbol{x}, \boldsymbol{y}) \mapsto \psi(\boldsymbol{y}^*\boldsymbol{x})$ is a positive hermitian linear form on A, i.e. it satisfies

$$\phi(\boldsymbol{x}^*\boldsymbol{y}) = \overline{\phi(\boldsymbol{y}^*\boldsymbol{x})} \qquad |\phi(\boldsymbol{y}^*\boldsymbol{x})|^2 \leq |\phi(\boldsymbol{x}^*\boldsymbol{x})||\phi(\boldsymbol{y}^*\boldsymbol{y})| \qquad \phi(\boldsymbol{x}^*) = \phi(\boldsymbol{x}^*e) = \overline{\phi(\boldsymbol{x})}$$
$$|\phi(e\boldsymbol{x})|^{\leq}\phi(e)\phi(\boldsymbol{x}^*\boldsymbol{x}) \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in A.$$

Hint: consider the bilinear form $\varphi : A \times A \to \mathbb{C}$, $(\boldsymbol{x}, \boldsymbol{y}) \mapsto \psi(\boldsymbol{x} + \boldsymbol{y}, \boldsymbol{x} + \boldsymbol{y})$ and show that it takes values in $\mathbb{R}$. Show also that it satisfies $\varphi(\boldsymbol{x}\boldsymbol{y}, \boldsymbol{z}) = \varphi(\boldsymbol{y}, \boldsymbol{x}^*\boldsymbol{z})$ for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in A$.

Remark 5.3.13: By comparing with formula (5.19), we see that the choice of a density operator on $B(\mathcal{H})$ gives rise to a state on the C^*-algebra $B(\mathcal{H})$. A *state in the sense of this section* is a linear functional on a C^*-algebra of quantum observables that assigns to every observable its expectation value. This expectation value is given by the density operator.

Hence, we can define expectation values *without making use of Hilbert spaces*. They are given as positive linear functionals on the C^*-algebra which contains the observables. The definition is elegant, but also worrying. It raises the question, if such expectation values automatically give rise to representations on Hilbert spaces or if, instead, there is a whole set of quantum systems which does not have a Hilbert space and which escaped the original formulation. We will see in the following that the answer to this is the first alternative and the two formulations of quantum mechanics are equivalent. For this, we introduce the concept of a representation of a C^*-algebra.

Definition 5.3.14: (Representation of C^*-algebra)

A *representation* of a (unital) C^*-algebra A is an algebra homomorphism $\pi : A \to B(\mathcal{H})$ into the the set of bounded linear operators on a Hilbert space $\mathcal{H}$ which is continuous with respect to the norm $||\ \ ||$ on A and the norm $||O|| = \sup\{||O\boldsymbol{x}|| \mid \boldsymbol{x} \in \mathcal{H}, ||\boldsymbol{x}|| = 1\}$ on $B(\mathcal{H})$ such that $\pi(\boldsymbol{x}^*) = \pi(\boldsymbol{x})^\dagger$ for all $\boldsymbol{x} \in A$ (and $\pi(e) = e$).

It can be shown that every state on a C^*-algebra A gives rise to a representation of A on a Hilbert space. This is called the GNS construction.

Definition 5.3.15: (GNS pair)

Let ψ be a positive linear functional on a C^*-algebra A. A *GNS pair* for ψ is a pair $(\pi, \boldsymbol{x})$, where π is a representation of A on a Hilbert space $\mathcal{H}$ and $\boldsymbol{x} \in \mathcal{H}$ a vector such that

- $\boldsymbol{x}$ is *cyclic*: $\overline{\{\pi(\boldsymbol{a})\boldsymbol{x} \mid \boldsymbol{a} \in A\}} = \mathcal{H}$
- $\psi(\boldsymbol{a}) = \langle \boldsymbol{x} | \pi(\boldsymbol{a})\boldsymbol{x} \rangle$ for all $\boldsymbol{a} \in A$.

Two GNS pairs $(\pi, \boldsymbol{x})$, $(\pi', \boldsymbol{x}')$ are called *equivalent* if there exists a unitary operator $U : \mathcal{H} \to \mathcal{H}'$ such that $\boldsymbol{x}' = U\boldsymbol{x}$ and $U \circ \pi(\boldsymbol{a}) = \pi'(\boldsymbol{a}) \circ U$ for all $\boldsymbol{a} \in A$. In other words: Two GNS pairs are equivalent if they are related by a unitary intertwiner between the associated representations.

Theorem 5.3.16: Every positive linear functional ψ on a unital C^*-algebra A has a GNS pair $(\pi, \boldsymbol{x})$ and any two GNS pairs for ψ are equivalent. A positive linear functional on A is a pure state if and only if the representation π is irreducible.

Proof: see p123 ff, p129 in Arveson: A Short Course on spectral theory. We only sketch the idea of the proof.

As the C^*-algebra A is the only vector space available in this definition, it seems plausible to construct the Hilbert space $\mathcal{H}$ from A. More precisely, one defines the Hilbert space $\mathcal{H}$ as the quotient A/N_ψ where $N_\psi = \{\boldsymbol{a} \in A \mid \psi(\boldsymbol{a}^*\boldsymbol{a}) = 0\}$. If we denote by $[\boldsymbol{a}] = \boldsymbol{a} + N_\psi$ the equivalence class $[\boldsymbol{a}] = \{\boldsymbol{a} + \boldsymbol{y} \mid \boldsymbol{y} \in N_\psi\}$, we can define the scalar product on $\mathcal{H}$ by setting $\langle \boldsymbol{a} + N_\psi, \boldsymbol{y} + N_\psi\rangle = \psi(\boldsymbol{y}^*\boldsymbol{a})$, $\boldsymbol{a}, \boldsymbol{y} \in A$. The cyclic vector $\boldsymbol{x}$ is the equivalence class $\boldsymbol{x} = e + N$. The representation is given by $\pi(\boldsymbol{a})(\boldsymbol{y} + N_\psi) = \boldsymbol{a} \cdot \boldsymbol{y} + N_\psi$. The rest of the proof consists of checking that these objects are well-defined and have the required properties. □

Thus, we find that every state on a C^*-algebra gives rise to a representation of A on a Hilbert space such that the state is realised as an expectation value in this representation. To show that the formulation of quantum mechanics in terms of C^*-algebras is indeed equivalent to the traditional formulation, it remains to show that states on A *exist*. This is the content of the theorem by Gelfand-Naimark. For a proof see for example section 4.8. in Arveson: A Short Course on spectral theory.

Theorem 5.3.17: (Gelfand-Naimark)

For every element $\boldsymbol{a} \in A$, there is a state $\psi_a : A \to \mathbb{C}$ such that $\psi_a(\boldsymbol{a}^*\boldsymbol{a}) = ||\boldsymbol{a}||^2$. Every C^*-algebra A can be represented as the algebra of bounded linear operators $B(\mathcal{H})$ on a Hilbert space $\mathcal{H}$ such that for any state $\psi : A \to \mathbb{C}$, $\langle \pi(\boldsymbol{a})\boldsymbol{x}|\boldsymbol{x}\rangle = \psi(\boldsymbol{a})$ and $\pi(\boldsymbol{a}^*)\boldsymbol{x} = \pi(\boldsymbol{a})^\dagger \boldsymbol{x}$ for all $\boldsymbol{x} \in \mathcal{H}$.

Remark 5.3.18: Note that this theorem does not assert that $\mathcal{H}$ is a *separable* Hilbert space. It can be shown, however, that for a C^*-algebra which is generated as a C^*-algebra by a finite or countably infinite set of elements, the Hilbert space is separable.

This shows that the two formulations of quantum mechanics - the one based on Hilbert spaces and the one based on C^*-algebras - are indeed equivalent. The concept of a Hilbert space is therefore not necessary as a starting point for the the formulation of a quantum theory. It is possible to formulate the quantum theory based on the observables. In this formulation, physical states appear as positive linear functionals of the observables and Hilbert spaces are obtained from their representations.

5.4 Canonical Quantisation

After investigating the formulation of quantum mechanics from two different viewpoints, we will now look at the *construction* of quantum theories from classical theories. It should be noted that this construction - usually referred to as *quantisation* is in general neither an algorithm nor an existence theorem. While it can often be shown that quantisations with the required physical properties exist in *specific cases*, this existence is not guaranteed a priori or in general. Moreover, the question of what constitutes an admissible quantum theory for a given classical theory is subtle. We will see in the following that even in the simplest possible case - the two-dimensional Heisenberg algebra, which describes the motion of a particle in one-dimensional space - the construction of a quantum theory and its existence is not as straightforward as it seems.

We start our investigation with the classical theory. The classical observables which parametrise the phase space form a Poisson algebra.

Definition 5.4.1: A Poisson algebra is an associative unital algebra A together with a bilinear map $\{\,,\,\}: A\times A\to A$, the *Poisson bracket* which

1. is *antisymmetric*: $\{f,g\} = -\{g,f\}$ for all $f,g\in A$
2. satisfies the *Leibnitz identity*: $\{f\cdot g,h\} = f\{g,h\} + g\{f,h\}$ for all $f,g,h\in A$
3. satisfies the *Jacobi identity*: $\{\{f,g\},h\} + \{\{h,f\},g\} + \{\{g,h\},f\} = 0\ \forall f,g,h\in A$.

Remark 5.4.2: Any Poisson algebra $(A,+,\cdot,\{\,,\,\})$ is also a Lie algebra with Lie bracket $\{\,,\,\}$ as well as a commutative associative algebra.

Example 5.4.3: For an open subset $U\subset\mathbb{R}^n$, the vector space $\mathcal{C}^\infty(U)$ of smooth functions with pointwise multiplication is a commutative associative algebra. If $U\subset\mathbb{R}^n$ is equipped with a symplectic form $\omega\in\Omega^2(U)$, the vector space $\mathcal{C}^\infty(U)$ has the structure of a Poisson algebra (see Lemma 1.4.25).

Example 5.4.4: For any submanifold $U\subset\mathbb{R}^n$, the cotangent bundle T^*U has the structure of a Poisson algebra. In terms of coordinate functions q_i, $i=1,...,\dim(U)$ on U and coordinates $p^i: T_qU\to\mathbb{R}$, $i=1,...,\dim(U)$ which are the one-forms dual to the basis vector fields $\partial_i: q\in U\to T_qU$, $p^i(\partial_j)=\delta_{ij}$, the Poisson structure takes the form

$$\{p^i,q_j\}=\delta_{ij}\qquad \{q_i,q_j\}=\{p^i,p^j\}=0\qquad i,j\in\{1,...,\dim(U)\}.$$

The coordinates q_i are called *positions*, the coordinates p^i conjugate *momenta*.

It can be shown that any Poisson structure on a submanifold $U\subset\mathbb{R}^n$ of even dimension which is given by a symplectic form looks like the Heisenberg algebra locally. This is the content of Darboux's theorem.

Lemma 5.4.5: (Darboux's theorem)

Let $\omega\in\Omega^2(U)$ be a symplectic two-form on a submanifold $U\subset\mathbb{R}^n$ of even dimension $\dim(U)=2n$. By Poincaré's lemma, for any star-shaped neighbourhood $V\subset U$ there exists a one-form $\theta\in\Omega^1(V)$, the *symplectic potential*, with $\omega=d\theta$. Then, there exists a set of coordinates $q_i,p^i: V\to\mathbb{R}$ such that

$$\theta=\sum_{i=1} q_i\, dp^i,$$

and the Poisson structure on V takes the form

$$\{p^i,q_j\}=\delta_{ij}\qquad \{q_i,q_j\}=\{p^i,p^j\}=0\qquad i,j\in\{1,...,\dim(U)\}.$$

Such coordinates are called *Darboux coordinates*. The coordinates q_i are called *positions*, the coordinates p^i conjugate *momenta*.

Remark 5.4.6: Darboux's theorem asserts that locally, all Poisson structures that are given by symplectic forms look alike. This is a consequence of the *antisymmetry* of the Poisson bracket or, equivalently, the antisymmetry of the symplectic form. In contrast, a *symmetric* positive definite bilinear form on the tangent bundle TU of a submanifold $U \subset \mathbb{R}^n$ corresponds to a Riemannian metric. It can be shown that such metrics can be brought into a standard form given by the Riemannian coordinates. However, this does not lead to an analogue of Darboux's theorem since the second order derivatives and higher order derivatives of the metric do not vanish.

Darboux's theorem provides a strong motivation for the study of the *Heisenberg* Poisson algebra.

Definition 5.4.7: (Heisenberg algebra)

The Heisenberg algebra H^n is the Poisson algebra $\mathcal{C}^\infty(T^*\mathbb{R}^n)$ with coordinates q_i on $\mathbb{R}^n$ and conjugate coordinates p^i on $T_q\mathbb{R}^n \cong \mathbb{R}^n$ and Poisson bracket

$$\{p_i, p_j\} = \{q_i, q_j\} = 0 \quad \forall i, j = 1, ..., n \qquad \{p_j, q_i\} = \delta_{ij}. \tag{5.33}$$

The algebra $\mathrm{Pol}(p^i, q_i)$ of polynomials in the coordinates p^i, q_j is a Poisson subalgebra of H^n, i.e. it is a subalgebra with respect to the associative algebra structure and $\{\mathrm{Pol}(p^i, q_i), \mathrm{Pol}(p^i, q_i)\} \subset \mathrm{Pol}(p^i, q_i)$.

We will now investigate the *canonical quantisation* of the Heisenberg algebra, which - at least locally - tells us something about the quantisation of physical phase spaces. It should be noted that canonical quantisation is *not* an algorithm or a theorem. Rather, it has the status of an idea or a principle. As we will see in the following, one needs to be precise in defining what *quantisation* means as well as in its implementation, and there is no guarantee that a mathematical object with the desired properties exists.

The basic idea or principle of canonical quantisation is the following: A Poisson algebra combines the structure of a (commutative, unital) algebra with the structure of a Lie algebra. The compatibility of these structures is given by the Leibnitz identity. In canonical quantisation this Poisson algebra is to be replaced by an (non-commutative) *associative* algebra such that *both*, the multiplicative structure and the Lie algebra structure of the Poisson algebra are obtained from its multiplication (the latter via the commutator).

In physics, this principle is often states as the requirement that there should be a one-to-one correspondence between multiplicative generators of these algebras (classical and quantum observables) and a relation between the Poisson bracket of the Poisson algebra and the commutator of the associative algebra

$$\{\,,\,\} \to \tfrac{1}{i\hbar}[\,,\,]$$

Moreover, in order to obtain self-adjoint and unitary operators, one demands that the associative algebra has the structure of a C^*-algebra and is represented on a Hilbert space. This is referred to as canonical quantisation. To make this notion precise, we use the following definition.

Definition 5.4.8: (Perfect quantisation)

A *perfect* quantisation of a commutative Poisson algebra A is a vector space homomorphism $Q : A \to \hat{A}$ into an associative, unital C^*-algebra $\hat{A} \subset B(\mathcal{H})$ of operators on a Hilbert space $\mathcal{H}$ such that:

1. Q is a Lie algebra homomorphism between the Lie algebras $(A, \{\,,\,\})$ and $(\hat{A}, \{\,,\,\}_\hbar)$, where the Lie bracket on $\hat{A}$ is given by $\{\boldsymbol{x}, \boldsymbol{y}\}_\hbar = \frac{1}{i\hbar}[\boldsymbol{x}, \boldsymbol{y}] = \frac{1}{i\hbar}(\boldsymbol{x} \circ \boldsymbol{y} - \boldsymbol{y} \circ \boldsymbol{x})$. In other words:
$$Q(\{f, g\}) = \tfrac{1}{i\hbar}[Q(f), Q(g)] = \tfrac{1}{i\hbar}(Q(f) \circ Q(g) - Q(g) \circ Q(f)) \qquad \forall f, g \in A. \tag{5.34}$$
2. The unit of the Poisson algebra is mapped to the unit of the C^*-algebra: $Q(1) = e$.
3. Q induces an irreducible representation of A via the inclusion $Q(A) \subset B(\mathcal{H})$.

In the following, we will often omit the map Q and write $\hat{f}$ instead of $Q(f)$ for all $f \in A$.

Remark 5.4.9: It is often useful to modify this definition to allow also unbounded operators with domains that are dense in $\mathcal{H}$.

We will now show that a perfect quantisation does not exist for the Heisenberg Poisson algebra. We consider the simplest case, namely the Heisenberg algebra H_1 of dimension two. As a first step, we obtain the following lemma.

Lemma 5.4.10: In any perfect quantisation Q of the Heisenberg algebra H_1, one has
$$\widehat{q^2} = \hat{q}^2 \qquad \widehat{p^2} = \hat{p}^2 \qquad \widehat{qp} = \tfrac{1}{2}(\hat{p}\hat{q} + \hat{q}\hat{p}). \tag{5.35}$$

Proof:

As $\{q^2, q\} = 0$, we have $[\widehat{q^2}, \hat{q}] = 0$. Condition (5.34) together with the condition that the representation on the Hilbert space $\mathcal{H}$ is irreducible implies that the operator $\widehat{q^2} \in B(\mathcal{H})$ is a function $A(\hat{q})$, which depends only on $\hat{q}$ but not on $\hat{p}$. As $\{p, q^2\} = -2q$, we have
$$[\hat{p}, A(\hat{q})] = -i\hbar\dot{A}(\hat{q}) = -2i\hbar\hat{q}.$$
This implies that $\widehat{q^2} = A(\hat{q})$ is of the form $\widehat{q^2} = \hat{q}^2 - 2\mathfrak{e}_-$, where $\mathfrak{e}_-$ is a hermitian operator which does not depend on $\hat{q}, \hat{p}$.

An analogous argument for $p, \hat{p}$ implies $\widehat{p^2} = \hat{p}^2 + 2\mathfrak{e}_+$, where $\mathfrak{e}_+$ is a hermitian operator which does not depend on $\hat{p}, \hat{q}$. The identity $\{q^2, p^2\} = 4pq$ implies
$$4i\hbar\widehat{pq} = [\widehat{q^2}, \widehat{p^2}] = [\hat{q}^2, \hat{p}^2] - 4[\mathfrak{e}_-, \mathfrak{e}_+] = 2i\hbar(\hat{q}\hat{p} + \hat{p}\hat{q}) - 4i\hbar\mathfrak{h},$$
where $i\hbar\mathfrak{h} = [\mathfrak{e}_-, \mathfrak{e}_+]$ is a hermitian operator which does not depend on $\hat{p}$, $\hat{q}$. One can then show that the operators $\mathfrak{e}_\pm$, $\mathfrak{h}$ satisfy
$$[\mathfrak{e}_-, \mathfrak{e}_+] = i\hbar\,\mathfrak{h} \qquad [\mathfrak{h}, \mathfrak{e}_\pm] = \pm 2i\hbar\,\mathfrak{e}_\pm.$$
This is the Lie bracket of $\mathfrak{sl}(2, \mathbb{C})$ that we encountered in the section on Lie algebras. However, there we found that the corresponding operators (matrices) were *anti-hermitian*. One can show that $\mathfrak{sl}(2, \mathbb{C})$ does not have a non-trivial representation in which these commutation relations hold for *hermitian* operators. Hence, the operators $\mathfrak{e}_\pm$, $\mathfrak{h}$ must be trivial, and we have
$$\widehat{q^2} = \hat{q}^2 \qquad \widehat{p^2} = \hat{p}^2 \qquad \widehat{qp} = \tfrac{1}{2}(\hat{p}\hat{q} + \hat{q}\hat{p}).$$
□

This lemma determines the quantisation map Q for polynomials of order one and two in the variables p, q. We can now use the correspondence between commutators and Poisson brackets to determine the quantisation map for polynomials of higher order. However, when doing this one obtains contradictions and finds that it is not possible to extend the perfect quantisation in this way. This is known as the *Gronewold van Hove paradox*.

Theorem 5.4.11: (Gronewold van Hove paradox)

A perfect quantisation does not exist for any Lie subalgebra of the Heisenberg algebra H^1 which contains polynomials in p and q of degree greater than two.

Proof:

Using the correspondence between Poisson bracket and commutator, we show that

$$\widehat{q^3} = \hat{q}^3 \qquad \widehat{p^3} = \hat{p}^3 \qquad \widehat{q^2p} = \tfrac{1}{2}(\hat{q}^2\hat{p} + \hat{p}\hat{q}^2) \qquad \widehat{qp^2} = \tfrac{1}{2}(\hat{p}^2\hat{q} + \hat{q}\hat{p}^2)$$

By induction, it then follows that for any real polynomial P

$$\widehat{P(q)} = P(\hat{q}) \qquad \widehat{P(p)} = P(\hat{p}) \qquad \widehat{P(q)p} = \tfrac{1}{2}(P(\hat{q})\hat{p} + \hat{p}P(\hat{q})) \qquad \widehat{P(p)q} = \tfrac{1}{2}(P(\hat{p})\hat{q} + \hat{q}P(\hat{p})).$$

This leads to a contradiction with the Poisson bracket.

$$\{q^3, p^3\} = 3\{q^2p, p^2q\}.$$

Applying the correspondence to the left-hand side yields

$$[\hat{q}^3, \hat{p}^3] = i\hbar(9\hat{q}^2\hat{p}^2 - 18i\hbar\hat{p}\hat{q} - 6\hbar^2)$$

while the right-hand side gives

$$3[\hat{q}^2\hat{p}, \hat{p}^2\hat{q}] = i\hbar(9\hat{p}^2\hat{q}^2 - 18i\hbar\hat{p}\hat{q} - 3\hbar^2)$$

This contradicts the identity $Q(\{f,g\}) = \frac{1}{i\hbar}(Q(g) \circ Q(g) = Q(g) \circ Q(f))$. □

Exercise 49: 1. Use the correspondence principle $\widehat{\{f,g\}} = \frac{1}{i\hbar}[\hat{f}, \hat{g}]$ and the Poisson bracket of the Heisenberg Poisson algebra

$$\{p, p\} = \{q, q\} = 0 \qquad \{p, q\} = 1.$$

to show that the quadratic relations

$$\widehat{q^2} = \hat{q}^2 \qquad \widehat{p^2} = \hat{p}^2 \qquad \widehat{qp} = \tfrac{1}{2}(\hat{p}\hat{q} + \hat{q}\hat{p}).$$

imply

$$\widehat{P(q)} = P(\hat{q}) \qquad \widehat{P(p)} = P(\hat{p}) \qquad \widehat{P(q)p} = \tfrac{1}{2}(P(\hat{q})\hat{p} + \hat{p}P(\hat{q})) \qquad \widehat{P(p)q} = \tfrac{1}{2}(P(\hat{p})\hat{q} + \hat{q}P(\hat{p})).$$

for all polynomials P with real coefficients.

2. Determine the commutators $[\hat{q}^3, \hat{p}^3]$, $[\hat{q}^2\hat{p}, \hat{p}^2\hat{q}]$ and the associated Poisson brackets $\{q^3, p^3\}$, $\{q^2p, p^2q\}$. Show that this leads to a contradiction with the correspondence principle.

Remark 5.4.12: Roughly speaking, what goes wrong when one tries to construct a perfect quantisation that contains higher order polynomials in the variables p, q is that each identity of the form $\{f(p,q), g(p,q)\} = h(p,q)$, where f, g, h are polynomials in p and q corresponds to *several* identities in the associative algebra. These identities are related by commuting the variables $\hat{q}$, $\hat{p}$. As the commutator of $\hat{p}$ and $\hat{q}$ is not trivial, this leads to a contradiction.

The essence of the Gronewold van Hove paradox is that demanding the existence of a perfect quantisation is too rigid for most physical systems. Instead of imposing that the identity

$$Q(\{f,g\}) = \tfrac{1}{i\hbar}[Q(f), Q(g)] = \tfrac{1}{i\hbar}(Q(f) \circ Q(g) - Q(g) \circ Q(f)) \tag{5.36}$$

holds for *all* $f, g \in A$, one usually only demands that this condition holds only for a linear subspace $A' \subset A$, which forms a Lie subalgebra of A and which is such that elements of A' *generate* A *multiplicatively.* In other words: $\{A', A'\} \subset A'$ and polynomials in A' should be dense in A.

Definition 5.4.13: (Quantisation)

An (imperfect) quantisation of a commutative Poisson algebra A is a vector space homomorphism $Q : A \to \hat{A}$ into an associative, unital C^*-algebra $\hat{A} \subset B(\mathcal{H})$ of operators on a Hilbert space $\mathcal{H}$ such that:

1. The unit of the Poisson algebra is mapped to the unit of the C^*-algebra: $Q(1) = e$.
2. Q induces an irreducible representation of A via $Q(A) \subset B(\mathcal{H})$.
3. There exists a linear subspace $A' \subset A$ which is a Lie subalgebra with respect to the Poisson bracket $\{A', A'\} \subset A'$ and is such that the polynomials $\mathrm{Pol}(B_{A'})$ in the elements of a Basis $B_{A'}$ of A' are dense in A. The restriction of quantisation map to A', $Q|_{A'} : A' \to Q(A')$, is a Lie algebra isomorphism from A' to a Lie subalgebra $Q(A') \subset \hat{A}$

$$Q(\{f,g\}) = \tfrac{1}{i\hbar}[Q(f), Q(g)] = \tfrac{1}{i\hbar}(Q(f) \circ Q(g) - Q(g) \circ Q(f)) \qquad \forall f, g \in A'. \tag{5.37}$$

Remark 5.4.14: If the correspondence requirement is weakened in this way, quantisations can be shown to exist for most physical systems. Note however, that for higher order polynomials in the elements of the basis $B_{A'}$, the correspondence condition only holds in a weakened form

$$[Q(f), Q(g)] = i\hbar Q(\{f,g\}) + O(\hbar^2) \tag{5.38}$$

where $O(\hbar^2)$ denotes linear combinations involving elements of $\hat{A}$ which are preceded by factors $\hbar^k$ with $k \geq 2$.