Mathematical Physics I and II
Mathematics Part

Melchior Wirth

February 27, 2025

Contents

Chapter 1

Local and Global Analysis

1.1 Fundamentals

1.1.1 Linear Theory

Definition 1.1.1. Let V be a vector space over $\mathbb{K}$. A finite family $(e_i)_{i\in I}$ in V is called a *finite basis* of V if for every $v \in V$ there exists a unique family $(\lambda^i)_{i\in I}$ in $\mathbb{K}$ such that

$$v = \sum_{i\in I} \lambda^i e_i.$$

The vector space V is called *finite-dimensional* if it has a finite basis, otherwise it is called *infinite-dimensional*. By convention, the empty family is a finite basis of the trivial vector space $\{0\}$.

Remark 1.1.2.

- We use $\mathbb{K}$ to denote a field that is either $\mathbb{R}$ or $\mathbb{C}$.
- If $(\lambda_i)_{i\in I}$ is a finite family in $\mathbb{K}$ and $(v_i)_{i\in I}$ is a finite family in V, then $\sum_{i\in I} \lambda_i v_i$ is called a *linear combination* of v_i, $i \in I$. The family $(v_i)_{i\in I}$ is called *linearly independent* if whenever $(\lambda_i)_{i\in I}$ is a finite family in $\mathbb{K}$ such that

  $$\sum_{i\in I} \lambda^i v_i = 0,$$

  then $\lambda^i = 0$ for all $i \in I$.

  An equivalent definition of a finite basis is a linearly indepenent finite family $(e_i)_{i\in I}$ such that every vector in V is a linear combination of e_i, $i \in I$.

- If $(e_i)_{i\in I}$ and $(f_j)_{j\in J}$ are finite bases of V, then I and J have the same cardinality (number of elements). This number is called the *dimension of* V and denoted by $\dim V$.

Example 1.1.3. The vector space $\mathbb{K}^n$ has the *canonical or standard basis* $(e_i)_{i=1}^n$, where $e_i = (0, \dots, 0, 1, 0, \dots, 0)$ with 1 in the i-th position.

Example 1.1.4. The vector space $\mathbb{K}[X]$ of polynomials over $\mathbb{K}$ has no finite basis. Indeed, any finite family $(p_i)_{i \in I}$ contains a polynomial of maximal degree d. Then every linear combination of p_i, $i \in I$, also has degree at most d. This mean that every polynomial with degree strictly larger than d is not a linear combination of p_i, $i \in I$.

Definition 1.1.5. If V and K are vector spaces over $\mathbb{K}$, a map $\varphi \colon V \to W$ is called *linear* if

$$\varphi(\lambda^1 v_1 + \lambda^2 v_2) = \lambda_1 \varphi(v_1) + \lambda_2 \varphi(v_2)$$

for all $\lambda^1, \lambda^2 \in \mathbb{K}$ and $v_1, v_2 \in V$.

If V and W are finite-dimensional, we write $\mathcal{L}(V, W)$ for the set of all linear maps from V to W.

Example 1.1.6. Rotations of the plane are linear maps. In contrast, translations are not linear (except for the trivial case, which is translation by 0).

Definition 1.1.7. If $(v_i)_{i=1}^n$ is a basis of V, $(w_j)_{j=1}^m$ is a basis of W and $\varphi \in \mathcal{L}(V, W)$, then the *transformation matrix* of φ with respect to the bases (v_i) and (w_j) is the matrix $A = (A^j{}_i) \in \mathbb{K}^{m \times n}$ whose entries satisfy

$$\varphi(v_i) = \sum_{j=1}^m A^j{}_i w_j.$$

If $V = W$ and $v_i = w_i$ for $1 \le i \le m$, we simply call A the transformation matrix of φ with respect to the basis $(v_i)_{i=1}^m$.

Remark 1.1.8. The transformation matrix $(A^j{}_i)$ depends not only on the linear map φ, but also on the bases (v_i) and (w_j) (and in particular on the order of the basis elements).

Example 1.1.9. The transformation matrix of rotation by $\pi/2$ (90°) with respect to the standard basis of $\mathbb{R}^2$ is given by $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

Lemma 1.1.10. *Let V and W be vector spaces over $\mathbb{K}$ with bases $(v_j)_{j=1}^m$, $(w_k)_{k=1}^n$. If $\varphi \in \mathcal{L}(V, W)$, write A_φ for the transformation matrix of φ with respect to the bases (v_j), (w_k).*

(a) The map $\mathcal{L}(V, W) \to \mathbb{K}^{n \times m}$, $\varphi \mapsto A_\varphi$ is a bijection.

(b) If $\varphi, \psi \in \mathcal{L}(V, W)$ and $\lambda, \mu \in \mathbb{K}$, then

$$A_{\lambda\varphi+\mu\psi} = \lambda A_\varphi + \mu A_\psi.$$

Let U be a vector space over $\mathbb{K}$ with basis $(u_i)_{i=1}^l$. If $\varphi \in \mathcal{L}(U, V)$ (resp. $\varphi \in \mathcal{L}(U, W)$), we write A_φ for the transformation matrix of φ with respect to the bases (u_i), (v_j) (resp. (u_i), (w_k)).

(c) If $\varphi \in \mathcal{L}(U, V)$, $\psi \in \mathcal{L}(V, W)$, then

$$A_{\psi\circ\varphi} = A_\psi A_\varphi.$$

Remark 1.1.11. For $\mathbb{K}^n$, we have a canonical basis (the standard basis). Thus linear maps from $\mathbb{K}^n$ to $\mathbb{K}^m$ can *canonically* be identified with $m \times n$ matrices.

Lemma 1.1.12 (Change of basis)**.** *Let V and W be vector spaces over $\mathbb{K}$, let $(v_i)_{i=1}^m$, $(\tilde{v}_j)_{j=1}^m$ be bases of V and let $(w_k)_{k=1}^n$, $(\tilde{w}_l)_{l=1}$ be bases of W. The basis change matrices $C \in \mathbb{K}^{m\times m}$ and $D \in \mathbb{K}^{n\times n}$ are the matrices with entries $C^j{}_i$ and $(D^l{}_k)$ determined by*

$$v_i = \sum_{j=1}^m C^j{}_i \tilde{v}_j,$$
$$w_k = \sum_{l=1}^n D^l{}_i \tilde{w}_l.$$

The matrices C and D are invertible. Moreover, if $\varphi \in \mathcal{L}(V, W)$, A (resp. B) denotes the transformation matrix of φ with respect to the bases (v_i), (w_k) (resp. $(\tilde{v}_j)$, $(\tilde{w}_l)$), then

$$B = DAC^{-1}.$$

Definition 1.1.13. Let $\mathrm{GL}(n, \mathbb{K}) = \{C \in \mathbb{K}^{n\times n} \mid C \text{ invertible}\}$. For $C \in \mathrm{GL}(n, \mathbb{K})$, the map

$$\mathbb{K}^{n\times n} \to \mathbb{K}^{n\times n},\ A \mapsto CAC^{-1}$$

is called *conjugation.* To matrices $A, B \in \mathbb{K}^{n\times n}$ are called *similar* if there exists $C \in \mathrm{GL}(n, \mathbb{K})$ such that $B = CAC^{-1}$.

Some operations on linear maps are easiest to define on the transformation matrices. To ensure that this does not depend on the choice of basis, the operation needs to be invariant under conjugation.

Example 1.1.14 (Trace). For $A = (A^j{}_i) \in \mathbb{K}^{n\times n}$, the trace is defined as $\mathrm{tr}(A) = \sum_{i=1}^n A^i{}_i$. The trace satisfies $\mathrm{tr}(CAC^{-1}) = \mathrm{tr}(C^{-1}CA) = \mathrm{tr}(A)$ for all $A \in \mathbb{K}^{n\times n}$, $C \in \mathrm{GL}(n, \mathbb{K})$.

Example 1.1.15 (Determinant). For $A = (A^j{}_i) \in \mathbb{K}^{n\times n}$, the determinant is defined as $\det(A) = \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^n A^{\sigma(i)}{}_i$. The determinant satisfies $\det(CAC^{-1}) = \det(C)\det(A)\det(C)^{-1} = \det(A)$.

Definition 1.1.16 (Dual space)**.** If V is a finite-dimensional vector space over $\mathbb{K}$, the *dual space* V^* of V is the vector space $\mathcal{L}(V, \mathbb{K})$.

If $(v_i)_{i=1}^n$ is a basis of V, then the dual basis $(v^i)_{i=1}^n$ is characterized by

$$v^i(v_j) = \delta^i{}_j, \quad 1 \leq i, j \leq n.$$

Remark 1.1.17. If $V = \mathbb{K}^n$, then V^* can canonically be identified with $1 \times n$ matrices or *row vectors*.

Lemma 1.1.18. *If V is a vector space over $\mathbb{K}$ with basis $(v_i)_{i=1}^n$, then the dual basis $(v^i)_{i=1}^n$ is a basis of V^*.*

Lemma 1.1.19 (Basis change for dual bases)**.** *If V is a vector space over $\mathbb{K}$ with bases $(v_i)_{i=1}^n$ and $(\tilde{v}_j)_{j=1}^n$ and C is the basis change matrix, then the basis change matrix for the dual bases (v^i) and $(\tilde{v}^j)$ is $(C^{\mathrm{T}})^{-1}$.*

Remark 1.1.20. We write V^{**} for $(V^*)^*$. There is a natural map

$$\chi\colon V \to V^{**},\ \chi(v)(\varphi) = \varphi(v).$$

If $(v_i)_{i=1}^n$ is a basis of V with dual basis $(v^i)_{i=1}^n$, then the dual basis of $(v^i)_{i=1}^n$ is given by $(\chi(v_i))_{i=1}^n$. In particular, χ is a bijection. We will use the map χ to identify V^{**} with V.

Definition 1.1.21 (Multilinear maps)**.** If $V_1, \dots, V_r$ and W are vector spaces over $\mathbb{K}$, a map

$$\varphi\colon V_1 \times \cdots \times V_r \to W$$

is called a *multinear map (of order r)* if for every $j \in \{1, \dots, r\}$ and all $v_i \in V_i$, $i \neq j$, the map

$$V_j \to W, v \mapsto \varphi(v_1, \dots, v_{j-1}, v, v_{j+1}, \dots, v_r)$$

is linear. If $W = \mathbb{K}$, we also call φ a *multilinear form* (of order r).

Remark 1.1.22. The direct product $V_1 \times \cdots \times V_r$ is again a vector space. Except for trivial cases, a linear map $\varphi\colon V_1 \times \cdots \times V_r \to W$ is not multilinear.

Example 1.1.23 (Determinant). If we view elements of $\mathbb{K}^n$ as row vectors (or column vectors), the map

$$(\mathbb{K}^n)^n \to \mathbb{K},\ (v_1, \dots, v_n) \mapsto \det((v_1\ \dots\ v_n))$$

is multilinear.

Example 1.1.24 (Inner product). If $\mathbb{K} = \mathbb{R}$, an inner product on V is a multilinear form of order 2 (or *bilinear form*).

Example 1.1.25. If $\varphi_1 \in V_1^*, \dots, \varphi_r \in V_r^*$, then the map

$$\varphi_1 \otimes \cdots \otimes \varphi_r \colon V_1 \times \cdots \times V_r \to \mathbb{K},\ (v_1, \dots, v_r) \mapsto \prod_{j=1}^{r} \varphi_j(v_j)$$

is multilinear.

Definition 1.1.26. We write $V_1^* \otimes \cdots \otimes V_r^*$ for the space of multilinear forms on $V_1 \times \cdots \times V_r$.

Lemma 1.1.27. *If $V_1, \dots, V_r$ are finite-dimensional vector spaces over $\mathbb{K}$, then $V_1^* \otimes \cdots \otimes V_r^*$ is a vector space of dimension* $\dim V_1 \cdot \ldots \cdot \dim V_r$ *over* $\mathbb{K}$.

Proof. We only consider the case $r = 2$ for ease of notation. Let $(v_i)_{i=1}^m$ be a basis of V_1 and $(w_j)_{j=1}^n$ a basis of V_2. We claim that $(v^i \otimes w^j)_{i,j}$ is a basis of $V_1^* \otimes V_2^*$.

If $\varphi \colon V_1 \times V_2 \to \mathbb{K}$ is bilinear, then

$$\begin{aligned}
\varphi\left(\sum_{i=1}^m \lambda^i v_i, \sum_{j=1}^n \mu^j w_j\right) &= \sum_{i=1}^m \sum_{j=1}^n \lambda^i \mu^j \varphi(v_i, w_j) \\
&= \sum_{i=1}^m \sum_{j=1}^n \lambda^i \mu^j \sum_{k=1}^m \sum_{l=1}^n \varphi(v_k, w_l) \delta^k{}_i \delta^l{}_j \\
&= \sum_{i=1}^m \sum_{j=1}^n \lambda^i \mu^j \sum_{k=1}^m \sum_{l=1}^n \varphi(v_k, w_l) v^k(v_i) w^l(w_i) \\
&= \sum_{k=1}^m \sum_{l=1}^n \varphi(v_k, w_l)(v^k \otimes w^l)\left(\sum_{i=1}^n \lambda^i v_i, \sum_{j=1}^m \mu^j w_j\right).
\end{aligned}$$

Hence every element of $V_1^* \otimes V_2^*$ is a linear combination of $(v^i \otimes w^j)_{i,j}$. Moreover, if $\lambda_{ij} \in \mathbb{K}$ such that

$$\sum_{i=1}^m \sum_{j=1}^n \lambda_{ij} v^i \otimes w^j = 0,$$

then

$$0 = \left(\sum_{i=1}^m \sum_{j=1}^n \lambda_{ij} v^i \otimes w^j\right)(v_k, w_l) = \lambda_{kl}$$

for every $k \in \{1, \dots, m\}$, $l \in \{1, \dots, n\}$. Hence $(v^i \otimes w^j)_{i,j}$ is linearly independent. $\square$

Remark 1.1.28. Using the canonical identification $V^{**} \cong V$ for finite-dimensional vector spaces, we can also make sense of $V_1 \otimes \cdots \otimes V_r$ as $V_1^{**} \otimes \cdots \otimes V_r^{**}$, which is the space of multilinear forms of order r on $V_1^* \times \cdots \times V_r^*$.

Definition 1.1.29 (Tensors)**.** If V is a finite-dimensional vector space over $\mathbb{K}$, an *r-contravariant and s-covariant tensor* or *(r, s)-tensor* is an element of

$$\underbrace{V^* \otimes \cdots \otimes V^*}_{r \text{ times}} \otimes \underbrace{V \otimes \cdots \otimes V}_{s \text{ times}}.$$

In other words, an (r, s)-tensor is a multilinear form on $V^r \times (V^*)^s$.

Example 1.1.30. If $\varphi \in \mathcal{L}(V, V)$, then

$$V \times V^* \to \mathbb{K},\ (v, f) \mapsto f(\varphi(v))$$

is multilinear, i.e., a $(1, 1)$-tensor. Conversely, every $(1, 1)$-tensor is of this form. This gives a canonical identification of the space of $(1, 1)$-tensor with $\mathcal{L}(V, V)$.

Definition 1.1.31 (Alternating forms)**.** Let V be a vector space over $\mathbb{K}$. A multilinear form $\omega\colon V^r \to \mathbb{K}$ is called *alternating* if

$$\omega(v_1, \ldots, v_r) = -\omega(v_1, \ldots, v_{j-1}, v_k, v_{j+1}, \ldots, v_{k-1}, v_j, v_{k+1}, \ldots, v_r)$$

for all $v_1, \ldots, v_r \in V$, $j, k \in \{1, \ldots, r\}$.

The set of all alternating r-forms on V is denoted by $\Lambda^r V^*$. Furthermore, we set $\Lambda^0 V^* = \mathbb{K}$.

Remark 1.1.32. In terms of the sign of a permutation, the defining property of an alternating form can be expressed as

$$\omega(v_1, \ldots, v_r) = \operatorname{sgn} \sigma\, \omega(v_{\sigma(1)}, \ldots, v_{\sigma(r)})$$

for all $v_1, \ldots, v_r \in V$, $\sigma \in S_r$.

Example 1.1.33. If we view elements of $\mathbb{K}^n$ as row vectors, the map

$$(\mathbb{K}^n)^n \to \mathbb{K},\ (v_1, \ldots, v_n) \mapsto \det\big(\big(v_1 \ldots v_n\big)\big)$$

is an alternating multilinear form.

Definition 1.1.34 (Wedge product)**.** Let V be a be a vector space over $\mathbb{K}$. The anti-symmetrization operator on $(V^*)^{\otimes r}$ is defined by

$$P_\wedge\colon (V^*)^{\otimes r} \to \Lambda^r V^*,\ \omega \mapsto \left((v_1, \ldots, v_r) \mapsto \frac{1}{r!} \sum_{\sigma \in S_r} \operatorname{sgn} \sigma\, \omega(v_{\sigma(1)}, \ldots, v_{\sigma(r)}) \right).$$

For $\alpha \in \Lambda^r V^*$ and $\beta \in \Lambda^s V^*$, the *wedge product* $\alpha \wedge \beta$ is defined as

$$\alpha \wedge \beta = \frac{(r+s)!}{r!s!} P_\wedge(\alpha \otimes \beta).$$

Lemma 1.1.35. *Let V be a vector space over $\mathbb{K}$.*

(a) The wedge product is associative: If $\alpha \in \Lambda^r V^*$, $\beta \in \Lambda^s V^*$ and $\gamma \in \Lambda^t V^*$, then

$$(\alpha \wedge \beta) \wedge \gamma = \alpha \wedge (\beta \wedge \gamma).$$

(b) If $v_1, \dots, v_r \in V$ and $\varphi_1, \dots, \varphi_r \in V^*$, then

$$(\varphi_1 \wedge \dots \wedge \varphi_r)(v_1, \dots, v_r) = \det((\varphi_j(v_k))_{j,k=1}^r).$$

(c) If $\alpha \in \Lambda^r V^*$, $\beta \in \Lambda^s V^*$, then

$$\alpha \wedge \beta = (-1)^{rs} \beta \wedge \alpha.$$

(d) If V has finite dimension n, then $\Lambda^r V^*$ is a vector space of dimension $\binom{n}{r}$. In particular, $\Lambda^r V^* = \{0\}$ for $r > n$.

Remark 1.1.36. One consequence of the previous result is that if $\dim V = n$, then $\Lambda^{n-r} V^* \cong \Lambda^r V$ and $\Lambda^0 V^* \cong \mathbb{K}$, $\Lambda^1 V^* \cong V^*$.

Example 1.1.37. If $V = \mathbb{K}^3$, then $\dim \Lambda^0 V^* = \dim \Lambda^3 V^* = 1$, $\dim \Lambda^1 V^* = \dim \Lambda^2 V^* = 3$. More explicitly, isomorphisms of $\Lambda^1 V^*$ and $\Lambda^2 V^*$ with $\mathbb{R}^3$ are given by

$$\varphi\colon \mathbb{R}^3 \to \Lambda^1 V^*, (v^1, v^2, v^3) \mapsto v^1 e^1 + v^2 e^2 + v^3 e^3$$
$$\psi\colon \mathbb{R}^3 \to \Lambda^2 V^*, (v^1, v^2, v^3) \mapsto v^1 e^2 \wedge e^3 + v^2 e^3 \wedge e^1 + v^3 e^1 \wedge e^2$$

Hence, whenever $v, w \in \mathbb{R}^3$, then there exists a unique vector $v \times w \in \mathbb{R}^3$ such that $\psi(v \times w) = \varphi(v) \wedge \varphi(w)$. This is the classical cross product of vectors.

Definition 1.1.38 (Orientation)**.** Let V be a finite-dimensional vector space over $\mathbb{R}$. Two bases $(v_j)_{j=1}^n$ and $(w_j)_{j=1}^n$ are said to have *the same orientation* if there exists $\lambda > 0$ such that $w_1 \wedge \dots \wedge w_n = \lambda v_1 \wedge \dots \wedge v_n$. Having the same orientation defines an equivalence relation with two equivalence classes on the set of all finite bases of V. An equivalence class is called an *orientation* of V.

Remark 1.1.39. In this definition, the order of the basis vectors matters. If we swap e_i and e_j for $i \neq j$, the orientation of the basis changes.

Remark 1.1.40. If V is a vector space over $\mathbb{K}$ and $(v_i)_{i=1}^n$ and $(w_i)_{i=1}^n$ are bases of V, then $v_1 \wedge \dots \wedge v_n$ and $w_1 \wedge \dots \wedge w_n$ are non-zero elements of $\Lambda^n V^*$. Since $\dim \Lambda^n V^* = 1$, there exists $\lambda \in \mathbb{K} \setminus \{0\}$ such that $w_1 \wedge \dots \wedge w_n = \lambda v_1 \wedge \dots \wedge v_n$.

This explains why it is only sensible to define an orientation for bases of *real* vector spaces: $\mathbb{R} \setminus \{0\}$ has two connected components, the positive and negative numbers, while $\mathbb{C} \setminus \{0\}$ is connected.

Example 1.1.41. The orientation of $\mathbb{R}^n$ containing the standard basis $(e_i)_{i=1}^n$ is called the *standard orientation* of $\mathbb{R}^n$.

1.1.2 Topology

Definition 1.1.42 (Metric space)**.** A *metric* on a set X is a map $d\colon X \times X \to [0, \infty)$ with the following three properties:

(a) Non-degeneracy: For all $x, y \in X$, $d(x, y) = 0$ if and only if $x = y$.

(b) Symmetry: $d(x, y) = d(y, x)$ for all $x, y \in X$.

(c) Triangle inequality: $d(x, y) \leq d(x, z) + d(z, y)$ for all $x, y, z \in X$.

A set with a metric is called a *metric space*.

Example 1.1.43 (Euclidean metric)*.* The Euclidean metric on $\mathbb{K}^n$ is defined by

$$d(x, y) = \left(\sum_{i=1}^{n} |x_i - y_i|^2 \right)^{1/2}, \quad x, y \in \mathbb{K}^n.$$

Example 1.1.44 (Discrete metric)*.* For any set X the discrete metric is defined by

$$d(x, y) = \begin{cases} 1 & \text{if } x \neq y, \\ 0 & \text{if } x = y, \end{cases} \quad x, y \in X.$$

Definition 1.1.45 (Topology)**.** A *topology* on a set X is a subset $\mathcal{T}$ of $\mathcal{P}(X)$ with the following three properties:

(a) $\emptyset, X \in \mathcal{T}$.

(b) If $U, V \in \mathcal{T}$, then $U \cap V \in \mathcal{T}$.

(c) If I is an arbitrary index set and $U_i \in \mathcal{T}$, $i \in I$, then $\bigcup_{i \in I} U_i \in \mathcal{T}$.

The elements of $\mathcal{T}$ are called *open subsets* of X. A set with a topology is called a *topological space*.

Every metric gives rise to a topology in the following way.

Proposition 1.1.46. *Let (X, d) be a metric space. For $x \in X$ and $r > 0$ let $B_r(x) = \{y \in X \mid d(x, y) < r\}$. The set*

$$\mathcal{T}_d = \{U \subset X \mid \forall x \in U \, \exists r > 0\colon B_r(x) \subset U\}$$

is a topology on X.

Proof. Clearly $\emptyset, X \in \mathcal{T}_d$. If $U, V \in \mathcal{T}_d$, let $x \in U \cap V$. By definition, there exists $r, s > 0$ such that $B_r(x) \subset U$, $B_s(x) \subset V$. Let $t = \min\{r, s\}$. Since $B_t(x) \subset B_r(x)$ and $B_t(x) \subset B_s(x)$, we have $B_t(x) \subset U \cap V$. Hence $U \cap V \in \mathcal{T}_d$.

Now let I be an arbitrary index set and $U_i \in \mathcal{T}_d$ for $i \in I$. If $x \in \bigcup_{i \in I} U_i$, then there exists $j \in I$ such that $x \in U_j$. By definition, there exists $r > 0$ such that $B_r(x) \subset U_j \subset \bigcup_{i \in I} U_i$ Thus $\bigcup_{i \in I} U_i \in \mathcal{T}_d$. □

Remark 1.1.47. Note that we did not even use the triangle inequality in the proof. However, without is, the topology defined in the previous proposition can be quite pathological.

Definition 1.1.48 (Open ball, neighborhood)**.** Let (X, d) be a metric space. If $x \in X$ and $r > 0$, then $B_r(x) = \{y \in X \mid d(x, y) < r\}$ is called the *open r-ball around x*. More generally, if $A \subset X$, the set $U_r(A) = \{y \in X \mid \exists a \in A \colon d(a, y) < r\}$ is called the *open r-neighborhood of A*.

Lemma 1.1.49. *If (X, d) is a metric space, $A \subset X$ and $r > 0$, then $U_r(A)$ is open.*

Proof. If $x \in U_r(A)$, then there exists $a \in A$ such that $d(x, a) < r$. Let $s = r - d(x, a)$. If $y \in X$ with $d(x, y) < s$, then

$$d(y, a) \leq d(y, x) + d(x, a) < r - d(x, a) + d(x, a) < r.$$

Hence $y \in U_r(A)$. It follows that $B_s(x) \subset U_r(A)$. Hence $U_r(A)$ is open. □

Definition 1.1.50 (Continuous map, homeomorphism)**.** Let X and Y be topological spaces. A map $f \colon X \to Y$ is called *continuous* if for every open subset V of Y the preimage $f^{-1}(V) = \{x \in X \mid f(x) \in V\}$ is open.

A map $f \colon X \to Y$ is called a *homeomorphism* if it is bijective and both f and f^{-1} are continuous. In this case, X and Y are called *homeomorphic*.

Remark 1.1.51. If $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ are topological spaces and $f \colon X \to Y$ is a homeomorphism, then $\mathcal{T}_X = \{f^{-1}(O) \mid O \in \mathcal{T}_Y\}$. In the other words, the sets X and Y and the respective topologies only differ by "renaming". In this sense, heomeomorphic spaces have the same topological properties.

Lemma 1.1.52. *Let X and Y be topological spaces. A map $f \colon X \to Y$ is continuous if and only if for every $x \in X$ and open neighborhood V of $f(x)$ there exists an open neighborhood U of x such that $f(U) \subset V$.*

Lemma 1.1.53. *Let (X, d) and (Y, ρ) be metric spaces. A map $f \colon X \to Y$ is continuous if and only if for every $x \in X$ and every sequence (x_n) in X such that $d(x_n, x) \to 0$ one has $\rho(f(x_n), f(x)) \to 0$.*

Definition 1.1.54 (Closed subsets)**.** A subset F of a topological space X is called *closed* if $X \setminus F$ is open.

Remark 1.1.55. There can be subsets that are neither open nor closed and subsets that are both open and closed. For example, consider $\mathbb{R}$ with the topology induced by the Euclidean metric. Then $[0, 1)$ is neither open nor closed and $\mathbb{R}$ itself is both open and closed.

Lemma 1.1.56. *If X is a topological space, the closed subsets of X have the following properties:*

(a) $\emptyset$ and X are closed.

(b) If $F, G \subset X$ are closed, then $F \cup G$ is closed.

(c) If I is an arbitrary index set and $F_i \subset X$ is closed for every $i \in I$, then $\bigcap_{i \in I} F_i$ is closed.

Definition 1.1.57 (Neighborhoods, convergence of sequences)**.** Let X be a topological space and $x \in X$. A subset V of X is called a *neighborhood* of x if there exists an open subset U of X such that $\{x\} \subset U \subset V$

Let (x_n) be a sequence in X and $x \in X$. We say that (x_n) converges to x and write $x_n \to x$ if for every open subset U of X that contains x there exists $N_0 \in \mathbb{N}$ such that $x_n \in U$ for every $n \geq N_0$.

Remark 1.1.58. A sequence can converge to several points. For example, if $\mathcal{T} = \{\emptyset, X\}$, then every sequence converges to every point in X. This happens because there are not enough open sets to separate the points. To avoid such pathologies, we will focus on topological spaces that satisfy a certain separation axiom.

Definition 1.1.59 (Hausdorff space)**.** A topological space X is called *Hausdorff* if for every pair of distinct points $x, y \in X$ there exist open neighborhoods U of x and V of y such that $U \cap V = \emptyset$.

Lemma 1.1.60. *If X is a Hausdorff topological space, (x_n) is a sequence in X and $x, y \in X$ such that $x_n \to x$ and $x_n \to y$, then $x = y$.*

Proof. Suppose for a contradiction that $x \neq y$. Let U be an open neighborhood of x and V an open neighborhood of y such that $U \cap V = \emptyset$. Since $x_n \to x$, there exists $N_0 \in \mathbb{N}$ such that $x_n \in U$ for $n \geq N_0$, and since $x_n \to y$, there exists $M_0 \in \mathbb{N}$ such that $x_n \in V$ for $n \geq M_0$. Hence $x_n \in U \cap V$ for $n \geq \max\{N_0, M_0\}$, a contradiction. $\square$

Lemma 1.1.61. *If (X, d) is a metric space, then the topology induced by d is Hausdorff.*

Proof. If $x, y \in X$ are distinct points and $r = d(x, y)$, then $B_{r/2}(x)$ and $B_{r/2}(y)$ are open subsets of X that contain x and y, respectively. If there were $z \in B_{r/2}(x) \cap B_{r/2}(y)$, we would have

$$r = d(x, y) \leq d(x, z) + d(z, y) < r/2 + r/2 = r,$$

a contradiction. Thus $B_{r/2}(x) \cap B_{r/2}(y) = \emptyset$. $\square$

Lemma 1.1.62. *If (X, d) is a metric space, then a sequence (x_n) in X converges to $x \in X$ with respect to the topology induced by d if and only if $d(x_n, x) \to 0$.*

Proof. If $x_n \to x$ and $\varepsilon > 0$, then there exists $N_0 \in \mathbb{N}$ such that $x_n \in B_\varepsilon(x)$ for every $n \geq N_0$ since $B_\varepsilon(x)$ is an neighborhood of x. Thus $d(x_n, x) < \varepsilon$ for $n \geq N_0$. As $\varepsilon > 0$ was arbitrary, we conclude $d(x_n, x) \to 0$.

Assume conversely that $d(x_n, x) \to 0$ and let U be an open neighborhood of x. By definition of the topology induced by d, there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subset U$. Since $d(x_n, x) \to 0$, there exists $N_0 \in \mathbb{N}$ such that $d(x_n, x) < \varepsilon$ for $n \geq N_0$. In other words, $x_n \in B_\varepsilon(x) \subset U$ for $n \geq N_0$. Thus $x_n \to x$. $\square$

Lemma 1.1.63. *Let (X, d) be a metric space. A subset F of X is closed if and only if for every sequence (x_n) in F that converges to some $x \in X$ one has $x \in F$.*

Proof. First assume that F is closed and (x_n) is a sequence in F that converges to $x \in X$. By definition, $X \setminus F$ is open. Suppose for a contradiction that $x \in X \setminus F$. As $X \setminus F$ is open, there exists $N_0 \in \mathbb{N}$ such that $x_n \in X \setminus F$ for every $n \geq N_0$, a contradiction.

Now assume conversely that for every sequence (x_n) in F that converges to $x \in X$ we have $x \in F$. We need to show that $X \setminus F$ is open. Let $x \in X \setminus F$. If $B_{1/n}(x) \cap F \neq \emptyset$ for every $n \in \mathbb{N}$, we can find a sequence (x_n) such that $x_n \in B_{1/n}(x)$ for every $n \in \mathbb{N}$. But then $x_n \to x$, which implies $x \in F$ by assumption, a contradiction. Thus $B_{1/n}(x) \subset X \setminus F$ for some $n \in \mathbb{N}$. Hence $X \setminus F$ is open by the definition of the topology induced by a metric. $\square$

Definition 1.1.64 (Compact space)**.** A topological space K is called *compact* if it is Hausdorff and for every index set I and every family $(U_i)_{i \in I} U_i$ of open subsets of K such that $K \subset \bigcup_{i \in I}$ there exists a finite subset J of I such that $K \subset \bigcup_{j \in J} U_j$.

Remark 1.1.65. A family $(U_i)_{i\in I}$ of open subsets of K such that $K \subset \bigcup_{i\in I} U_i$ is called an *open covering.* If $J \subset I$, then the family $(U_j)_{j\in J}$ is called a *subcovering.* With this terminology, a topological space is compact if it is Hausdorff and every open covering has a finite subcovering.

Definition 1.1.66 (Subspace topology, compact subset)**.** If $(X, \mathcal{T})$ is a topological space and $A \subset X$, the *subspace topology* on A is defined as

$$\mathcal{T}_A = \{U \cap A \mid U \in \mathcal{T}\}.$$

A subset K of X is called *compact* if it is a compact topological space in the subspace topology.

Proposition 1.1.67. *Let X be a Hausdorff topological space and $K \subset X$ a compact subset.*

(a) K is a closed subset of X.

(b) If X is a metric space with the topology induced by the metric, then K is also bounded.

Proof. (a) We have to prove that $X \setminus K$ is open. Let $x \in X \setminus K$. By the Hausdorff property, for every $y \in K$ there exist open subsets U_y, V_y of X such that $x \in U_y$, $y \in V_y$ and $U_y \cap V_y = \emptyset$. Then $V_y \cap K$ is an open subset of K for the subspace topology and $K \subset \bigcup_{y\in K} V_y \cap K$.

Since K is compact, there exist $n \in \mathbb{N}$ and $y_1, \ldots, y_n \in K$ such that $K \subset \bigcup_{j=1}^n V_{y_i} \cap K$. Let $W_x = \bigcap_{j=1}^n U_{y_j}$, which is an open subset of X containing x. Moreover, $W_x \cap K \subset \bigcap_{j=1}^n U_{y_j} \cap \bigcup_{j=1}^n V_{y_j} = \emptyset$. Thus $W_x \subset X \setminus K$.

As $x \in X \setminus K$ was arbitrary, we conclude $X \setminus K = \bigcup_{x\in X\setminus K} W_x$. Hence $X \setminus K$ is open as union of open sets.

(b) The family $(B_1(y) \cap K)_{y\in K}$ is an open covering of K. Since K is compact, there exists $n \in \mathbb{N}$ and $y_1, \ldots, y_n \in K$ such that $K \subset \bigcup_{j=1}^n B_1(y_j)$. Hence, if $z \in K$ is arbitrary, then there exists $j \in \{1, \ldots, n\}$ such that $z \in B_1(y_j)$ and thus

$$d(z, y_1) \leq d(z, y_j) + d(y_j, y_1) < 1 + \max_{1\leq k\leq n} d(y_k, y_1) =: R.$$

Therefore $K \subset B_R(y_1)$, which implies that K is bounded. □

Lemma 1.1.68. *If K is a compact topological space and $C \subset K$ is closed, then C is compact.*

Theorem 1.1.69. *A metric space X is compact if and only if every sequence in X has a convergent subsequence.*

Lemma 1.1.70. *If K is a compact topological space and (C_n) is a sequence of non-empty closed subsets of K such that $C_{n+1} \subset C_n$, then $\bigcap_{n=1}^{\infty} C_n$ is a non-empty closed subset of K.*

Proof. Suppose for a contradiction that $\bigcap_{n=1}^{\infty} C_n = \emptyset$. Then $(K \setminus C_n)_{n \in \mathbb{N}}$ is an open covering of K. Since K is compact, there exist $m \in \mathbb{N}$ and $j_1, \dots, j_m \in \mathbb{N}$ such that

$$K \subset \bigcup_{k=1}^{m} K \setminus C_{j_k} = K \setminus \bigcap_{k=1}^{m} C_{j_k} = K \setminus C_{\max\{j_1,\dots,j_m\}}.$$

But this implies $C_{\max\{j_1,\dots,j_m\}} = \emptyset$, a contradiction. Hence $\bigcap_{n=1}^{\infty} C_n \neq \emptyset$. Moreover, the subset if closed as intersection of closed subsets. □

Definition 1.1.71. Let X be a topological space and $A \subset X$. The closure $\overline{A}$ of A is defined as

$$\overline{A} = \bigcap_{\substack{C \supset A \\ C \text{ closed}}} C.$$

Remark 1.1.72. As intersection of closed sets, the closure of A is closed and thus the smallest closed set containing A. In particular, the set A itself is closed if and only if $\overline{A} = A$.

Lemma 1.1.73. *If X is a metric space and $A \subset X$, then*

$$\overline{A} = \{x \in X \mid \exists \text{ sequence } (x_n) \text{ in } A\colon x_n \to x\}.$$

Proof of Theorem 1.1.69. First assume that K is compact and (x_n) is a sequence in K. Let $C_n = \overline{\{x_k \mid k \geq n\}}$. By definition, C_n is closed and $C_{n+1} \subset C_n$. By Lemma 1.1.70, the intersection $\bigcap_{n=1}^{\infty} C_n$ is non-empty.

Let $x \in \bigcap_{n=1}^{\infty} C_n$. We construct a subsequence (x_{j_n}) of (x_n) inductively. Let $j_1 = 1$. Now assume that we are given $j_1, \dots, j_n$ with $d(x, x_{j_n}) < \frac{1}{n}$. Since $x \in C_{j_n+1} = \overline{\{x_k \mid k \geq j_n + 1\}}$, we can find $k \geq j_n + 1$ with $d(x, x_k) < \frac{1}{n+1}$ by Lemma 1.1.73. Set $j_{n+1} = k$. By construction, $d(x, x_{j_n}) < \frac{1}{n}$, hence $x_{j_n} \to x$.

The converse direction is harder and will not be discussed in this course. □

Proposition 1.1.74. *Let (X, d) be a metric space such that the closed balls $\bar{B}_r(x) = \{y \in X \mid d(x, y) \leq r\}$ are compact for every $x \in X$, $r > 0$. Then every bounded closed subset of X is compact.*

Proof. If $A \subset X$ is bounded, then there exists $x \in X$ and $r > 0$ such that $A \subset \bar{B}_r(x)$. If A is furthermore closed, then it is compact as a closed subset of a compact space by Lemma 1.1.68. □

Remark 1.1.75. The property that closed balls are compact is satisfied for $\mathbb{K}^n$ with the Euclidean metric or more generally for any finite-dimensional normed space. Note however that in general, a bounded closed subset of a metric space need not be compact.

Proposition 1.1.76. *If K is a compact topological space, Y is a topological space and $f\colon K \to Y$ is continuous, then $f(K)$ is compact.*

Proof. Let $(V_i)_{i\in I}$ be an open covering of $f(K)$. Then $(f^{-1}(V_i))_{i\in I}$ is an open covering of K. Since K is compact, there exists a finite subcovering $(f^{-1}(V_j))_{j\in J}$. Thus $(V_j)_{j\in J}$ is a finite open covering of $f(K)$. $\square$

Corollary 1.1.77. *If K is a compact topological space and $f\colon X \to \mathbb{R}$ is a continuous map, then f attains its maximum and minimum.*

Proof. The image $f(K)$ is a bounded and closed subset of $\mathbb{R}$ by the previous result and thus has a minimum and maximum. $\square$

Remark 1.1.78. Unless stated otherwise, the topology on $\mathbb{K}^n$ is always taken to be the standard topology, i.e., the topology induced by the Euclidean metric.

1.1.3 Differentiability

Definition 1.1.79 (Norm)**.** Let V be a vector space over $\mathbb{K}$. A function $\|\cdot\|\colon V \to [0,\infty)$ is called a *norm* if it satisfies the following three properties:

(a) Non-degeneracy: $\|v\| = 0$ if and only if $v = 0$.

(b) Positive homogeneity: $\|\lambda v\| = |\lambda|\|v\|$ for all $\lambda \in \mathbb{K}$, $v \in V$.

(c) Triangle inequality: $\|v + w\| \leq \|v\| + \|w\|$ for all $v, w \in V$.

A vector space with a norm is called a *normed space.*

Example 1.1.80. For $p \in [1,\infty)$, the p-norm on $\mathbb{K}^n$ is defined by

$$\|\cdot\|_p\colon \mathbb{K}^n \to [0,\infty),\ \|x\|_p = \left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p}.$$

For $p = \infty$ we set $\|x\|_\infty = \max_{1\leq j\leq n}|x_j|$.

Definition 1.1.81 (Bounded linear maps)**.** If $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ are normed spaces, a linear map $L\colon V \to W$ is called *bounded* if there exists $C > 0$ such that
$$\|Lv\|_W \leq C\|v\|_V$$
for all $v \in V$. The set of all bounded linear maps from V to W is denoted by $\mathcal{L}(V, W)$. The operator norm on $\mathcal{L}(V, W)$ is defined by

$$\|\cdot\|_{\mathrm{op}}\colon \mathcal{L}(V, W) \to [0, \infty),\ \|\varphi\|_{\mathrm{op}} = \sup\|\varphi(v)\|_W : \|v\|_V \leq 1\}.$$

Remark 1.1.82. If V is finite-dimensional, then every linear map from V to W is bounded. Hence this notation is consistent with the notation we introduced for finite-dimensional vector spaces.

Remark 1.1.83. If $(V, \|\cdot\|)$ is a normed space, then the metric induced by $\|\cdot\|$ is defined as $d(v, w) = \|v - w\|$. Then $\|\cdot\|$ also induces a topology on V, namely the topology induced by d as discussed in the previous section. Thus it makes sense to speak of complete normed spaces, open subsets etc.

Definition 1.1.84. Let V be a vector space over $\mathbb{K}$. Two norm $\|\cdot\|_1$, $\|\cdot\|_2$ on V are said to be *equivalent* if there exists $C > 0$ such that

$$C^{-1}\|v\|_1 \leq \|v\|_2 \leq C\|v\|_1$$

for all $v \in V$.

Proposition 1.1.85. *Let V be a vector space over $\mathbb{K}$. The norms $\|\cdot\|_1$, $\|\cdot\|_2$ on V are equivalent if and only if they induce the same topology.*

Theorem 1.1.86. *Any two norms on a finite-dimensional vector space are equivalent.*

Definition 1.1.87 (Derivative)**.** Let V, W be normed spaces and $U \subset V$ an open subset. A function $f\colon U \to W$ is called *differentiable* at $p \in U$ if there exists a continuous linear map $L \in \mathcal{L}(V, W)$ such that

$$\lim_{h \to 0} \frac{\|f(p + h) - f(p) - L[h]\|}{\|h\|} = 0.$$

If f is differentiable at p, the linear map L in the definition is unique and is called the *derivative of f at p* and denoted by $Df(p)$.

If f is differentiable at each point $p \in U$, we say that f is differentiable. In this case, the map

$$Df\colon U \to \mathcal{L}(V, W),\ p \mapsto Df(p)$$

is called the *derivative* of f.

Remark 1.1.88.

- An equivalent way to describe the differentiability of f at p is to say that the remainder term

$$R(h) = f(p+h) - f(p) - L[h]$$

satisfies $\lim_{h\to 0} \frac{\|R(h)\|}{\|h\|} = 0$. This property is also expressed as $R(h) = o(\|h\|)$ as $h \to 0$.

- The affine map $x \mapsto f(p) + Df(p)(x-p)$ is the best affine approximation of f at p. In this sense, the derivative is the "linearization" of f at p.

Example 1.1.89. If $L \in \mathcal{L}(V, W)$, then L is differentiable and $DL(p) = L$ for all $p \in V$. Indeed,

$$\frac{\|L(p+h) - L(p) - L(h)\|}{\|h\|} = 0.$$

Lemma 1.1.90. *Let V, W be normed spaces, $U \subset V$ an open subset and $p \in U$. If $f \colon U \to W$ is differentiable at p, then f is continuous at p.*

Proof. With the remainder term from the previous remark we have

$$\|f(p+h) - f(p)\| \leq \|r(h)\| + \|L(h)\| \leq \|r(h)\| + \|L\|_{\mathrm{op}}\|h\|.$$

Since $r(h) = o(\|h\|)$, we have $\|r(h)\| \to 0$ as $h \to 0$. Therefore, $\|f(p+h) - f(p)\| \to 0$ as $\|h\| \to 0$. □

Proposition 1.1.91. *Let V_1, V_2, W be normed spaces, $U_1 \subset V_1$, $U_2 \subset V_2$ open subsets and $f \colon U_1 \to U_2$, $g \colon U_2 \to W$ continuous maps. If f is differentiable at p and g is differentiable at $f(p)$, then $g \circ f$ is differentiable at p and*

$$D(g \circ f)(p) = Dg(f(p)) \circ Df(p).$$

Proof. We already have a candidate for the map L in the definition of differentiability, namely $L = Dg(f(p)) \circ Df(p)$. Let

$$\begin{aligned} r(x) &= \frac{f(x) - f(p) - Df(p)[x-p]}{\|x-p\|}, \ x \neq p, \\ s(y) &= \frac{g(y) - g(f(p)) - Dg(f(p))[y - f(p)]}{\|y - f(p)\|}, \ y \neq f(p), \end{aligned}$$

and set $r(p) = 0$, $s(f(p)) = 0$. Since f and g are differentiable at p and $f(p)$, respectively, the functions r and s are continuous in p and $f(p)$, respectively.

Define $t\colon U_1 \to W$ by $t(p) = 0$ and

$$t(x) = Dg(f(p))[r(x)] + s(f(x)) \left\| Df(p)\frac{x-p}{\|x-p\|} + r(x) \right\|$$

for $x \neq p$. Then t is continuous in p and we have

$$\begin{aligned} g(f(x)) &= g(f(p)) + Dg(f(p))[Df(p)[x-p]] + Dg(f(p))r(x)\|x-p\| \\ &\quad + s(f(x))\|Df(p)[x-p] + r(x)\|x-p\|\| \\ &= g(f(p)) + L[x-p] + t(x)\|x-p\|. \end{aligned}$$

As t is continuous in p, we conclude that $g \circ f$ is differentiable at p with derivative $D(g \circ f)(p) = L$. □

Definition 1.1.92. Let $V_1, \ldots, V_r$ and W be normed spaces. A multilinear map $\varphi\colon V_1 \times \cdots \times V_r \to W$ is called *bounded* if there exists $C > 0$ such that

$$\|\varphi(v_1, \ldots, v_r)\|_W \leq C\|v_1\|_{V_1} \ldots \|v_r\|_{V_r}.$$

We write $\mathcal{L}^r(V_1, \ldots, V_r; W)$ for the space of all bounded multilinear maps from $V_1 \times \cdots \times V_r$ to W. If $V_1 = \cdots = V_r$, we simply write $\mathcal{L}^r(V; W)$.

On $\mathcal{L}^r(V_1, \ldots, V_r; W)$ one defines a norm by

$$\|\varphi\| = \sup\{\|\varphi(v_1, \ldots, v_r)\|_W : \|v_1\|_{V_1}, \ldots, \|v_r\|_{V_r} \leq 1\}.$$

Remark 1.1.93 (Contraction of multilinear maps). If $\varphi \in \mathcal{L}^{r+1}(V, W)$ and $v \in V$, then $\varphi[h, \cdot\,] \in \mathcal{L}^r(V; W)$ and $\|\varphi[h, \cdot\,]\| \leq \|\varphi\|\|h\|$.

Definition 1.1.94 (Derivatives of higher order)**.** Let V, W be normed spaces, $U \subset V$ an open subset, $p \in U$ and $r \in \mathbb{N}$. We say that f is $r+1$ times differentiable at p if there exists $\varepsilon > 0$ such that f is r times differentiable on $B_\varepsilon(p)$ and there exists $L \in \mathcal{L}^{r+1}(V; W)$ such that

$$\lim_{h \to 0} \frac{\|D^r f(p+h) - D^r f(p) - L[h, \cdot\,]\|}{\|h\|} = 0.$$

In this case, the map $L \in \mathcal{L}^{r+1}(V; W)$ is unique and denoted by $D^{r+1}f(p)$. If f is $r+1$ times differentiable at every point $p \in U$, we say that f is $r+1$ times differentiable on U and call the map $D^{r+1}f\colon U \to \mathcal{L}^{r+1}(V; W)$ the $(r+1)$-th derivative of f on U.

Remark 1.1.95.

- This is a recursive definition: To define what it means for f to be $r+1$ times differentiable on U, we need to know what it means for f to be r times differentiable on U and that the r-th derivative of f at p is an element of $\mathcal{L}^r(V; W)$. This is fine since after r steps we arrive at the notion of differentiability we have defined before.

- A function f is $r+1$ times differentiable at p if it is r-times differentiable on an open ball around p and the r-th derivative itself is differentiable. Essentially, $D^{r+1}f$-is the derivative of the $D^r f$ in this case, up to an identification of $\mathcal{L}(V, \mathcal{L}^r(V; W))$ and $\mathcal{L}^{r+1}(V; W)$.

- Recall that if V is finite-dimensional and $W = \mathbb{K}$, we denoted $\mathcal{L}^r(V; \mathbb{K})$ by $V^* \otimes \cdots \otimes V^*$ (r factors). Hence the r-th derivative at p is a $(r, 0)$-tensor.

Example 1.1.96. Let $f\colon \mathbb{R}^n \to \mathbb{R}$, $x \mapsto x^{\mathrm{T}}x = \sum_{j=1}^n (x^j)^2$. Let $p \in \mathbb{R}^n$. If $Df(p)$ exists, it is an element of $\mathcal{L}(\mathbb{R}^n, \mathbb{R}) = (\mathbb{R}^n)^*$. We identify elements of $(\mathbb{R}^n)^*$ with row vectors.

We have

$$f(p+h) - f(p) = (p+h)^{\mathrm{T}}(p+h) - p^{\mathrm{T}}p = \underbrace{2p^{\mathrm{T}}h}_{\text{linear in } h} + \underbrace{h^{\mathrm{T}}h}_{\text{higher order in } h}.$$

This suggests $Df(p) = 2p^{\mathrm{T}}$. In fact,

$$\frac{|f(p+h) - f(p) - 2p^{\mathrm{T}}h|}{\|h\|_2} = \frac{|h^{\mathrm{T}}h|}{\|h\|_2} \leq \|h\|_2 \overset{\|h\|_2 \to 0}{\to} 0,$$

which confirms our guess.

Now on to the second derivative. If $D^2f(p)$ exists, it is an element of $\mathcal{L}^2(\mathbb{R}^n; \mathbb{R})$, i.e., a bilinear map from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}$. We have

$$Df(p+h)[k] - Df(p)[k] = 2(p+h)^{\mathrm{T}}k - 2p^{\mathrm{T}}k = 2h^{\mathrm{T}}k.$$

This suggests $D^2f(p)[h, k] = 2h^{\mathrm{T}}k$. In fact,

$$\frac{\|Df(p+h)[k] - Df(p)[k] - 2h^{\mathrm{T}}k\|}{\|h\|_2} = 0,$$

which confirms our guess again. Since $D^2f(p)$ does not depend on p, the higher derivatives are constant.

Definition 1.1.97 (Continuously differentiable functions)**.** Let V, W be normed spaces and $U \subset V$ an open subset. We say that a function $f\colon U \to W$ is r *times continuously differentiable* if it is r times differentiable and $D^r f\colon U \to \mathcal{L}^r(V; W)$ is continuous. We write $C^r(U; W)$ for the space of all r times continuously differentiable functions from U to W.

A function $f\colon U \to W$ is called *smooth* if it is r times differentiable for all $r \in \mathbb{N}$. The set of all smooth functions from U to W is denoted by $C^\infty(U; W)$.

Remark 1.1.98. The derivatives of lower order are automatically continuous so that we need to require continuity only for the derivative of highest order.

Remark 1.1.99. Let $U \subset \mathbb{R}^n$ be open and $f\colon U \to \mathbb{R}^m$. There exist functions $f^1, \dots, f^m\colon U \to \mathbb{R}$ such that $f(x) = (f^1(x), \dots, f^m(x))$ for all $x \in U$. The function f is smooth if and only if $f^1, \dots, f^n$ are smooth.

Example 1.1.100. Polynomials of n variables are smooth. These are maps of the form $f\colon \mathbb{K}^n \to \mathbb{K}$, $x \mapsto \sum_{\alpha\in\mathbb{N}^n,\, |\alpha|\le k} a_\alpha x^\alpha$ with $a_\alpha \in \mathbb{K}$ are smooth. Here we use multi-index notation: If $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$, then $x^\alpha = x_1^{\alpha_1} \dots x_n^{\alpha_n}$.

Definition 1.1.101 (Directional derivative)**.** Let V, W be normed spaces, $U \subset V$ an open subset, $f\colon U \to W$ a function and $h \in V$. If the limit

$$\partial_h f(p) = \lim_{t\to 0} \frac{f(p+th) - f(p)}{t}$$

exists, it is called the *directional derivative of f at p in the direction h.*

In the special case when $V = \mathbb{R}^n$ and h is the standard basis vector e_j, we write $\frac{\partial f}{\partial x^j}(p)$ for $\partial_{e_j} f(p)$ and call it a *partial derivative.*

Proposition 1.1.102. *Let V, W be normed spaces, $U \subset V$ an open subset, $f\colon U \to W$ and $p \in U$. If f is differentiable at p, then the directional derivative $\partial_h f(p)$ exists for all $h \in V$ and satisfies*

$$\partial_h f(p) = Df(p)[h].$$

Proof. We have

$$\left\| \frac{f(p+th) - f(p)}{t} - Df(p)[h] \right\| = \frac{\|f(p+th) - f(p) - Df(p)[th]\|}{\|th\|} \|h\|.$$

Since f is differentiable at p, this expression converges to 0 as $t \to 0$. □

The converse of this proposition is not true. There exist functions with directional derivatives in all directions at a point that are still not differentiable at that point. The situation is different if one additionally assumes continuity of the directional derivatives in the following sense.

Proposition 1.1.103. *Let $U \subset \mathbb{R}^m$ be open. A function $f\colon U \to \mathbb{R}^n$ is continuously differentiable if and only if the partial derivatives $\partial_{x^1} f$, $\partial_{x^m} f$ exist and are continuous. In this case, the matrix of the derivative $Df(p)$ with respect to the canonical bases is given by $(\frac{\partial f^j}{\partial x^k}(p))_{j,k}$, where $f = (f^1, \dots, f^n)$.*

Remark 1.1.104. The matrix $(\frac{\partial f^j}{\partial x^k}(p)) = j, k$ is called the *Jacobian* or *Jacobi matrix* of f at p.

It is customary (for reasons that should become clear later) to write $\frac{\partial}{\partial x^j}$ for the standard basis vector e_j and dx^j for the dual basis vector e^j. With this notation, one can write the derivative of a differentiable function $f\colon U \to \mathbb{R}$ as

$$Df = \sum_{j=1}^{m} \frac{\partial f}{\partial x^j} dx^j.$$

This expression is called the *total differential* of f. One often writes df instead of Df if the codomain is $\mathbb{R}$.

1.1.4 Implicit and inverse function theorem

Definition 1.1.105 (Cauchy sequence)**.** Let (X, d) be a metric space. A sequence (x_n) in X is called *Cauchy sequence* if for every $\varepsilon > 0$ there exists $N_0 \in \mathbb{N}$ such that $d(x_m, x_n) < \varepsilon$ for $m, n \geq N_0$.

Proposition 1.1.106. *Every convergent sequence in a metric space is a Cauchy sequence.*

Proof. Let (x_n) be a convergent sequence in the metric space (X, d) with limit x and let $\varepsilon > 0$. Since $x_n \to x$, there exists $N_0 \in \mathbb{N}$ such that $d(x_n, x) < \varepsilon/2$ for all $n \geq N_0$. Hence if $m, n \geq N_0$, we have

$$d(x_m, x_n) \leq d(x_m, x) + d(x, x_n) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \qquad \square$$

The converse is not true. As a simple example take the sequence $(1/n)_{n\in\mathbb{N}}$. This sequence converges to $\{0\}$ (with respect to the Euclidean metric on $\mathbb{R}$). But if we take $\mathbb{R} \setminus \{0\}$ with the Euclidean metric, then $(1/n)$ is still a Cauchy sequence, but it does not have a limit in $\mathbb{R} \setminus \{0\}$. This suggests that the lack of convergence of Cauchy sequences is related to some points "missing" in the space. This motivates the following definition.

Definition 1.1.107 (Complete metric space)**.** A metric space is called *complete* if every Cauchy sequence converges.

Proposition 1.1.108. *Every compact metric space is complete.*

Proof. Let (X, d) be a compact metric space and (x_n) a Cauchy sequence in X. Since X is compact, there exists $x \in X$ and a subsequence (x_{n_k}) of (x_n) such that $x_n \to x$. Let $\varepsilon > 0$. Since (x_n) is a Cauchy sequence, there exists $N_0 \in \mathbb{N}$ such that $d(x_m, x_n) < \varepsilon/2$ for $m, n \geq N_0$. Moreover, since $x_{n_k} \to x$, there exists $K_0 \in \mathbb{N}$ such that $d(x_{n_k}, x) < \varepsilon/2$. If $n \geq N_0$ and $k \geq K_0$, $n_k \geq N_0$, we have

$$d(x_n, x) \leq d(x_n, x_{n_k}) + d(x_{n_k}, x) < \varepsilon. \qquad \square$$

Proposition 1.1.109. *Every closed subset of a complete metric space is complete.*

Proof. Let (X, d) be a complete metric space and C a closed subset of X. If (x_n) is a Cauchy sequence in C, then since (X, d) is complete, there exists $x \in X$ such that $x_n \to x$. Since C is closed, we have $x \in C$. $\square$

Definition 1.1.110. Let (X_1, d_1) and (X_2, d_2) be metric spaces and $L > 0$. A map $f \colon X_1 \to X_2$ is called *Lipschitz (continuous) with constant L* if

$$d_2(f(x), f(y)) \leq Ld(x, y)$$

for all $x, y \in X$.

Proposition 1.1.111. *Every Lipschitz continuous map between metric spaces is continuous.*

Proof. Let (X, d), (Y, ρ) be metric spaces and $f \colon X \to Y$ a Lipschitz continuous map with Lipschitz constant L. If $x \in X$ and (x_n) is a sequence in X such that $x_n \to x$, then

$$d(f(x), f(x_n)) \leq Ld(x, x_n) \to 0. \qquad \square$$

Proposition 1.1.112. *Every continuous linear map between normed spaces is Lipschitz continuous.*

Theorem 1.1.113 (Banach fixed-point theorem)**.** *Let (X, d) be a non-empty complete metric space and $L < 1$. If $F \colon X \to X$ is a Lipschitz continuous map with Lipschitz constant L, then there exists a unique point $x_* \in X$ such that $F(x_*) = x_*$.*

Proof. Choose $x_0 \in X$ and define recursively a sequence (x_n) by $x_1 = F(x_0)$, $x_{n+1} = F(x_n)$. We have

$$d(x_{n+1}, x_n) \leq Ld(x_n, x_{n-1}) \leq \cdots \leq L^n d(x_1, x_0).$$

Thus

$$\begin{aligned} d(x_{n+k}, x_n) &\leq d(x_{n+k}, x_{n+k-1}) + \cdots + d(x_{n+1}, x_n) \\ &\leq (L^{n+k-1} + \cdots + L^n) d(x_1, x_0) \\ &\leq L^n \frac{1 - L^k}{1 - L} d(x_1, x_0) \\ &\leq \frac{L^n}{1 - L} d(x_1, x_0). \end{aligned}$$

Since $L^n \to 0$ as $n \to \infty$, the sequence (x_n) is a Cauchy sequence. As (X, d) is assumed to be complete, there exists $x_* \in X$ such that $x_n \to x_*$. Since F is continuous, it follows that

$$x_* = \lim_{n\to\infty} x_n = \lim_{n\to\infty} F(x_{n-1}) = F(x_*).$$

This settles the existence of a fixed point. For uniqueness, let $y_* \in X$ such that $F(y_*) = y_*$. Since f is L-Lipschitz, we have

$$d(x_*, y_*) = d(F(x_*), F(y_*)) \leq Ld(x_*, y_*).$$

Since $L < 1$, it follows that $x_* = y_*$. □

Definition 1.1.114 (Banach space)**.** A complete normed space is called *Banach space.*

Remark 1.1.115. Here completeness refers to the metric induced by the norm.

Theorem 1.1.116 (Mean value theorem)**.** *Let X, Y be Banach spaces, $a, b \in X$ and U an open subset of X that contains $\{ta + (1-t)b \mid t \in [0,1]\}$. If $f\colon U \to Y$ is differentiable, then*

$$\|f(a) - f(b)\|_Y \leq \sup_{t\in[0,1]} \|Df(ta + (1-t)b)\|_{\mathrm{op}} \|a - b\|_X.$$

Corollary 1.1.117. *Let X, Y be Banach spaces and U a convex open subset of X. If $f\colon U \to Y$ is differentiable and $\sup_{p\in U}\|Df(p)\| < \infty$, then f is L-Lipschitz with $L = \sup_{p\in U}\|Df(p)\|$.*

Theorem 1.1.118 (Inverse function theorem)**.** *Let X, Y be Banach spaces, $W \subset Y$ open, $f\colon W \to Y$ continuously differentiable and $x_0 \in W$. If the derivative $Df(x_0)\colon X \to Y$ of f at x_0 is bijective with $Df(x_0)^{-1} \in \mathcal{L}(Y, X)$, then there exist an open neighborhood U of x_0 and an open neighborhood V of $f(x_0)$ such that f restricts to a bijective map from U to V and the inverse $f|_U^{-1}\colon V \to U$ is continuously differentiable.*

Remark 1.1.119. In general, if X, Y are Banach spaces and $A \in \mathcal{L}(X, Y)$ is bijective, then $A^{-1} \in \mathcal{L}(Y, X)$. However, if Y is infinite-dimensional, this is a deep theorem that will be covered later in this course.

Proof of Theorem 1.1.118. We can assume without loss of generality that $x_0 = 0$ and $f(x_0) = 0$. Otherwise replacing f by $Df(0)^{-1}f$, we can also assume $X = Y$ and $Df(0) = \mathrm{id}$.

Let $r > 0$ such that the closed ball $\bar{B}_r(0)$ is contained in W. For $y \in X$ let

$$g_y \colon \bar{B}_r(0) \to X,\ g_y(x) = x - f(x) + y.$$

Note that $x_* \in \bar{B}_r(0)$ is a fixed point of g_y if and only if $f(x) = y$.

We want to show that for r sufficiently small, $g_y(\bar{B}_r(0)) \subset \bar{B}_r(0)$ and g_y is L-Lipschitz with $L < 1$ in order to apply the Banach fixed-point theorem.

Since Df is continuous at 0, we can assume without loss of generality that $\|\mathrm{id} - Df(x)\|_{\mathrm{op}} \leq 1/2$ for $x \in \bar{B}_r(0)$. By the mean-value theorem, if $\|x\| \leq r$ and $\|y\| \leq r/2$, then

$$\|g_y(x)\| \leq \|y\| + \sup_{x' \in \bar{B}_r(0)} \|\mathrm{id} - Df(x')\|_{\mathrm{op}} \|x\| \leq r.$$

Thus g_y maps $\bar{B}_r(0)$ into itself. Moreover, if $x_1, x_2 \in \bar{B}_r(0)$, then

$$\|g_y(x_1) - g_y(x_2)\| \leq \sup_{x' \in \bar{B}_r(0)} \|\mathrm{id} - Df(x')\|_{\mathrm{op}} \|x_1 - x_2\| \leq \frac{1}{2}.$$

Hence g_y is a 1/2-Lipschitz map from $\bar{B}_r(0)$ to itself for every $y \in \bar{B}_{r/2}(0)$. By the Banach fixed-point theorem, for every $y \in \bar{B}_{r/2}(0)$ there exists a unique $x \in \bar{B}_r(0)$ with $f(x) = y$. Thus, if we let $V = B_{r/2}(0)$ and $U = f^{-1}(V) \cap B_r(0)$, the map $f \colon U \to V$ is bijective.

It remains to show that $f|_U^{-1}$ is continuously differentiable. The bound $\|\mathrm{id} - Df(x)\|_{\mathrm{op}} \leq 1/2$ for $x \in U$ implies that $Df(x)$ is bijective and $Df(x)^{-1} \in \mathcal{L}(Y, X)$. We will prove that later in the course (Neumann series).

Let us first prove that $f|_U^{-1}$ is continuous. For $x_1, x_2 \in U$ we have

$$\begin{aligned}
\|x_1 - x_2\| &\leq \|f(x_1) - f(x_2)\| + \sup_{x' \in \bar{B}_r(0)} \|\mathrm{id} - Df(x')\|_{\mathrm{op}} \|x_1 - x_2\| \\
&\leq \|f(x_1) - f(x_2)\| + \frac{1}{2}\|x_1 - x_2\|.
\end{aligned}$$

Hence $\|x_1 - x_2\| \leq 2\|f(x_1) - f(x_2)\|$. Thus $f|_U^{-1}$ is 2-Lipschitz.

To show that $f|_U^{-1}$ is differentiable, let $y, y' \in V$ and $x = f|_U^{-1}(y)$, $x' = f|_U^{-1}(y')$. We have

$$\begin{aligned}
&\frac{\|f^{-1}(y') - f^{-1}(y) - Df(x)^{-1}[y - y']\|}{\|y - y'\|} \\
&\quad \leq \|Df(x)^{-1}\|_{\mathrm{op}} \frac{\|Df(x)[x - x'] - f(x) - f(x')\|}{\|x - x'\|} \frac{\|f^{-1}(y) - f^{-1}(y')\|}{\|y - y'\|} \\
&\quad \leq \frac{\|Df(x)^{-1}\|_{\mathrm{op}}}{2} \frac{\|Df(x)[x - x'] - f(x) - f(x')\|}{\|x - x'\|}.
\end{aligned}$$

If $y' \to y$, then $x' \to x$ by the continuity of $f|_U^{-1}$. Thus $f|_U^{-1}$ is differentiable with $Df|_U^{-1}(f(x)) = Df(x)^{-1}$. Continuity of $Df|_U^{-1}$ follows from the continuity of the inverse map, which will also be shown later in the course. □

Remark 1.1.120. In the situation of the inverse function theorem, if f is r times continuously differentiable, then the local inverse $f|_U^{-1}$ is also r times continuously differentiable. In particular, if f is smooth and $Df(x)$ is bijective for all $x \in W$, then f is a *local diffeomorphism*.

Remark 1.1.121. If $X = \mathbb{R}^m$ and $Y = \mathbb{R}^n$, then $Df(x_0)\colon \mathbb{R}^m \to \mathbb{R}^n$ can only be bijective if $m = n$, and this is the case if and only if $\det Df(x_0) \neq 0$.

Remark 1.1.122. If one only assumes that f is differentiable (not necessarily continuously differentiable) and $Df(x)$ is bijective for all x in a neighborhood of x_0, then the conclusion of the inverse function theorem still holds (except that the inverse is only differentiable, not necessarily continuously differentiable). This little known result relies on the Brouwer fixed-point theorem instead of the Banach fixed-point theorem.

A close relative of the inverse function theorem is the implicit function theorem. To state it, we need the following bit of notation. Let X, Y be Banach spaces. There are several ways to turn the cartesian product $X \times Y$ into a Banach space, for example by defining

$$\|(x,y)\|_{X\times Y} = \left(\|x\|_X^2 + \|y\|_Y^2\right)^{1/2}.$$

In this way, if $\mathbb{K}^m$ and $\mathbb{K}^n$ are endowed with the Euclidean norm, then $\mathbb{K}^m \times \mathbb{K}^n$ carries the Euclidean norm, too. This norm on the product has the property that the projection mappings

$$\pi_X\colon X \times Y \to X,\ (x,y) \mapsto x,$$
$$\pi_Y\colon X \times Y \to Y,\ (x,y) \mapsto y$$

are continuous.

Let Z be another Banach space and $U \subset X \times Y$ and open subset. If $f\colon U \to Z$ is a (continuously) differentiable function, then for every $x_0 \in \pi_X^{-1}(U)$ and $y_0 \in \pi_Y^{-1}(U)$ the functions

$$f(\,\cdot\,, y_0)\colon \pi_X^{-1}(U) \to Z, x \mapsto f(x, y_0)$$
$$f(x_0, \cdot\,)\colon \pi_Y^{-1}(U) \to Z,\ y \mapsto f(x_0, y)$$

are (continuously) differentiable. We will denote their derivatives by $D_1 f(\,\cdot\,, y_0)$ and $D_2 f(x_0, \cdot\,)$, respectively.

Theorem 1.1.123 (Implicit function theorem)**.** *Let X, Y, Z be Banach spaces, $W \subset X \times Y$ open and $F\colon W \to Z$ continuously differentiable. If $(x_0, y_0) \in W$ such that $D_2F(x_0, y_0)$ is bijective with bounded inverse, then there exist neighborhoods U of (x_0, y_0) and V of x_0 and a unique function $f\colon V \to Y$ such that*

$$\{(x_0, y_0) \in U \mid F(x, y) = F(x_0, y_0)\} = \{(x, f(x)) \mid x \in V\}.$$

Moreover, the function f is continuously differentiable and

$$Df(x) = -(D_2F(x, f(x)))^{-1} D_1F(x, f(x)).$$

for all $x \in V$.

Remark 1.1.124. If $g\colon A \to B$ is a function, the preimage $g^{-1}(b)$ is called a level set of g. The set $\{(a, g(a)) \mid a \in A\}$ is called the graph of g. Thus the implicit function theorem states that if $D_2F(x_0, y_0)$ is bijective, then the level set of F is locally the graph of a function.

Remark 1.1.125. If $X = \mathbb{R}^l$, $Y = \mathbb{R}^m$ and $Z = \mathbb{R}^n$, the invertibility of $D_2f(x_0, y_0)$ implies $m = n$ and is equivalent to $\det D_2f(x_0, y_0) \neq 0$, as in the case of the inverse function theorem.

In a typical application of the implicit function theorem, one chooses the decomposition of the domain into a cartesian product based on the function at hand in the following sense: Let $U \subset \mathbb{R}^m$ be open and $F\colon U \to \mathbb{R}^n$ continuously differentiable. If $q_0 \in \operatorname{im} F$ such that $DF(p_0)$ is surjective ($\operatorname{rk} DF(p_0) = n$) for every $p_0 \in F^{-1}(q_0)$, one calls q_0 a *regular value* of F.

Let $X = \ker DF(p_0)$ and $Y = \{y \in \mathbb{R}^m \mid x \cdot y = 0 \text{ for all } x \in X\}$. Then $X \cap Y = \{0\}$, $X + Y = \mathbb{R}^m$ and the norm on the product described above is exactly the norm of $\mathbb{R}^m$. Since $DF(p_0)$ is surjective, $DF(p_0)|_Y$ is bijective. Hence one can apply the implicit function theorem.

1.2 Vector Fields and Flows

1.2.1 Vector Fields

From now on we focus on finite-dimensional real vector spaces, which we take to be $\mathbb{R}^n$ for some $n \in \mathbb{N}$ without loss of generality. It is convenient to work with functions that have derivatives of arbitrary order.

Definition 1.2.1. Let X, Y be normed spaces and let $U \subset X$, $V \subset Y$ be open. A map $f\colon U \to V$ is called a *diffeomorphism* if it is smooth, bijective and has a smooth inverse.

Remark 1.2.2. More generally, one can define *diffeomorphisms of class C^r* as r times continuously differentiable bijective maps with inverse in C^r. In this sense, the diffeomorphisms from the previous definition are diffeomorphisms of class C^∞. As we will work exclusively in the category of smooth manifolds and smooth maps later, we drop the suffix "of class C^∞" and simply say diffeomorphism.

Definition 1.2.3. Let U be an open subset of $\mathbb{R}^n$ and $p \in U$. A linear map $X_p \colon C^\infty(U) \to \mathbb{R}$ is called *derivation at p* if it satisfies the Leibniz rule (or product rule)

$$X_p(fg) = f(p)X_p(g) + X_p(f)g(p)$$

for all $f, g \in C^\infty(U)$.

Example 1.2.4. The partial derivative $\frac{\partial}{\partial x^j}\big|_p$ is a derivation.

Lemma 1.2.5 (Hadamard)**.** *Let $U \subset \mathbb{R}^n$ be open and $p \in U$. If $f \in C^\infty(U)$, then there exists an open subset V of U containing p and there exist $g_1, \dots, g_n \in C^\infty(V)$ such that*

$$f(x) = f(p) + \sum_{j=1}^{n} g_j(x)(x_j - p_j)$$

for all $x \in V$ and $g_j(p) = \frac{\partial f}{\partial x^j}(p)$.

Proof. Let $r > 0$ so that $B_r(p) \subset U$. For $x \in B_r(p)$ let

$$h \colon [0,1] \to \mathbb{R},\ h(t) = f(tx + (1-t)p).$$

By the chain rule, h is differentiable on $[0,1]$ and

$$h'(t) = \sum_{j=1}^{n} \frac{\partial f}{\partial x^j}(tx + (1-t)p)(x_j - p_j).$$

Thus

$$\begin{aligned}
f(p) - f(x) &= h(1) - h(0) \\
&= \int_0^1 h'(t)\, dt \\
&= \sum_{j=1}^{n} (x_j - p_j) \int_0^1 \frac{\partial f}{\partial x^j}(tx + (1-t)p)\, dt.
\end{aligned}$$

Let

$$g_j\colon B_r(p) \to \mathbb{R},\ g_j(x) = \int_0^1 \frac{\partial f}{\partial x^j}(tx + (1-t)p)\,dt.$$

It is not hard to see that g_j is smooth. Moreover,

$$f(x) = f(p) + \sum_{j=1}^{n}(x_j - p_j)g_j(x)$$

holds by definition. The last property follows by directly by plugging in p in the definition of g_j. $\square$

Definition 1.2.6 (Support of a function)**.** Let X be a topological space and $f\colon X \to \mathbb{K}$ a map. The *support* $\operatorname{supp} f$ of f is the set $\overline{\{x \in X \mid f(x) \neq 0\}}$.

Lemma 1.2.7 (Bump functions)**.** *If $C \subset U \subset V \subset \mathbb{R}^n$ with C closed and U, V open in $\mathbb{R}^n$, then there exists $f \in C^\infty(V)$ such that $0 \leq f \leq 1$, $f(x) = 1$ if $x \in C$ and $\operatorname{supp} f \subset U$.*

Proposition 1.2.8 (Locality of derivations)**.** *Let $U \subset V \subset \mathbb{R}^n$ be open subsets and $p \in U$. If $f \in C^\infty(V)$ is constant on U and $X_p\colon C^\infty(V) \to \mathbb{R}$ is a derivation at p, then $X_p(f) = 0$.*

Proof. First assume that f is constant on all of U. Since $f^2 = f(p)f$, we have

$$f(p)X_p(f) = X_p(f^2) = 2f(p)X_p(f).$$

If $f(p) \neq 0$, this implies $X_p(f) = 0$. If $f(p) = 0$, then $f = 0$ and $X_p(f) = 0$ follows from linearity.

Now let us prove the general case. By the first part, we may assume that $f = 0$ on U. By the previous lemma, there exists $g \in C^\infty(V)$ with $g(p) = 1$ and $\operatorname{supp} g \subset U$. In particular, $fg = 0$. By the product rule,

$$0 = X_p(fg) = f(p)X_p(g) + X_p(f)g(p) = X_p(f). \qquad \square$$

Theorem 1.2.9. *Let $U \subset \mathbb{R}^n$ be open and $p \in U$. Then set of derivations at p forms an n-dimensional vector space with basis $\frac{\partial}{\partial x^1}\big|_p, \ldots, \frac{\partial}{\partial x^n}\big|_p$.*

Proof. Let X_p be a derivation at p. There is $r > 0$ and smooth functions $g_1, \ldots, g_n \in C^\infty(B_r(p), \mathbb{R}^n)$ such that

$$f(x) = f(p) + \sum_{j=1}^{n}(x_j - p_j)g_j(x)$$

for $x \in B_r(p)$. Let $\psi \in C^\infty(U)$ with $\psi|_{B_{r/4}(p)} = 1$ and $\operatorname{supp}\psi \subset B_{r/2}(p)$ and define

$$\tilde{g}_j : U \to \mathbb{R},\ \tilde{f}(x) = \begin{cases} \psi(x)g_j(x) & \text{if } x \in B_r(p), \\ 0 & \text{otherwise.} \end{cases}$$

As $\tilde{g}_j$ vanishes on the complement of $B_{r/2}(p)$ and is smooth on $B_r(p)$, it is smooth on all of U.

Let

$$\tilde{f} : U \to \mathbb{R},\ \tilde{f}(x) = f(p) + \sum_{j=1}^{n} (x_j - p_j)\tilde{g}_j(x).$$

By definition, $\tilde{f}|_{B_{r/4}(p)} = f|_{B_{r/4}(p)}$ and $\tilde{g}_j(p) = g_j(p) = \frac{\partial f}{\partial x^j}(p)$. The previous lemma implies $X_p(f) = X_p(\tilde{f})$.

Let $\pi^j : U \to \mathbb{R}$, $x \mapsto x^j$. By the Leibniz rule,

$$X_p(f) = \sum_{j=1}^{n} X_p((\pi^j - p^j)\tilde{g}_j) = \sum_{j=1}^{n} X_p(\pi^j)\tilde{g}_j(p) = \sum_{j=1}^{n} X_p(\pi^j)\frac{\partial f}{\partial x^j}(p).$$

It follows that $X_p = \sum_{j=1}^n X_p(\pi^j) \frac{\partial}{\partial x^j}\big|_p$. Linear independence of the derivations $\frac{\partial}{\partial x^1}\big|_p, \dots, \frac{\partial}{\partial x^n}\big|_p$ is not hard to see. $\square$

Remark 1.2.10. A rephrasing of the previous result is that the derivations at p are in one-to-one correspondence with n-tuples $(v^1, \dots, v^n) \in \mathbb{R}^n$ via

$$(v^1, \dots, v^n) \mapsto \sum_{j=1}^{n} v^j \partial_{x^j}\big|_p.$$

Note that the right side is nothing but the directional derivative $\partial_v|_p$ for $v = (v^1, \dots, v^n)$. Hence the derivations at p are exactly the directional derivatives at p.

Definition 1.2.11 (Vector field)**.** Let $U \subset \mathbb{R}^n$ be an open subset. A *vector field on* U is a linear operator $X : C^\infty(U) \to C^\infty(U)$ that satisfies the Leibniz rule (or product rule)

$$X(fg) = fX(g) + X(f)g$$

for all $f, g \in C^\infty(U)$. The space of vector fields on U is denoted by $\mathcal{X}(U)$.

Lemma 1.2.12 (Locality of vector fields)**.** *If U, V are open subsets of $\mathbb{R}^n$ with $U \subset V$ and $X \in \mathcal{X}(V)$, then $X(f)|_U = X(g)|_U$ for all $f, g \in C^\infty(V)$ with $f|_U = g|_U$.*

Theorem 1.2.13. *Let $U \subset \mathbb{R}^n$ be open. The map*

$$C^\infty(U, \mathbb{R}^n) \to \mathcal{X}(U),\ (f^1, \ldots, f^n) \mapsto \sum_{j=1}^{n} f^j \frac{\partial}{\partial x^j}$$

is a linear bijection.

Lemma 1.2.14 (Lie bracket)**.** *Let $U \subset \mathbb{R}^n$ be open and $X, Y \in \mathcal{X}(U)$. The map*

$$[X, Y] \colon C^\infty(U) \to C^\infty(U),\ f \mapsto X(Y(f)) - Y(X(f))$$

is a vector field, called the Lie bracket *of X and Y.*

Proof. We only need to verify the Leibniz rule. For $f, g \in C^\infty(U)$ we have

$$\begin{aligned}
[X, Y](fg) &= X(Y(fg)) - Y(X(fg)) \\
&= X(fY(g) + Y(f)g) - Y(fX(g) - X(f)g) \\
&= X(f)Y(g) + fX(Y(g)) + X(Y(f))g + Y(f)X(g) \\
&\quad - Y(f)X(g) - fY(X(g)) - Y(X(f))g - X(f)Y(g) \\
&= f(X(Y(g)) - Y(X(g))) + (X(Y(f)) - Y(X(f)))g \\
&= f[X, Y](g) + [X, Y](f)g.
\end{aligned}$$

□

Lemma 1.2.15. *Let $U \subset \mathbb{R}^n$ be open. The Lie bracket $[\,\cdot\,, \cdot\,] \colon \mathcal{X}(U) \times \mathcal{X}(U) \to \mathcal{X}(U)$ is bilinear and satisfies the* Jacobi identity

$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0$$

for all $X, Y, Z \in \mathcal{X}(U)$.

Definition 1.2.16 (Push-forward)**.** Let U, V be open subsets of $\mathbb{R}^n$ and $\Phi \colon U \to V$ a smooth bijective map with smooth inverse. If $X \in \mathcal{X}(U)$, then the *push-forward* of X under Φ is defined as

$$\Phi_* X \colon C^\infty(V) \to C^\infty(V),\ f \mapsto X(f \circ \Phi) \circ \Phi^{-1}.$$

Lemma 1.2.17. *Let U, V be open subsets of $\mathbb{R}^n$ and $\Phi \colon U \to V$ a smooth bijective map with smooth inverse. If $X, Y \in \mathcal{X}(U)$, then $\Phi_*[X, Y] = [\Phi_* X, \Phi_* Y]$.*

1.2.2 Flows

Definition 1.2.18 (Integral curve)**.** Let $U \subset \mathbb{R}^n$ be open and $X \in \mathcal{X}(U)$. A continuously differentiable map $\gamma \colon I \to U$ defined on an open interval I is called an *integral curve* or *flow curve* of X if

$$\dot{\gamma}(t) = X_{\gamma(t)}$$

for all $t \in I$.

Remark 1.2.19. Recall that every vector field X on U can be written as $X = \sum_j f^j \frac{\partial}{\partial x^j}$ with $f_1, \ldots, f_n \in C^\infty(U)$. Hence the equation $\dot\gamma(t) = X(\gamma(t))$ can be rewritten as

$$\begin{aligned} \dot x^1(t) &= f^1(x^1(t), \ldots, x^n(t)) \\ &\vdots \\ \dot x^n(t) &= f^n(x^1(t), \ldots, x^n(t)) \end{aligned}$$

with $\gamma(t) = \sum_{j=1}^n x^j(t) \frac{\partial}{\partial x^j}$.

Theorem 1.2.20 (Picard–Lindelöf)**.** *Let $I = (a,b) \subset \mathbb{R}$, $U \subset \mathbb{R}^n$ open and $X\colon I \times U \to \mathbb{R}^n$ continuous in the first variable and Lipschitz continuous in the second variable. For each $(t_0, x_0) \in I \times U$ there exists a unique maximal interval $(t^-(t_0,x_0), t^+(t_0,x_0))$ containig t_0 and a unique curve $\gamma \in C^1((t^-,t^+),U)$ that satisfies*

$$\begin{aligned} \dot\gamma(t) &= X(t, \gamma(t)), \quad t \in (t^-(t_0,x_0), t^+(t_0,x_0)), \\ \gamma(t_0) &= x_0. \end{aligned}$$

Remark 1.2.21. The Picard–Lindelöf theorem deals with general non-autonomous equations, i.e., equations for which the right side depends explicitly on t. The flow equation defining integral curves of a vector field in contrast is autonomous. In this case, one can always assume $t_0 = 0$ by shifting the time parameter. In this case, we simply write $t^\pm(x_0)$ for $t^\pm(0, x_0)$.

Definition 1.2.22 (Local Flow)**.** Let $U \subset \mathbb{R}^n$ be open and $X \in \mathcal{X}(U)$. Let $V = \bigcup_{x \in U} (t^-(x), t^+(x)) \times \{x\} \subset \mathbb{R} \times U$. A map

$$\Phi_X \colon V \to U,\ (t,x) \mapsto \Phi_X^t(x)$$

is called a *local flow* for X if

- $\Phi_x^0 = \mathrm{id}_U$,
- $t \mapsto \Phi_X^t(x)$ is an integral curve of X for every $x \in U$.

In this case, the vector field X is called the *infinitesimal generator* of Φ_X.

If $t^-(x) = -\infty$, $t^+(x) = \infty$ for every $x \in U$, then Φ_X is called a *global flow.*

Remark 1.2.23. By the Picard–Lindelöf theorem, the local flow of a vector field exists and is unique.

Proposition 1.2.24. *Let $U \subset \mathbb{R}^n$ be open and $X \in \mathcal{X}(U)$. The local flow Φ_X has the following properties.*

(a) The map $\Phi_X \colon V \to U$ is smooth.

(b) For every $(t_0, x_0) \in V$ there exists an open neighborhood U_{x_0} of x_0 such that $\Phi_X^{t_0} \colon U_{x_0} \to \Phi_X^{t_0}(U_{x_0})$ is a diffeomorphism.

(c) If $s, t \in \mathbb{R}$ and $x \in U$ such that both sides are well-defined, then

$$\Phi_X^s(\Phi_X^t(x)) = \Phi_X^{s+t}(x).$$

Remark 1.2.25. Let us consider the case $U = \mathbb{R}^n$, $V = \mathbb{R} \times \mathbb{R}^n$. In particular, the flow is a global flow. In this case, $\Phi_X^t \colon \mathbb{R}^n \to \mathbb{R}^n$ is a diffeomorphism for all $t \in \mathbb{R}$ and the semigroup property

$$\Phi_X^s(\Phi_X^t(x)) = \Phi_X^{s+t}(x)$$

holds for all $s, t \in \mathbb{R}$, $x \in \mathbb{R}^n$.

Let $\mathrm{Diffeo}(\mathbb{R}^n)$ denote the set of all diffeomorphisms from $\mathbb{R}^n$ to itself. The global flow Φ_X induces a group homomorphism

$$\Phi_X \colon \mathbb{R} \to \mathrm{Diffeo}(\mathbb{R}^n),\ t \mapsto \Phi_X^t.$$

In a suitable sense, this map is also smooth.

Remark 1.2.26. The flow equation implies that the push-forward of X under its flow satisfies $(\Phi_X^t)_* X = X$, at least if Φ_X is a global flow.

Definition 1.2.27. Let $U \subset \mathbb{R}^n$ be open and $X, Y \in \mathcal{X}(U)$. The *Lie derivative* of Y along X is defined as

$$\mathcal{L}_X Y = \lim_{t \to 0} \frac{(\Phi_X^t)_* Y - Y}{t}.$$

Remark 1.2.28. Technically, this definition is not quite correct: For every given $t \in \mathbb{R}$, the set of $x \in U$ such that $t \in (t^-(x), t^+(x))$ may be a proper subset of U, and the set on which Φ_X^t acts as a diffeomorphism may be even smaller. This difficulty can be overcome by defining $(\mathcal{L}_X Y)_p$ for a point $p \in U$ with Y restricted to a neighborhood of p on which Φ_X^t exists and acts as a diffeomorphism. Then one has to check that this definition is independent of the choice of this neighborhood. We will not concern ourselves with these technical difficulties here.

Proposition 1.2.29. *If $U \subset \mathbb{R}^n$ is open and $X, Y \in \mathcal{X}(U)$, then $\mathcal{L}_X Y = [X, Y]$.*

Proposition 1.2.30. *Let $U \subset \mathbb{R}^n$ be open and $X, Y \in \mathcal{X}(U)$. The local flows Φ_X, Φ_Y commute for small times, that is, for every $x \in U$ there exists $\varepsilon > 0$ such that*

$$\Phi_X^s(\Phi_Y^t(x)) = \Phi_Y^t(\Phi_X^s(x)), \quad s, t \in (-\varepsilon, \varepsilon)$$

if and only if $[X, Y] = 0$.

The vector fields $\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}$ commute in the sense of the previous proposition as a consequence of Schwarz's theorem. In fact, up to a change of coordinates, all full rank systems of commuting vector fields are locally of this form.

Theorem 1.2.31. *Let $U \subset \mathbb{R}^n$ be open. If $X_1, \ldots, X_n \in \mathcal{X}(U)$ and $p \in U$ satisfy*

(a) $[X_i, X_j] = 0$ for all $i, j \in \{1, \ldots, n\}$,

(b) $(X_1)_p \ldots, (X_n)_p$ are linearly independent,

then there exists an open neighborhood U_p of p and a diffeomorphism $\varphi \colon U_p \to V$ onto an open subset of V such that $\varphi_* X_j = \frac{\partial}{\partial y^j}$ for $j \in \{1, \ldots, n\}$. Here $\frac{\partial}{\partial y^1}, \ldots, \frac{\partial}{\partial y^n}$ denotes the basis vector fields on V.

Proof. Recall that there are functions ${f^k}_j \in C^\infty(U)$ such that $X_j = \sum_{k=1}^n {f^k}_j \frac{\partial}{\partial x^k}$. Let

$$F \colon U \to \mathbb{R}^{n \times n},\ F(x) = ({f^k}_j(x))_{j,k=1}^n.$$

By assumption, $\det F(p) \neq 0$. Since F is smooth, there exists an open neighborhood W_p of p such that $\det F(x) \neq 0$ for $x \in U_p$, that is, $(X_1)_x, \ldots, (X_n)_x$ are linearly independent for $x \in W_p$.

For $\varepsilon > 0$ sufficiently small we define

$$\Psi \colon (-\varepsilon, \varepsilon)^n \to U,\ (t_1, \ldots, t_n) \mapsto (\Phi_{X_1}^{t_1} \circ \cdots \circ \Phi_{X_n}^{t_n})(p).$$

By the previous proposition, the order of the vector fields $X_1, \ldots, X_n$ does not matter in this definition. Hence we can compute

$$\begin{aligned}\frac{\partial}{\partial t_i}\Psi(t_1, \ldots, t_n) &= \frac{\partial}{\partial t_i}\Phi_{X_i}^{t_i}(\Phi_{X_1}^{t_1} \circ \Phi_{t_{j-1}}^{X_{j-1}} \circ \Phi_{t_{j+1}}^{X_{j+1}} \circ \cdots \circ \Phi_{X_n}^{t_n})(p) \\ &= (X_i)_{\Psi(t_1, \ldots, t_n)}.\end{aligned}$$

Thus $D\Psi = F \circ \Psi$. In particular, $D\Psi(t_1, \ldots, t_n)$ is invertible for $(t_1, \ldots, t_n)$ in a neighborhood of 0. By the inverse function theorem, there exist neighborhoods V of 0 and U_p of p such that $\Psi(V) = U_p$ and $\Psi|_V$ is a diffeomorphism onto its image. Let $\varphi = (\Psi|_V)^{-1}$, which is a diffeomorphism from U_p to V by definition. It follows from the previous computation that $\varphi_* X_i = \frac{\partial}{\partial t_i}$. $\square$

1.2.3 Tensor fields

Let V be a (finite-dimensional) vector space over $\mathbb{K}$. We write $T_{r,s}(V)$ for the set of all (r,s)-tensors.

Definition 1.2.32 (Tensor field)**.** Let $U \subset \mathbb{R}^n$ be open. An (r,s)-*tensor field* on U is a smooth map from U to $T_{r,s}(\mathbb{R}^n)$.

Remark 1.2.33. To speak of a smooth map with values in $T_{r,s}(\mathbb{R}^n)$, we need a norm on it. Fortunately, $T_{r,s}(\mathbb{R}^n)$ is finite-dimensional and thus all norms are equivalent. Hence it does not matter which one we choose for the definition of smoothness.

Remark 1.2.34. Note that $T_{0,1}(V)$ is canonically isomorphic to V. Thus the space of $(0,1)$-tensor fields on U is canonically isomorphic to $C^\infty(U,\mathbb{R}^n)$, which in turn is canonically isomorphic to $\mathcal{X}(U)$.

Remark 1.2.35. Let $\Omega^1(U)$ denote the set of all $(1,0)$-tensor fields on U. We have a pairing between $C^\infty(U,\mathbb{R}^n)$ and $\Omega^1(U)$ given by

$$(\cdot|\cdot)\colon C^\infty(U,\mathbb{R}^n) \times \Omega^1(U) \to C^\infty(U),\, (X,\omega) \mapsto (p \mapsto \omega_p(X_p)).$$

For every $\omega \in \Omega^1(U)$ the map $(\,\cdot\,|\omega)$ is $C^\infty(U)$-linear, that is, it satisfies

$$(fX|\omega) = f(X|\omega)$$

for all $f \in C^\infty(U)$, $X \in C^\infty(U,\mathbb{R}^n)$. Conversely, for any $C^\infty(U)$-linear map $T\colon C^\infty(U,\mathbb{R}^n) \to C^\infty(U)$ there exists $\omega \in \Omega^1(U)$ such that $T(X) = (X|\omega)$ for all $X \in C^\infty(U,\mathbb{R}^n)$.

Therefore $\Omega^1(U)$ can be identified with the set of all $C^\infty(U)$-linear maps from $\mathcal{X}(U)$ to $C^\infty(U)$.

More generally, there is a canonical identification between (r,s)-tensor fields on U and $C^\infty(U)$-multilinear maps from $\mathcal{X}(U)^r \times \Omega^1(U)^s \to C^\infty(U)$.

Definition 1.2.36 (Connection)**.** Let $U \subset \mathbb{R}^n$ be open. A *connection* on U is a bilinear map

$$\nabla\colon \mathcal{X}(U) \times \mathcal{X}(U) \to C^\infty(U),\, (X,Y) \mapsto \nabla_X Y$$

with the following two properties:

(a) $\nabla_{fX} Y = f\nabla_X Y$ for all $f \in C^\infty(U)$, $X, Y \in \mathcal{X}(U)$.

(b) $\nabla_X(fY) = f\nabla_X Y + X(f)Y$ for all $f \in C^\infty(U)$, $X, Y \in \mathcal{X}(U)$.

The *torsion* T_∇ of ∇ is defined as

$$T_\nabla\colon \mathcal{X}(U) \times \mathcal{X}(U) \to \mathcal{X}(U),\ (X,Y) \mapsto \nabla_X Y - \nabla_Y X - [X,Y]$$

and the *curvature* R_∇ is defined as

$$R_\nabla\colon \mathcal{X}(U)^3 \to \mathcal{X}(U),\ (X,Y,Z) \mapsto \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z.$$

Remark 1.2.37. A $C^\infty(U)$-multilinear map $T\colon \mathcal{X}(U)^r \to \mathcal{X}(U)$ can be identified with a $C^\infty(U)$-multilinear map $\tilde{T}\colon \mathcal{X}(U)^r \times \Omega^1(U) \to C^\infty(U)$ via

$$\tilde{T}(X_1,\dots,X_r,\omega) = (T(X_1,\dots,X_r)|\omega).$$

In this sense, the torsion of a connection can be viewed as a $(2,1)$-tensor and the curvature can be viewed as a $(3,1)$-tensor.

1.3 Global Analysis – Manifolds Part I

1.3.1 Topological and smooth manifolds

Definition 1.3.1 (Basis of a topology)**.** Let $(X,\mathcal{T})$ be a topological space. A subset $\mathcal{B}$ of $\mathcal{T}$ is called a *basis of the topology* $\mathcal{T}$ if for every $U \in \mathcal{T}$ and every $x \in U$ there exists $B \in \mathcal{B}$ with $x \in B$ and $B \subset U$. The space $(X,\mathcal{T})$ is called *second-countable* if it has a countable basis.

Remark 1.3.2. A rewording of this definition is that $\mathcal{B}$ is a basis if every open set is a union of elements of $\mathcal{B}$.

Proposition 1.3.3. *(a) If $(X,\mathcal{T})$ is a second-countable topological space and $Y \subset X$, then the subspace topology on Y is second-countable.*

(b) A metric space (X,d) is second-countable if and only if there exists a countable subset D of X with $\overline{D} = X$.

Example 1.3.4. The set $\mathbb{Q}^n \subset \mathbb{R}^n$ has closure $\mathbb{R}^n$ and is countable. By (b), the space $\mathbb{R}^n$ (with the Euclidean topology) is second-countable, and by (a), the same is true for every subset of $\mathbb{R}^n$.

Definition 1.3.5 (Topological manifold)**.** A topological space M is called a *topological manifold* of dimension n if it satisfies the following three properties:

(a) M is a Hausdorff space.

(b) M is locally Euclidean, that is, for every $x \in M$ there exists an open subset U_x of M containing x that is homeomorphic to an open subset of $\mathbb{R}^n$.

(c) The topology on M is second-countable.

Remark 1.3.6. If M is a subset of $\mathbb{R}^n$ (with the subspace topology), we only need to check property (b) as (a) and (c) are automatically satisfied.

Example 1.3.7. The sphere $S^n = \{(x_0, \dots, x_n) \in \mathbb{R}^{n+1} \mid \sum_j x_j^2 = 1\} \subset \mathbb{R}^{n+1}$ with the subspace topology is a topological manifold of dimension n. For simplicity, let consider $n = 1$. The stereographic projection

$$\varphi_N \colon S^1 \setminus \{(0,1)\} \to \mathbb{R},\ (x, y) \mapsto \frac{2x}{1-y}$$

is a continuous bijective map with continuous inverse

$$\varphi_N^{-1} \colon \mathbb{R} \to S^1 \setminus \{(0,1)\},\ s \mapsto \left(\frac{4s}{4+s^2}, \frac{s^2-4}{4+s^2}\right),$$

hence a homeomorphism.

Likewise, the stereographic projection

$$\varphi_S \colon S^1 \setminus \{(0,-1)\} \to \mathbb{R},\ (x, y) \mapsto \frac{2x}{1+y}$$

is a homeomorphism. Thus every point (x, y) in S^1 has an open neighborhood (namely $S^1 \setminus \{(0,1)\}$ if $(x, y) \neq (0, 1)$ and $S^1 \setminus \{(0,-1)\}$ if $(x, y) = (0, 1)$) that is homeomorphic to $\mathbb{R}$.

Example 1.3.8. The intersection of two lines, for example $M = \mathbb{R} \times \{0\} \cup \{0\} \times \mathbb{R}$ is not a topological manifold: The open subset $(0, \infty) \times \{0\}$ of M is homeomorphic to $(0, \infty) \subset \mathbb{R}$, hence M could only be a topological manifold of dimension 1. However, the open subset $U = (-1, 1) \times \{0\} \cup \{0\} \times (-1, 1)$ is not homeomorphic to any subset of $\mathbb{R}$. To see this, one needs a concept not covered in this course, namely connectedness. The subset U is connected, but if you remove $\{(0, 0)\}$, what remains has four connected components. Connected subsets of $\mathbb{R}$ however are intervals, and if you remove a point from an interval, you end up with a space with at most two connected components. Since homeomorphisms preserve connected components, U cannot be homeomorphic to a subset of $\mathbb{R}^n$.

Definition 1.3.9 (Chart, atlas)**.** Let M be a topological manifold of dimension n. A *chart* is a pair (U, φ) consisting of an open subset U of M and a homeomorphism φ from U onto an open subset of $\mathbb{R}^n$. An *atlas* is a family $((U_i, \varphi_i))_{i\in I}$ of charts such that $M \subset \bigcup_{i\in I} U_i$.

If $i, j \in I$ with $U_i \cap U_j \neq \emptyset$, the *transition map* φ_{ij} is defined by

$$\varphi_{ij}\colon \varphi_i(U_i \cap U_j) \to \varphi_j(U_i \cap U_j),\ \varphi_{ij} = \varphi_j \circ \varphi_i^{-1}.$$

Definition 1.3.10 (Smooth atlas, smooth structure)**.** Let M be a topological manifold of dimension n. An atlas $((U_i, \varphi_i))_{i\in I}$ is called a *smooth atlas* if the transition map φ_{ij} is smooth for all $i, j \in I$ such that $U_i \cap U_j \neq \emptyset$.

Two smooth atlases $\mathcal{A}$ and $\mathcal{B}$ are called *equivalent* if $\mathcal{A} \cup \mathcal{B}$ is again a smooth atlas. An equivalence class of smooth atlases is called a *smooth structure* on M and a topological manifold with a smooth structure is called a *smooth manifold*.

Example 1.3.11. Every open subset U of $\mathbb{R}^n$ is a topological manifold of dimension n. It admits a smooth atlas containing only the singly chart (U, id). The induced smooth structure is called the *standard smooth structure* on U. Unless otherwise stated, we always consider the standard smooth structure on open subsets of $\mathbb{R}^n$.

Example 1.3.12. Every homeomorphism $\varphi\colon \mathbb{R} \to \mathbb{R}$ gives rise to a chart $(\mathbb{R}, \varphi)$. If f or f^{-1} is not smooth, then the atlas with the single chart $(\mathbb{R}, \varphi)$ is not equivalent to the atlas with the single chart $(\mathbb{R}, \mathrm{id})$.

Example 1.3.13. The n-sphere $S^n = \{x \in \mathbb{R}^{n+1} : \|x\|_2 = 1\}$ is a topological manifold of dimension n. As in Example 1.3.7 let us consider the case $n = 1$. For the atlas given by $(S^1\backslash\{(0, 1)\}, \varphi_N)$ and $(S^1\backslash\{(0, -1)\}, \varphi_S)$ the transition map given by

$$\varphi_{NS}\colon \mathbb{R} \setminus \{0\} \to \mathbb{R} \setminus \{0\},\ \varphi_{NS} = \varphi_S \circ \varphi_N^{-1}$$

evaluates to

$$\varphi_S(\varphi_N^{-1}(s)) = \varphi_S\left(\frac{4s}{4+s^2}, \frac{s^2-4}{4+s^2}\right) = \frac{4}{s}$$

for $s \in \mathbb{R} \setminus \{0\}$. Hence φ_{NS} is a smooth map.

Definition 1.3.14 (Smooth map)**.** Let M, N be topological manifolds and $((U_i, \varphi))_{i\in I}$, $((V_j, \psi_j))_{j\in J}$ atlases of M and N, respectively. A map $f\colon M \to N$ is called *smooth* if for all $i \in I$, $j \in J$ with $U_i \cap f^{-1}(V_j) \neq \emptyset$ the map

$$\psi_j \circ f \circ \varphi_i^{-1}|_{U_i\cap f^{-1}(V_j)}\colon \varphi_i(U_i \cap f^{-1}(V_j)) \to \psi_j(V_j)$$

is smooth. The set of all smooth maps from M to N is denoted by $C^\infty(M, N)$. We simply write $C^\infty(M)$ for $C^\infty(M, \mathbb{R})$.

The map f is called a *diffeomorphism* if it is bijective and both f and f^{-1} are smooth.

Remark 1.3.15. This definition of smooth map depends on the chosen atlases. However, if we replace $((U_i, \varphi))_{i\in I}$ and $((V_j, \psi_j))_{j\in J}$ by equivalent atlases, we end up with the same smooth maps. In particular, there is a well-defined notion of smooth maps between smooth manifolds.

Remark 1.3.16. If $U \subset \mathbb{R}^m$ and $V \subset \mathbb{R}^n$ with the standard smooth structures, then a map $f: U \to V$ is smooth in the sense of this definition if and only if it is smooth in the sense of the previous definition as map between open subsets of normed spaces.

1.3.2 Submanifolds of $\mathbb{R}^n$

Definition 1.3.17 (Smooth submanifold of $\mathbb{R}^n$)**.** A subset M of $\mathbb{R}^m$ is called a *smooth n-dimensional submanifold of* $\mathbb{R}^m$ if there exists a family $(V_i, \Psi_i)_{i\in I}$, with $V_i \subset \mathbb{R}^n$ open and $\Psi_i: V_i \to \mathbb{R}^m$ a smooth map for all $i \in I$, that satisfies the following properties:

(a) Ψ_i is injective and $\Psi_i(V_i) \subset M$ for all $i \in I$.

(b) $\operatorname{rk} D\Psi_i(x) = n$ for all $i \in I$, $x \in V_i$.

(c) For every $i \in I$, $x_0 \in V_i$ there exist open neighborhoods V_{x_0} of x_0, $U_{\Psi_i(x_0)}$ of $\Psi_i(x_0)$ such that

$$\Psi_i|_{V_{x_0}}: V_{x_0} \to U_{\Psi_i(x_0)} \cap M$$

is a homeomorphism.

(d) $M = \bigcup_{i\in I} \Psi_i(V_i)$.

Remark 1.3.18. The maps Ψ_i are called *parametrizations* of M. Note that they go in the opposite direction (from $\mathbb{R}^n$ into the manifold) from charts.

Remark 1.3.19. Every smooth n-dimensional submanifold of $\mathbb{R}^n$ is a smooth n-dimensional manifold with smooth atlas $(U_{\Psi_i(x_0)} \cap M, \Psi_i|_{V_{x_0}}^{-1})_{i\in I, x_0\in V_i}$. The converse is also true in the following sense: Every smooth n-dimensional manifold can be smoothly embedded into $\mathbb{R}^{2n}$. This is known as Whitney's embedding theorem. Note however, that this embedding is by no means canonical.

Proposition 1.3.20. *Let $F: \mathbb{R}^m \to \mathbb{R}^{m-n}$ be a smooth map. If $c \in \mathbb{R}^{m-n}$ is a regular value of F, then $F^{-1}(c)$ is a smooth n-dimensional submanifold of $\mathbb{R}^m$.*

Example 1.3.21. Let $F\colon \mathbb{R}^{n+1} \to \mathbb{R}$, $x \mapsto \sum_{j=1}^{n+1}(x^j)^2$. Clearly, F is smooth. We have $DF(x) = \sum_{j=0}^{n} 2x^j dx^j$. By the rank-nulllity theorem, any non-zero linear functional on $\mathbb{R}^{n+1}$ is surjective. Hence 1 is a regular value of F. It follows that

$$S^n = F^{-1}(1) = \{x \in \mathbb{R}^{n+1} \mid (x^1)^2 + \cdots + (x^{n+1})^2 = 1\}$$

is smooth n-dimensional submanifold of $\mathbb{R}^{n+1}$.

Example 1.3.22. Let $\mathbb{R}^{n\times n}_{\mathrm{sym}}$ denote the symmetric $n \times n$ matrices. Observe that $\mathbb{R}^{n\times n}_{\mathrm{sym}}$ is a real vector space of dimension $\frac{1}{2}n(n+1)$. Let

$$F\colon \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}_{\mathrm{sym}},\ A \mapsto A^{\mathrm{T}}A.$$

Again, it is not hard to see that F is smooth and $DF(A)[H] = A^{\mathrm{T}}H + H^{\mathrm{T}}A$. If $A^{\mathrm{T}}A = 1$ and $B \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$, then

$$DF(A)[AB/2] = \frac{1}{2}A^{\mathrm{T}}AB + B^{\mathrm{T}}A^{\mathrm{T}}A = B.$$

Thus $DF(A)$ is surjective whenever $A^{\mathrm{T}}A = 1_n$, which means that 1_n is a regular value of F.

It follows that

$$\mathrm{O}_n(\mathbb{R}) = F^{-1}(1_n) = \{A \in \mathbb{R}^{n\times n} \mid A^{\mathrm{T}}A = 1_n\}$$

is a smooth submanifold of $\mathbb{R}^{n\times n}$ of dimension $\frac{1}{2}n(n-1)$.

1.3.3 Tangent spaces

Definition 1.3.23 (Tangent vector)**.** Let M be a smooth manifold and $p \in M$. A *tangent vector at* p is a linear map $X_p\colon C^\infty(M) \to \mathbb{R}$ that satisfies the Leibniz rule (or product rule)

$$X_p(fg) = X_p(f)g(p) + f(p)X_p(f)$$

for all $f, g \in C^\infty(M)$. The set of all tangent vectors at p is denoted by T_pM.

Remark 1.3.24. If M is an open subset of $\mathbb{R}^n$ (with the standard smooth structure), then a tangent vector at p is exactly what we called a derivation at p in a prior section.

Lemma 1.3.25. *Let M be a smooth manifold and $p \in M$. If $f \in C^\infty(M)$ is constant on an open set containing p and $X_p \in T_pM$, then $X_p(f) = 0$.*

As a consequence of the previous lemma, whenever U is an open neighborhood of $p \in M$ and $X_p \in T_pM$, then X_p restricts to a derivation on $C^\infty(U)$ at p and this restriction is consistent in the following sense: If V is another open neighborhood of $p \in M$ and $f \in C^\infty(M)$, then $X_p|_{C^\infty(U)}(f|_U) = X_p|_{C^\infty(V)}(f|_V)$.

For open subsets of $\mathbb{R}^n$, we have seen that tangent vectors at a point are the same as directional derivatives at that point. The same is true for smooth manifolds, but we need a more sophisticated concept of directional derivative as there are no longer "straight lines".

Proposition 1.3.26. *Let M be a smooth manifold and $p \in M$. For every $X_p \in T_pM$ there exists $\varepsilon > 0$ and a smooth map $\gamma\colon (-\varepsilon, \varepsilon) \to M$ such that $\gamma(0) = p$ and*

$$X_p(f) = \frac{d}{dt}(f \circ \gamma)(0)$$

for all $f \in C^\infty(M)$.

Proof. Let (U, φ) be a chart (from a smooth atlas in the smooth structure of M) with $p \in U$ and let $q = \varphi(p)$. Consider the map

$$\tilde{X}_q\colon C^\infty(\varphi(U)) \to \mathbb{R},\ f \mapsto X_p(f \circ \varphi).$$

It is easy to see that $\tilde{X}_q$ is a derivation at q. Hence there exist $v \in \mathbb{R}^n$ such that $\tilde{X}_q = \partial_v|_q$. For $\varepsilon > 0$ sufficiently small let

$$\gamma\colon (-\varepsilon, \varepsilon) \to M,\ t \mapsto \varphi^{-1}(q + tv).$$

Then $\gamma(0) = p$ and

$$\frac{d}{dt}(f \circ \gamma)(0) = \frac{d}{dt}(f \circ \varphi^{-1})(q + tv)(0) = \tilde{X}_q(f \circ \varphi^{-1}) = X_p(f). \qquad \square$$

Remark 1.3.27. Unlike in the case of open subsets of $\mathbb{R}^n$, the correspondence in the previous proposition is not one-to-one. A trivial observation is that if we restrict a smooth curve γ to a smaller interval, we still get the same tangent vector. But beyond that, there can be "genuinely different" curves that result in the same tangent vector, for example $\gamma_1, \gamma_2\colon (-1, 1) \to \mathbb{R}$ with $\gamma_1(t) = t^2$ and $\gamma_2(t) = -t^4$. The reason is that we do not have a canonical choice of a curve in a given direction like the straight lines in Euclidean space.

Example 1.3.28. Let M be a smooth n-dimensional submanifold of $\mathbb{R}^m$. Every smooth map $\gamma\colon (-\varepsilon, \varepsilon) \to M$ can be viewed as a map with values in $\mathbb{R}^m$. As such, the derivative at zero is a linear map from $\mathbb{R}$ to $\mathbb{R}^m$, which can be identified with a vector in $\mathbb{R}^m$.

If U is an open neighborhood of M and $f\colon U \to \mathbb{R}$ is smooth, then the tangent vector from the previous proposition satisfies

$$X_p(f) = \frac{d}{dt}(f \circ \gamma)(0) = Df(p)[\dot{\gamma}(0)] = \partial_{\dot{\gamma}(0)} f(p).$$

Hence every element of T_pM is a directional derivative in a direction "tangent" to M. This explains the notion of tangent space in this abstract setting.

Definition 1.3.29 (Differential)**.** Let M, N be smooth manifolds, $p \in M$ and $\varphi \in C^\infty(M, N)$. For $X_p \in T_pM$ the *pushforward* $\varphi_* X_p \in T_{\varphi(p)}N$ is defined by

$$\varphi_* X_p \colon C^\infty(N) \to \mathbb{R},\ g \mapsto X_p(g \circ \varphi).$$

The *differential* of φ at p is the map

$$D\varphi(p)\colon T_pM \to T_{\varphi(p)}N,\ X_p \mapsto \varphi_* X_p.$$

If $N = \mathbb{R}$, we also write $d\varphi(p)$ for $D\varphi(p)$.

Example 1.3.30. Let $U \subset \mathbb{R}^m$, $V \subset \mathbb{R}^n$ be open and $\varphi\colon U \to V$ be smooth. For $p \in U$ the spaces T_pU and $T_{\varphi(p)}V$ are the directional derivatives in p and $\varphi(p)$, respectively.

If $h \in \mathbb{R}^m$, then

$$\varphi_* \partial_h|_p \colon C^\infty(V) \to \mathbb{R},\ g \mapsto \partial_h(g \circ \varphi)|_p = \partial_{D\varphi(p)[h]} g|_{\varphi(p)},$$

where we used $D\varphi(p)$ to denote the derivative of φ at p defined before.

Hence if we make the identification $\mathbb{R}^m \xrightarrow{\cong} T_pU$, $h \mapsto \partial_h|_p$ and likewise for $T_{\varphi(p)}V$, then the differential $D\varphi(p)$ is the same as the derivative of φ at p.

Definition 1.3.31 (Tangent bundle)**.** Let M be a smooth n-dimensional manifold. The *tangent bundle* TM of M is defined as

$$TM = \bigcup_{p \in M} \{p\} \times T_pM.$$

The canonical projection $\pi\colon TM \to M$ is defined by $\pi(p, X_p) = p$ for all $p \in M$, $X_p \in T_pM$.

Proposition 1.3.32. *Let M be a smooth n-dimensional manifold with atlas $((U_i, \varphi_i))_{i \in I}$. For $i \in I$ let*

$$\tilde{\varphi}_i \colon \pi^{-1}(U_i) \to \varphi_i(U) \times \mathbb{R}^n,\ (p, v) \mapsto (\varphi_i(p), D\varphi_i(p)[v]).$$

The set

$$\mathcal{T}_{TM} = \{W \subset TM \mid \tilde{\varphi}_i(W \cap \pi^{-1}(U_i)) \text{ open in } \mathbb{R}^n \times \mathbb{R}^n \text{ for all } i \in I\}$$

is a second countable Hausdorff topology on TM that makes TM into an n-dimensional topological manifold.

Moreover, $((\pi^{-1}(U_i), \tilde{\varphi}_i))_{i\in I}$ is a smooth atlas on TM and equivalent atlases on M gives rise to equivalent atlases on TM.

Remark 1.3.33.

- Intuitively, the tangent bundle is the collection of tangent spaces, assembled in a way that the tangent space varies smoothly with the base point. This proposition makes this intuition rigorous.
- As a consequence of the previous proposition, the tangent bundle of a smooth n-dimensional manifold carries a canonical structure of a smooth $2n$-dimensional manifold.
- The tangent bundle carries additional structure: For every $p \in M$, the set $\pi^{-1}(p) = T_pM$ is an n-dimensional vector space, and whenever (φ, U) is a chart of M, then $D\varphi_i(p)$ is a bijective linear map from T_pM onto $\mathbb{R}^n$. This can be summarized as saying that TM is a *smooth vector bundle* of rank n over M.
- The space of *smooth sections* in the vector bundle TM is defined as

  $$\Gamma(TM) = \{s \in C^\infty(M, TM) \mid \pi \circ s = \mathrm{id}_M\}.$$

  This space of smooth sections can be canonically identified with $\mathcal{X}(M)$.

Definition 1.3.34 (Cotangent bundle)**.** Let M be a smooth n-dimensional manifold. If $p \in M$, we write T_p^*M for $(T_pM)^*$. The *cotangent bundle* T^*M of M is defined as

$$T^*M = \bigcup_{p\in M} \{p\} \times T_p^*M.$$

The canonical projection $\pi\colon T^*M \to M$ is defined as $\pi(p, \omega_p) = p$ for all $p \in M$, $\omega_p \in T_pM$.

Remark 1.3.35. The cotangent bundle can be turned into a smooth $2n$-dimensional manifold and a smooth vector bundle of rank n over M in much the same way as TM, one just has to adapt the definition of $\tilde{\varphi}_i$ to be

$$\tilde{\varphi}_i\colon \pi^{-1}(U_i) \to \varphi_i(U) \times \mathbb{R}^n,\ (p, \omega) \mapsto (\varphi_i(p), \omega \circ D\varphi_i(p)^{-1}).$$

Definition 1.3.36 (Differential 1-form)**.** Let M be a smooth n-dimensional manifold. The set of sections $\Gamma(T^*M)$ is denoted by $\Omega^1(M)$. An element $\omega \in \Omega^1(M)$ is called a *differential* 1*-form*.

Remark 1.3.37. We have seen that $T_p\mathbb{R}^n$ can be canonically identified with $\mathbb{R}^n$. Thus, if $f \in C^\infty(M)$ and $p \in M$, then $df(p)$ can be viewed as a linear map from T_pM to $\mathbb{R}$, or, in other words, an element of T_p^*M. One can check that

$$df\colon M \to T^*M,\, p \mapsto (p, df(p))$$

is smooth. In other words, $df \in \Gamma(T^*M) = \Omega^1(M)$.

Remark 1.3.38. Let M be a smooth n-dimensional manifold and (φ, U) a chart. We write $(x^1, \dots, x^n)$ for the coordinate maps of φ, that is, $\varphi(p) = (x^1(p), \dots, x^n(p))$ for $p \in U$. The maps $x^1, \dots, x^n$ are smooth maps from U to $\mathbb{R}$, hence $dx^1(p), \dots, dx^n(p) \in T_p^*M$. This is consistent with our notation of the dual basis used earlier. In fact, since φ is a diffeomorphism onto its image, $dx^1(p), \dots, dx^n(p)$ form a basis of T_p^*M for every $p \in U$.

1.4 Global Analysis – Manifolds Part II

1.4.1 Differential forms

Definition 1.4.1 (Exterior power of the cotangent bundle)**.** Let M be an n-dimensional smooth manifold. For $r \in \mathbb{N}$ we define r-th exterior power of the cotangent bundle as

$$\Lambda^r T^*M = \bigcup_{p \in M} \{p\} \times \Lambda^r T_p^*M.$$

Moreover, we let $\Lambda T_p^*M = \bigoplus_{k=0}^n \Lambda^r T_p^*M$ and

$$\Lambda T^*M = \bigcup_{p \in M} \{p\} \times \Lambda T_p^*M.$$

Remark 1.4.2. Just like the cotangent bundle, the r-th exterior power of the cotangent bundle has a natural structure of a smooth manifold such that the projection map $\pi\colon \Lambda^r T^*M \to M$, $(p, \alpha_p) \mapsto p$ is smooth and the fibers $\pi^{-1}(p)$ are real vector spaces, in this case of dimension $\binom{n}{r}$. In other words, $\Lambda^r T^*M$ is a smooth vector bundle (of rank $\binom{n}{r}$) over M. The same is true for ΛT^*M (with rank 2^n).

Remark 1.4.3. Recall that we defined $\Lambda^0 T_p^*M = \mathbb{R}$. Thus $\Omega^0(M) = C^\infty(M)$.

Definition 1.4.4 (Differential form)**.** Let M be a smooth n-dimensional manifold and $r \in \mathbb{N}$. The set of smooth sections $\Gamma(\Lambda^r T^*M)$ is denoted by $\Omega^r(M)$. An element of $\Omega^r(M)$ is called a *differential r-form* on M. The set of smooth sections $\Gamma(\Lambda T^*M)$ is denoted by $\Omega(M)$. An element of $\Omega(M)$ is called *differential form* on M.

Remark 1.4.5. Note that not every differential is a differential r-form for some r. For example, if $f \in C^\infty(M)$ and $\omega \in \Omega^1(M)$, then $f + \omega$ is a differential form, but not a differential r-form for any r unless $\omega = 0$ or $f = 0$.

Example 1.4.6. Let $U \subset \mathbb{R}^n$ be open. With the notation from the previous section,

$$\Omega^r(U) = \left\{ \sum_{1 \leq j_1 < \cdots < j_r \leq n} \alpha_{j_1,\ldots,j_r} dx^{j_1} \wedge \cdots \wedge dx^{j_r} : \alpha_{j_1,\ldots,j_r} \in C^\infty(U) \right\}.$$

Remark 1.4.7. The wedge product on alternating forms can be extended to a wedge product on differential forms by defining $(\alpha \wedge \beta)_p = \alpha_p \wedge \beta_p$ for $\alpha \in \Omega^r(M)$, $\beta \in \Omega^s(M)$, $p \in M$. Likewise, we can define the product of a smooth function and a differential form by $(f\alpha)_p = f(p)\alpha(p)$.

Definition 1.4.8 (Pull-back)**.** Let M, N be smooth manifolds and $\varphi \in C^\infty(M, N)$. The *pull-back* operation φ^* is defined as

$$\varphi^* \colon \Omega^r(N) \to \Omega^r(M), (\varphi^*\omega)_p(v_1, \ldots, v_r) = \omega_{\varphi(p)}(D\varphi(p)[v_1], \ldots, D\varphi(p)[v_r])$$

for $p \in M$, $v_1, \ldots, v_r \in T_pM$.

Lemma 1.4.9. *Pull-back is compatible with wedge products: If M, N are smooth manifolds, $\varphi \in C^\infty(M, N)$ and $\alpha \in \Omega^r(N)$, $\beta \in \Omega^s(N)$, then*

$$\varphi^*(\alpha \wedge \beta) = \varphi^*\alpha \wedge \varphi^*\beta.$$

Proof. If $p \in M$, then

$$\begin{aligned}
\varphi^*(\alpha \wedge \beta)_p &= (\alpha_{\varphi(p)} \wedge \beta_{\varphi(p)}) \circ D\varphi(p)^{\otimes(r+s)} \\
&= \frac{(r+s)!}{r!s!} P_\wedge(\alpha_{\varphi(p)} \otimes \beta_{\varphi(p)}) \circ D\varphi(p)^{\otimes(r+s)} \\
&= \frac{(r+s)!}{r!s!} P_\wedge((\alpha_{\varphi(p)} \circ D\varphi(p)^{\otimes r}) \otimes (\beta_{\varphi(p)} \circ D\varphi(p)^{\otimes s})) \\
&= (\varphi^*\alpha)_p \wedge (\varphi^*\beta)_p.
\end{aligned}$$ ☐

As discussed before, the differential df of a smooth function is a differential 1-form. One can extend this notion of differential to produce a differential $(r+1)$-form from a differential r-form in the following way.

Proposition 1.4.10 (Exterior derivative)**.** *Let M be a smooth n-dimensional manifold. There exists a unique linear map $d\colon \Omega(M) \to \Omega(M)$ such that*

- $d(\Omega^r(M)) \subset \Omega^{r+1}(M)$ *for all* $r \in \{0, \dots, n\}$,
- $df = Df$ *for all* $f \in \Omega^0(M)$,
- $d(f\alpha) = fd\alpha + df \wedge \alpha$ *for all* $f \in \Omega^0(M)$, $\alpha \in \Omega^r(M)$, $r \in \{0, \dots, n\}$.
- $d^2 = 0$.

The map d is called the exterior derivative.

It has the following properties:

- *If* $\varphi \in C^\infty(M, N)$, *then* $d \circ \varphi^* = \varphi^* \circ d$.
- *If* $\alpha \in \Omega^r(M)$, $\beta \in \Omega^s(M)$, *then* $d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^r \alpha \wedge d\beta$.

Remark 1.4.11.

- As the differential, the exterior derivative is local: If $\alpha, \beta \in \Omega^r(M)$ coincide on an open subset U of M, then $d\alpha$ and $d\beta$ coincide on U.
- The graded Leibniz rule together with locality give a recipe to compute the exterior derivative in local coordinates: If (φ, U) is a chart, then

$$d\left(\sum_{j_1<\dots<j_r} f_{j_1\dots j_r} dx^{j_1} \wedge \dots \wedge dx^{j_r}\right) = \sum_{i=1}^n \sum_{j_1<\dots<j_r} \frac{\partial f_{j_1\dots j_r}}{\partial x^i} dx^i \wedge dx^{j_1} \wedge \dots \wedge dx^{j_r}$$

on U.

Definition 1.4.12 (Contraction)**.** Let M be a smooth n-dimensional manifold, $\alpha \in \Omega^{r+1}(M)$ and $X \in \mathcal{X}(M)$. The *contraction* of X into α is the r-form $i_X\alpha$ defined by

$$(i_X\alpha)_p = \alpha_p(X_p, \cdot\,)$$

for $p \in M$.

Example 1.4.13. If $f \in C^\infty(M)$ and $X \in \mathcal{X}(M)$, then $i_X(df) = df(X) = X(f)$.

Definition 1.4.14 (Lie derivative)**.** Let M be a smooth n-dimensional manifold, $X \in \mathcal{X}(M)$ with local flow Φ_X and $\alpha \in \Omega^r(M)$. The *Lie derivative* of α along X is defined as

$$L_X\alpha = \lim_{t\to 0} \frac{(\Phi_X^t)^*\alpha - \alpha}{t}.$$

Remark 1.4.15. As in the case of vector fields, some care has to be taken in this definition as Φ_X^t is usually not a global diffeomorphism. Note that in contrast to the Lie derivative of vector fields, we take the pull-back instead of the pushforward along the flow.

As the Lie derivative of a vector field along a vector field, there is a simpler algebraic formula for the Lie derivative of a differential form.

Proposition 1.4.16 (Cartan's homotopy formula)**.** *Let M be a smooth n-dimensional manifold. If $X \in \mathcal{X}(M)$ and $\alpha \in \Omega^r(M)$, then*

$$L_X\alpha = d(i_X\alpha) + i_X(d\alpha).$$

1.4.2 The de Rham complex and vector calculus

Let M be an n-dimensional smooth manifold. The exterior derivative can be diagrammatically written as

$$\{0\} \to \Omega^0(M) \xrightarrow{d} \Omega^1(M) \xrightarrow{d} \dots \xrightarrow{d} \Omega^n(M) \to \{0\},$$

where the composition of any two adjacent arrows is zero. This diagram is abbreviated as $(\Omega^\bullet(M), d)$ and called the *de Rham complex*. It is a fundamental example of a *cochain complex*.

Definition 1.4.17 (Closed and exact forms, de Rham cohomology)**.** Let M be an n-dimensional smooth manifold. A differential form $\alpha \in \Omega(M)$ is called *closed* if $d\alpha = 0$ and *exact* if there exists $\beta \in \Omega(M)$ such that $\alpha = d\beta$.

The r-th *de Rham cohomology group* $H^r_{\mathrm{dR}}(M)$ is defined as

$$H^r_{\mathrm{dR}}(M) = (\ker d \cap \Omega^r(M))/(\operatorname{ran} d \cap \Omega^r(M)).$$

Remark 1.4.18. Since $d^2 = 0$, every exact differential form is closed and thus $\operatorname{ran} d \cap \Omega^r(M) \subset \ker d \cap \Omega^r(M)$.

Example 1.4.19. If $U \subset \mathbb{R}^n$ is convex, then $H^r_{\mathrm{dR}}(U) = \{0\}$ for every $r \in \{0, \dots, n\}$.

Example 1.4.20. Every differential 1-form on S^1 is closed since $\Omega^2(S^1) = \{0\}$. However, not every 1-form is exact: Recall that we can identify T_vS^1 with $\{w \in \mathbb{R}^2 \mid v \in S^1, v \cdot w = 0\}$. Let

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

and

$$\omega_v \colon T_vS^1 \to \mathbb{R},\ w \mapsto Av \cdot w.$$

One can show that $\omega \in \Omega^1(S^1)$ and there exists no function $f \in C^\infty(S^1)$ such that $df = \omega$. This is easy to see once one has Stokes's theorem at hand: If $\omega = df$, then $\int_{S^1} \omega = 0$, which is not the case.

Note that this is a global statement: For every chart (φ, U) there exists a function $f \in C^\infty(U)$ such that $\omega|_U = df|_U$, but such a function cannot be extended to a smooth function on S^1 that still obeys this identity.

Definition 1.4.21 (Tensor field)**.** For $r, s \in \mathbb{N}$ we define

$$(T^*M)^{\otimes r} \otimes (TM)^{\otimes s} = \bigcup_{p \in M} \{p\} \times (T_p^*M)^{\otimes r} \otimes (T_pM)^{\otimes s}$$

and

$$\pi\colon (T^*M)^{\otimes r} \otimes (TM)^{\otimes s} \to M,\ (p, T_p) \mapsto p.$$

There is a canonical structure of a smooth manifold on $(T^*M)^{\otimes r} \otimes (TM)^{\otimes s}$ such that π is smooth and $(T^*M)^{\otimes r} \otimes (TM)^{\otimes s}$ becomes a smooth vector bundle over M. A smooth section $T \in \Gamma((T^*M)^{\otimes r} \otimes (TM)^{\otimes s})$ is called an (r, s)-*tensor field* on M.

Example 1.4.22.

- A $(0, 1)$ tensor field is a vector field and a $(1, 0)$ tensor field is a differential 1-form. Every differential r-form is a $(r, 0)$ tensor field, but not vice versa – tensor fields do not have to be alternating.

Definition 1.4.23. Let M be a smooth manifold. A *Riemannian metric* on M is a $(2, 0)$ tensor field such that g_p is an inner product on T_pM for every $p \in M$. A *Riemannian manifold* is a pair (M, g) consisting of a smooth manifold M and a Riemannian metric g on M.

Example 1.4.24. Let M be a smooth n-dimensional submanifold of $\mathbb{R}^m$. Recall that one can identify T_pM with the subspace

$$\{\dot\gamma(0) \mid \gamma \in C^\infty((-\varepsilon, \varepsilon), \mathbb{R}^m),\ \gamma(0) = p, \operatorname{im}\gamma \subset M\}$$

of $\mathbb{R}^m$. A Riemannian metric on M is defined by

$$g_p\colon T_pM \times T_pM \to \mathbb{R},\ (v, w) \mapsto \langle v, w \rangle,$$

where $\langle\,\cdot\,,\,\cdot\,\rangle$ is any inner product on $\mathbb{R}^m$.

Definition 1.4.25 (Musical isomorphisms)**.** Let (M, g) be a Riemannian metric. For $p \in M$, the musical isomorphism $^\flat$ is defined as

$$^\flat\colon T_pM \to T_p^*M,\ v \mapsto v^\flat = g_p(v, \cdot).$$

Moreover, $^\sharp$ is the inverse of $^\flat$.

These isomorphisms are extended to maps between $\mathcal{X}(M)$ and $\Omega^1(M)$ by

$$(X^\flat)_p = (X_p)^\flat, \qquad (\omega^\sharp)_p = (\omega_p)^\sharp$$

for $X \in \mathcal{X}(M)$, $\omega \in \Omega^1(M)$, $p \in M$.

Remark 1.4.26. If $v^\flat = 0$,then $g_p(v, v) = 0$, hence $v = 0$. Thus $^\flat$ is injective, and since T_pM and T_p^*M have the same dimension, also surjective. Therefore $^\flat$ is really an isomorphism and the inverse $^\sharp$ is well-defined.

Remark 1.4.27. Let (φ, U) be a chart with $\varphi = (x^1, \dots, x^n)$ and $p \in U$. If g is a Riemannian metric and $g_p\left(\frac{\partial}{\partial x^j}|_p, \frac{\partial}{\partial x^k}|_p\right) = g_{jk}(p)$, then

$$\left(\sum_{j=1}^n \lambda^j \frac{\partial}{\partial x^j}\bigg|_p\right)^\flat = \sum_{j,k=1}^n g_{jk}(p)\lambda^j dx^k|_p$$

$$\left(\sum_{j=1}^n \mu_j dx^j|_p\right)^\sharp = \sum_{j,k=1}^n g^{jk}(p)\mu_j \frac{\partial}{\partial x^k}\bigg|_p$$

Here $(g^{jk}(p))_{jk}$ denotes the inverse matrix of $(g_{jk}(p))_{jk}$. This is what is known as “lowering and raising the indices” in physics.

Definition 1.4.28 (Gradient)**.** Let (M, g) be a Riemannian manifold. The gradient of $f \in C^\infty(M)$ is defined as $\nabla_g f = (df)^\sharp$.

Remark 1.4.29. With the notation from Remark 1.4.27, the gradient of $f \in C^\infty(M)$ is given by

$$(\nabla_g f)_p = \sum_{j,k} g^{jk}(p)\frac{\partial f(p)}{\partial x^j}\frac{\partial}{\partial x^k}\bigg|_p.$$

Definition 1.4.30 (Volume form)**.** Let M be a smooth n-dimensional manifold. A *volume form* on M is a differential n-form $\mathrm{vol} \in \Omega^n(M)$ such that $\mathrm{vol}_p \neq 0$ for all $p \in M$.

Remark 1.4.31. Not every smooth manifold admits a volume form. Those that do are called *orientable*.

Lemma 1.4.32. *Let M be a smooth n-dimensional manifold. If $\mathrm{vol} \in \Omega^n(M)$ is a volume form, then the maps*

$$C^\infty(M) \to \Omega^n(M), f \mapsto f\mathrm{vol}$$
$$\mathcal{X}(M) \to \Omega^{n-1}(M), X \mapsto i_X\mathrm{vol}$$

are isomorphisms.

Definition 1.4.33 (Divergence)**.** Let M be a smooth n-dimensional manifold with volume form $\mathrm{vol} \in \Omega^n(M)$. The divergence operator is defined by

$$\operatorname{div}\colon \mathcal{X}(M) \to C^\infty(M),\ (\operatorname{div} X)\mathrm{vol} = d(i_X \mathrm{vol})$$

Moreover, if $n = 3$ and g is a Riemannian metric on M, then the *curl operator* is defined by

$$\mathrm{curl}_g\colon \mathcal{X}(M) \to \mathcal{X}(M),\ i_{\mathrm{curl}_g X}\mathrm{vol} = d(X^\flat).$$

Remark 1.4.34. The previous lemma shows that div and curl_g are well-defined. Note that we only need a volume form to defined div, while we need a Riemannian metric to define curl_g, and the latter only makes sense in dimension 3.

Let (M, g) be a 3-dimensional Riemannian manifold. The definitions of ∇_g, curl_g and div can be summarized by the following commutative diagram:

$$\begin{tikzcd}
\{0\} \arrow[r] & \Omega^0(M) \arrow[r, "d"] & \Omega^1(M) \arrow[r, "d"] & \Omega^2(M) \arrow[r, "d"] & \Omega^3(M) \arrow[r] & \{0\} \\
\{0\} \arrow[r] & C^\infty(M) \arrow[r, "\nabla_g"] \arrow[u, "\mathrm{id}"] & \mathcal{X}(M) \arrow[r, "\mathrm{curl}_g"] \arrow[u, "\flat"] & \mathcal{X}(M) \arrow[r, "\mathrm{div}"] \arrow[u, "i_\bullet \mathrm{vol}"] & C^\infty(M) \arrow[r] \arrow[u, "\bullet\,\mathrm{vol}"] & \{0\}
\end{tikzcd}$$

In particular, $d^2 = 0$ implies $\mathrm{curl}_g \circ \nabla_g = 0$ and $\operatorname{div} \circ \mathrm{curl}_g = 0$.

Example 1.4.35. Let $U \subset \mathbb{R}^n$ be open and let g be the Riemannian metric on U described in Example 1.4.24. Note that with this choice of Riemannian metric, we have

$$g_p\left(\left.\frac{\partial}{\partial x^j}\right|_p, \left.\frac{\partial}{\partial x^k}\right|_p\right) = \delta_{jk}$$

for all $j, k \in \{1, \ldots, n\}$.

The musical isomorphisms are given by

$$\begin{aligned}
{}^\flat\colon T_pU \to T_p^*U,\ \sum_{j=1}^n \lambda^j \left.\frac{\partial}{\partial x^j}\right|_p &\mapsto \sum_{j=1}^n \lambda^j dx^j|_p \\
{}^\sharp\colon T_p^*U \to T_pU,\ \sum_{j=1}^n \mu_j dx^j|_p &\mapsto \sum_{j=1}^n \mu_j \left.\frac{\partial}{\partial x^j}\right|_p.
\end{aligned}$$

Thus

$$\nabla_g f = \sum_{j=1}^n \frac{\partial f}{\partial x^j}\frac{\partial}{\partial x^j}.$$

Moreover, if we take the volume form $dx^1 \wedge \cdots \wedge dx^n$, we obtain

$$\operatorname{div}\left(\sum_{j=1}^{n} f^j \frac{\partial}{\partial x^j}\right) = \sum_{j=1}^{n} \frac{\partial f^j}{\partial x^j}.$$

Finally, if $n = 3$, then

$$\operatorname{curl}_g\left(\sum_{j=1}^{3} f^j \frac{\partial}{\partial x^j}\right) = (\partial_2 f^3 - \partial_3 f^2)\partial_1 + (\partial_3 f^1 - \partial_1 f^3)\partial_2 + (\partial_1 f^2 - \partial_2 f^1)\partial_3,$$

where we used the shortened notation $\partial_j = \frac{\partial}{\partial x^j}$.

1.4.3 Integration of differential forms

Recall that two bases $(v_j)_{1\le j\le n}$ and $(w_j)_{1\le j\le n}$ of a real vector space V have the same orientation if there exists $\lambda > 0$ such that $v^1 \wedge \cdots \wedge v^n = \lambda w^1 \wedge \cdots \wedge w^n$. This defines an equivalence relation on the set of (ordered) bases of V with two equivalence classes. An orientation of V is a choice of an equivalence class.

Definition 1.4.36. A linear map $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ is called *orientation-preserving* if it is invertible and $(\varphi(e_j))_{1\le j\le n}$ has the same orientation as $(e_j)_{1\le j\le n}$.

Let $U, V \subset \mathbb{R}^n$ be open. A diffeomorphism $\varphi\colon U \to V$ is called *orientation-preserving* if $D\varphi(p)$ is orientation-preserving for every $p \in U$.

Lemma 1.4.37. *A linear map $\varphi\colon \mathbb{R}^n \to \mathbb{R}^n$ is orientation-preserving if and only if* $\det \varphi > 0$.

Definition 1.4.38 (Orientable manifold)**.** Let M be a smooth manifold. A smooth atlas $((\varphi_i, U_i))_{i\in I}$ is called an *orientation* of M if for all $i, j \in I$ such that $U_i \cap U_j \neq \emptyset$ the transition maps

$$\varphi_{ij}\colon \varphi_i(U_i \cap U_j) \to \varphi_j(U_i \cap U_j),\ \varphi_{ij} = \varphi_j \circ \varphi_i^{-1}$$

are orientation-preserving diffeomorphisms. A smooth manifold with a choice of an orientation is called *oriented manifold*. A chart (ψ, V) in an oriented manifold is called *positively oriented* if the union of $((\varphi_i, U_i))_{i\in I}$ and (ψ, V) is again an orientation.

If M admits an orientation, it is called *orientable*.

Remark 1.4.39. In the light of the previous lemma, an atlas is an orientation if and only if the differential of the transition maps has positive determinant at every point.

As remarked before, orientable manifolds are exactly those that admit a volume form. We can prove one of these two implications now.

Lemma 1.4.40. *If a smooth manifolds admits a volume form, then it is orientable.*

Proof. Let M be a smooth n-dimensional manifold and $\mathrm{vol} \in \Omega^n(M)$ a volume form. For any chart (φ, U) the pull-back $(\varphi^{-1})^*\mathrm{vol}$ is of the form $g dx^1 \wedge \cdots \wedge dx^n$ for some $g \in C^\infty(\varphi(U))$. Since vol vanishes nowhere, the function g vanishes nowhere. Moreover, we can change the sign of g by switching two coordinates. Thus for every point $p \in M$ there exists a chart (φ_p, U_p) such that $(\varphi^{-1})^*\mathrm{vol} = g dx^1 \wedge \cdots \wedge dx^n$ with $g > 0$. It is not hard to check that $((\varphi_p, U_p))_{p \in U}$ is an orientation of M. □

We want to define the integral of differential forms. To avoid running into integrability issues, we will restrict ourselves to differential forms with compact support.

Definition 1.4.41 (Support of a differential form)**.** Let M be a smooth manifold. If $\omega \in \Omega(M)$, the *support* of ω is defined as

$$\operatorname{supp} \omega = \overline{\{p \in M \mid \omega_p \neq 0\}}.$$

The set of all compactly supported differential forms (resp. different r-forms) is denoted by $\Omega_c(M)$ (resp. $\Omega_c^r(M)$). We also write $C_c^\infty(M)$ for $\Omega_c^0(M)$.

Definition 1.4.42 (Integral of top-level differential forms)**.** Let $U \subset \mathbb{R}^n$ be open. We define the integral of compactly supported n-forms on U by

$$\int_U : \Omega_c^n(U) \to \mathbb{R},\ f dx^1 \wedge \cdots \wedge dx^n \mapsto \int_U f\, d(x^1, \ldots, x^n),$$

where the integral on the right side is the usual Riemann (or Lebesgue) integral of functions on U.

Remark 1.4.43. Every differential form $\omega \in \Omega^n(U)$ is of the form $f dx^1 \wedge \cdots \wedge dx^n$ with $f \in C^\infty(U)$. Moreover, the support of $f dx^1 \wedge \cdots \wedge dx^n$ is easily seen to be the same as the support of f. Thus the integral of compactly supported n-forms on U is well-defined.

Remark 1.4.44. For the definition of the integral of n-forms, it is important that we express ω as $f dx^1 \wedge \cdots \wedge dx^n$ and not $f dx^2 \wedge dx^1 \wedge \cdots \wedge dx^n$, for example. This means that the integral takes the orientation of U into account.

Lemma 1.4.45. *Let $U, V \subset \mathbb{R}^n$ be open. If $\varphi\colon U \to V$ is an orientation-preserving diffeomorphism and $\omega \in \Omega_c^n(V)$, then*

$$\int_V \omega = \int_U \varphi^*\omega.$$

Proof. We write $(x^1, \dots, x^n)$ for the coordinates of U and $(y^1, \dots, y^n)$ for the coordinates of V. Say $\omega = g dy^1 \wedge \cdots \wedge dy^n$ with $g \in C_c^\infty(V)$. We need to compute the pull-back $\varphi^*\omega$. Since $\dim \Lambda^n T_p^* U = 1$, we know that

$$(\varphi^*\omega)_p = f(p) dx^1 \wedge \cdots \wedge dx^n|_p$$

for some $f \in C^\infty(U)$. By definition of the pullback,

$$\begin{aligned}(\varphi^*\omega)_p[\partial_{x^1}, \dots, \partial_{x^n}] &= g(\varphi(p)) dy^1(D\varphi(p)\partial_{x^1}) \wedge \cdots \wedge dy^n(D\varphi(p)\partial_{x^n}) \\ &= g(\varphi(p)) \det((dy^j(D\varphi(p)\partial_{x^k}))_{j,k}).\end{aligned}$$

The matrix $(dy^j(D\varphi(p)\partial_{x^k}))_{j,k})$ is the representation matrix of $D\varphi(p)$ in the standard basis. Thus $f(p) = g(\varphi(p)) \det D\varphi(p)$ and hence

$$(\varphi^*\omega)_p = \det D\varphi(p) g(\varphi(p)) dx^1 \wedge \cdots \wedge dx^n.$$

Therefore

$$\int_V \omega = \int_V g d(y^1, \dots, y^n) = \int_U \det D\varphi(p) g(\varphi(p))\, d(x^1, \dots, x^n) = \int_U \varphi^*\omega$$

by the transformation formula. □

In general, a manifold cannot be covered by a single chart. For most operations we have defined so far, this was no problem: These operations were local so that we could restrict our attention to the domain of chart. This is not true for the integral, which takes all values of a function (or later a differential form) into account. To deal with this issue, we need some more technical tools.

Definition 1.4.46 (Locally finite covering)**.** Let X be a topological space. A family $(U_i)_{i \in I}$ of open subsets of X is called *locally finite* if every point $p \in X$ has a neighborhood V_p such that the set $\{i \in I \mid U_i \cap V_p \neq \emptyset\}$ is finite.

A family $(V_j)_{j \in J}$ of open subsets of X is called a *refinement* of $(U_i)_{i \in I}$ if for every $j \in J$ there exists $i \in I$ such that $V_j \subset U_i$.

Proposition 1.4.47. *Let M be a topological manifold. Every family $(U_i)_{i \in I}$ of open subsets of M such that $M = \bigcup_{i \in I} U_i$ admits a locally finite refinement $(V_j)_{j \in J}$ such that $M = \bigcup_{j \in J} V_j$.*

Remark 1.4.48. Topological spaces with the property that every open cover has a locally finite refinement are called *paracompact*. Hence the previous proposition says that every topological manifold is paracompact.

Definition 1.4.49 (Partition of unity)**.** Let M be a smooth manifold and $(U_i)_{i\in I}$ be a locally finite open cover of M. A family $(\psi_i)_{i\in I}$ in $C^\infty(M)$ is called a *partition of unity* subordinate to the cover $(U_i)_{i\in I}$ if

- $\operatorname{supp}\psi_i \subset U_i$ for all $i \in I$,
- $\psi_i(p) \geq 0$ for all $i \in I$, $p \in M$,
- $\sum_{i\in I} \psi_i(p) = 1$ for all $p \in M$.

Remark 1.4.50. Note that for every $p \in M$ there are only finitely many indices $i \in I$ such that $\psi_i(p) \neq 0$. Hence there is no problem defining the sum in the third bullet point.

Theorem 1.4.51. *Let M be a smooth manifold. Every locally finite open cover of M admits a subordinate partition of unity.*

Theorem 1.4.52 (Integral of differential n-forms)**.** *Let M be a smooth oriented n-dimensional manifold. There exists a unique linear map*

$$\int_M : \Omega_c^n(M) \to \mathbb{R}$$

such that

$$\int_M \omega = \int_{\varphi(U)} (\varphi^{-1})^*\omega$$

for every positively oriented chart (φ, U) and every $\omega \in \Omega_c^n(U)$.

Proof. Let $((\varphi_i, U_i))_{i\in I}$ be a smooth atlas of positively oriented charts. To show uniqueness, let $\int_M$, $\int_M'$ be linear maps that satisfy the conditions from the theorem. If $\omega \in \Omega_c^n(M)$, then there exists a finite subset $J \subset I$ such that $\operatorname{supp}\omega \subset \bigcup_{j\in J} U_j$. Let $(\psi_j)_{j\in J}$ be a partition of unity subordinate to $(U_j)_{j\in J}$. We have

$$\int_M \omega = \int_M \sum_{j\in J} \psi_j\omega = \sum_{j\in J} \int_{\varphi_j(U_j)} (\varphi_j^{-1})^*(\psi_j\omega) = \sum_{j\in J} \int_M' \psi_j\omega = \int_M' \omega.$$

This settles uniqueness.

The existence part of the statement takes a bit more work. Essentially, one has to show that the expression

$$\int_M \omega = \sum_{j\in J} \int_{\varphi_j(U_j)} (\varphi_j^{-1})^*(\psi_j \omega)$$

is independent of the atlas and the partition of unity. We will not go into details here. □

Remark 1.4.53. The proof contains a recipe for computing the integral: Let $\omega \in \Omega_c^n(M)$. Choose a positively oriented atlas $((\varphi_i, U_i))_{i\in I}$. Since $\operatorname{supp}\omega$ is compact, there exists a finite subset J of I such that $\operatorname{supp}\omega \subset \bigcup_{j\in J} U_j$. Let $(\psi_j)_{i\in J}$ be a partition of unity subordinate to $(U_j)_{j\in J}$. Then

$$\int_M \omega = \sum_{j\in J} \int_M \psi_j \omega = \sum_{j\in J} \int_{\varphi_j(U_j)} (\varphi_j^{-1})^* \omega,$$

where the integrals on the right side are determined by our previous definition.

1.4.4 Manifolds with boundary and Stokes's theorem

For the formulation of Stokes' theorem, we need objects that are not quite manifolds in our sense, like the closed unit disk or more generally the closed unit ball in $\mathbb{R}^n$. Unlike topological manifolds, that are locally homeomorphic to open subsets of $\mathbb{R}^n$, topological manifolds are locally homeomorphic to a closed half space in $\mathbb{R}^n$.

Definition 1.4.54 (Manifold with boundary)**.** Let $\mathbb{H}^n = \{(x^1, \ldots, x^n) \in \mathbb{R}^n \mid x_n \geq 0\}$. A topological space M is called an *n-dimensional topological manifold with boundary* if it satisfies the following three properties:

(a) M is a Hausdorff space.

(b) For every $x \in M$ there exists an open neighborhood U_x of x that is homeomorphic to an open subset of $\mathbb{H}^n$.

(c) The topology on M is second-countable.

Remark 1.4.55. Note that the only difference to the definition of a topological boundary is in bullet point (b), where $\mathbb{R}^n$ is replaced by $\mathbb{H}^n$.

Lemma 1.4.56. *Every topological manifold is a topological manifold with boundary.*

Proof. It suffices to show that every open subset U of $\mathbb{R}^n$ is homeomorphic to an open subset of $\mathbb{H}^n$. For that purpose consider the map

$$\varphi\colon \mathbb{R}^n \to \mathbb{H}^n,\ (x^1, \dots, x^n) \mapsto (x^1, \dots, x^{n-1}, \exp(x^n)).$$

This map is continuous and has image $\mathbb{H}^n_+ = \{(x^1, \dots, x^n) \in \mathbb{R}^n \mid x^n > 0\}$. The inverse of Φ on $\mathbb{H}^n_+$ is given by

$$\varphi^{-1}\colon \mathbb{H}^n_+ \to \mathbb{R}^n,\ (x^1, \dots, x^n) \mapsto (x^1, \dots, x^{n-1}, \log x^n),$$

which is also continuous. In particular, φ restricts to a homeomorphism from U to $\varphi(U)$. □

Definition 1.4.57 (Interior, boundary)**.** Let M be an n-dimensional topological manifold with boundary. The *interior* $\operatorname{int} M$ of M is the set of all points $x \in M$ that have a neighborhood that is homeomorphic to an open subset of $\mathbb{R}^n$. The complement $M \setminus \operatorname{int} M$ is called the *boundary* of M and denoted by ∂M.

Remark 1.4.58. There are also topological notions of interior and boundary of a set. These are *relative* notion, i.e., if $A \subset X$ a, $B \subset Y$ and $\varphi\colon A \to B$ is a homeomorphism, then φ does not necessarily map the interior (or boundary) of A onto the interior (or boundary) of B. Consequently, the topological notions of interior and boundary are not the same as the interior and boundary of a manifold with boundary.

Example 1.4.59. The interior and boundary of $\mathbb{H}^n$ are $\mathbb{H}^n_+$ and $\mathbb{R}^n \times \{0\}$. If we view $\mathbb{H}^n$ as subspace of $\mathbb{R}^n$, then the manifold interior and boundary of $\mathbb{H}^n$ coincide with the topological interior and boundary.

Example 1.4.60. The closed unit ball $\bar{B}_1(0) \subset \mathbb{R}^n$ is an n-dimensional manifold with boundary. The interior of $\bar{B}_1(0)$ is the open unit ball $B_1(0)$ and the boundary is the unit sphere S^{n-1}.

Lemma 1.4.61. *The boundary of an n-dimensional topological manifold with boundary is an $(n-1)$-dimensional topological manifold (or empty) and the interior is an n-dimensional topological manifold.*

Proof. The Hausdorff property and second countable of ∂M are inherited from M. To show that ∂M is locally Euclidean, let $x \in \partial M$. By definition, there exists an open neighborhood U of x in M, an open subset V of $\mathbb{H}^n$ and a homeomorphism $\varphi\colon U \to V$. By definition of the boundary and interior, we have $\varphi(U \cap \partial M) = V \cap (\mathbb{R}^{n-1} \times \{0\})$. Since $\mathbb{R}^{n-1} \times \{0\}$ is homeomorphic to $\mathbb{R}^{n-1}$, the open neighborhood $U \cap \partial M$ of x in ∂M is homeomorphic to an open subset of $\mathbb{R}^{n-1}$. The proof that $\operatorname{int} M$ is an n-dimensional topological manifold is similar. □

To define a smooth manifold with boundary, we proceed exactly as before for manifolds with boundary. To do so, we need a notion of smoothness for maps defined on open subsets of $\mathbb{H}^n$, which are not necessarily open in $\mathbb{R}^n$.

Definition 1.4.62 (Smooth map)**.** Let $A \subset \mathbb{R}^m$. A map $f\colon A \to \mathbb{R}^n$ is called *smooth* if there exists an open neighborhood U of A in $\mathbb{R}^m$ and a smooth map $g\colon U \to \mathbb{R}^n$ such that $g|_A = f$.

If $A \subset \mathbb{R}^m$ is open, we can just take $U = A$ and $g = f$ and recover the previous definition of smooth maps.

Definition 1.4.63 (Smooth manifold with boundary)**.** Let M be an n-dimensional topological manifold with boundary. A chart is a pair (φ, U) consisting of an open subset U of M and a homeomorphism φ from U onto an open subset of $\mathbb{H}^n$. A *smooth atlas* is a family $((\varphi_i, U_i))_{i\in I}$ of charts such that $M = \bigcup_{i\in I} U_i$ and the transition maps

$$\varphi_{ij}\colon \varphi_i(U_i \cap U_j) \to \varphi_j(U_i \cap U_j),\ \varphi_{ij} = \varphi_j^{-1} \circ \varphi_i$$

are smooth for all $i, j \in I$ with $U_i \cap U_j \neq \emptyset$.

Two smooth atlases $\mathcal{A}$ and $\mathcal{B}$ are called *equivalent* if $\mathcal{A} \cup \mathcal{B}$ is again a smooth atlas. An equivalence class of smooth atlases is called a *smooth structure* on M and a topological manifold with boundary equipped with a smooth structure is called a *smooth manifold with boundary*.

If M, N are smooth manifolds with boundary and $((U_i, \varphi_i))_{i\in I}$, $((V_j, \psi_j))_{j\in J}$ are smooth atlases, then a map $\varphi\colon M \to N$ is called *smooth* if $\psi_j \circ \varphi \circ \varphi_i^{-1}$ is smooth for all $i \in I$, $j \in J$ with $U_i \cap \varphi^{-1}(V_j) \neq \emptyset$. As before, we write $C^\infty(M, N)$ for the set of all smooth maps from M to N and $C^\infty(M)$ for $C^\infty(M, \mathbb{R})$.

Lemma 1.4.64. *If M is a topological manifold with boundary and $((\varphi_i, U_i))_{i\in I}$ is a smooth atlas, then $((\varphi_i|_{U_i\cap\partial M}, U_i\cap\partial M))_{i\in I}$ is a smooth atlas for ∂M such that the inclusion map $i\colon \partial M \to M$ is smooth. Similarly, $((\varphi_i|_{U_i\cap \operatorname{int} M}, U_i \cap \operatorname{int} M))_{i\in I}$ is a smooth atlas for $\operatorname{int} M$.*

The definition of the tangent space of a smooth manifold with boundary is exactly the same as before.

Definition 1.4.65 (Tangent vector, tangent space)**.** Let M be a smooth manifold with boundary and $p \in M$. A *tangent vector* at p is a linear map $X_p\colon C^\infty(M) \to \mathbb{R}$ that satisfies the Leibniz rule

$$X_p(fg) = f(p)X_p(g) + X_p(f)g(p)$$

for all $f, g \in C^\infty(M)$. The set of all tangent vectors at p is denoted by T_pM and called the *tangent space* of M at p.

Remark 1.4.66. One can also characterize tangent vectors in terms of curves. However, for points on the boundary, it does not suffice to consider curves defined on open intervals around 0, one has to allow for curves defined on half-open intervals containing 0.

Definition 1.4.67 (Inward and outward pointing vectors)**.** Let M be a smooth n-dimensional manifold with boundary and $p \in \partial M$. A tangent vector $X_p \in T_pM$ is called *tangential to* ∂M if there exists a smooth curve $\gamma\colon (-\varepsilon, \varepsilon) \to \partial M$ for some $\varepsilon > 0$ such that $\gamma(0) = p$ and

$$X_p(f) = \left.\frac{d}{dt}\right|_{t=0} f(\gamma(t))$$

for all $f \in C^\infty(M)$.

A tangent vector $X_p \in T_pM$ is called *outward pointing* (resp. *inward pointing*) if it is not tangent to ∂M and there exists a smooth curve $\gamma\colon (-\varepsilon, 0] \to M$ (resp. $\gamma\colon [0, \varepsilon) \to M$) for some $\varepsilon > 0$ such that $\gamma(0) = p$ and

$$X_p(f) = \left.\frac{d}{dt}\right|_{t=0} f(\gamma(t))$$

for all $f \in C^\infty(M)$.

Lemma 1.4.68. *Let M be a smooth n-dimensional manifold with boundary and $p \in \partial M$. A tangent vector $X_p \in T_pM$ belongs to the image of $Di(p)$ if and only if it is tangential to ∂M. Moreover, $T_pM \setminus \operatorname{im} Di(p)$ is the disjoint union of the inward and the outward pointing tangent vectors at p.*

Differential forms and the *exterior derivative* on smooth manifolds with boundary can be defined just as before for manifolds without boundary. To define integration of differential forms and state Stokes's theorem, we also need a notion of orientation of smooth manifolds with boundary.

Definition 1.4.69 (Orientation, orientable manifold)**.** Let M be a smooth manifold with boundary. A smooth atlas $((\varphi_i, U_i))_{i\in I}$ is called an *orientation* if $((\varphi_i|_{U_i \cap \operatorname{int} M}, U_i \cap \operatorname{int} M))_{i \in I}$ is an orientation for $\operatorname{int} M$. If M admits an orientation, then it is called *orientable*.

Lemma 1.4.70. *Let M be a smooth oriented dimensional manifold with boundary of even (resp. odd) dimension. There exists a unique orientation on the boundary ∂M such that for every positively oriented chart (φ, U) of M with $U \cap \partial M \neq \emptyset$ the chart $(\varphi|_{U\cap\partial M}, U \cap \partial M)$ is positively oriented (resp. negatively oriented).*

Remark 1.4.71. The geometric interpretation of this orientation on ∂M is the following: At any point $p \in \partial M$, if one takes a positively oriented basis of $T_p\partial M$ and adds an outward pointing vector as last basis element, one gets a positively oriented basis of T_pM. Additionally, this orientation convention makes Stokes's theorem true.

The orientation from the previous lemma is called the *induced orientation* on ∂M. The *integral* of a compactly supported differential n-form on a smooth oriented n-dimensional manifold with boundary can be defined analogously to the case of manifolds without boundary using positively oriented atlases and partitions of unity.

Theorem 1.4.72 (Stokes). *Let M be a smooth oriented n-dimensional manifold with boundary and endow ∂M with the induced orientation. If $\omega \in \Omega_c^{n-1}(M)$, then $i^*\omega \in \Omega_c^{n-1}(\partial M)$ and*

$$\int_{\partial M} i^*\omega = \int_M d\omega.$$

Proof. We first prove the result for $M = \mathbb{H}^n$ with standard orientation.

An $(n-1)$-form on $\mathbb{H}^n$ is of the form

$$\omega = \sum_{j=1}^n f_j dx^1 \wedge \cdots \wedge \widehat{dx^j} \wedge \cdots \wedge dx^n$$

with $f_1, \ldots, f_n \in C^\infty(\mathbb{H}^n)$. It has compact support if and only if $f_1, \ldots, f_n$ have compact support.

We have

$$\begin{aligned} d\omega &= \sum_{j=1}^n (-1)^{j-1} \frac{\partial f_j}{\partial x^j} dx^1 \wedge \cdots \wedge dx^n \\ i^*\omega &= f_n|_{\partial\mathbb{H}^n} dx^1 \wedge \cdots \wedge dx^{n-1}. \end{aligned}$$

By the definition of the integral of differential forms,

$$\begin{aligned} \int_{\mathbb{H}^n} d\omega &= \sum_{j=1}^n \int_0^\infty \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty (-1)^{j-1} \frac{\partial f_j}{\partial x^j}\, dx^1 \ldots\, dx^{n-1}\, dx^n, \\ \int_{\partial\mathbb{H}^n} i^*\omega &= (-1)^n \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty f_n(x^1, \ldots, x^{n-1}, 0)\, dx^1 \ldots\, dx^{n-1} \end{aligned}$$

If $j \in \{1, \ldots, n-1\}$ and $\operatorname{supp} f_j \subset [-R, R]^{n-1} \times [0, R]$, then we have for $j \leq n-1$ that

$$\int_{-\infty}^\infty \frac{\partial f_j}{\partial x^j}\, dx^j = f_n|_{x^j=R} - f_n|_{x^j=-R} = 0$$

and
$$\int_0^\infty \frac{\partial f_n}{\partial x^n}\, dx^n = f_n|_{x^n=R} - f_n|_{x_n=0} = -f_n|_{x^n=0}$$
by the fundamental theorem of calculus.

Therefore
$$\int_{\mathbb{H}^n} d\omega = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (-1)^n f_n(x^1, \ldots, x^{n-1}, 0)\, dx^1 \ldots dx^{n-1} = \int_{\partial\mathbb{H}^n} i^*\omega.$$
Now let M be an arbitrary smooth oriented n-dimensional manifold with boundary and $\omega \in \Omega_c^{n-1}(M)$. Let $((\varphi_i, U_i))_{i\in I}$ be a positively oriented atlas of M. Since $\operatorname{supp}\omega$ is compact, there exists $J \subset I$ finite such that $\operatorname{supp}\omega \subset \bigcup_{j\in J} U_j$. Let $(\psi_j)_{j\in J}$ be a partition of unity subordinate to $(U_j)_{j\in J}$. We have
$$\begin{aligned}
\int_M d\omega &= \sum_{j\in J} \int_{U_j} d(\psi_j\omega) \\
&= \sum_{j\in J} \int_{\varphi_j(U_j)} (\varphi_j^{-1})^* d(\psi_j\omega) \\
&= \sum_{j\in J} \int_{\mathbb{H}^n} (\varphi_j^{-1})^* d(\psi_j\omega) \\
&= \sum_{j\in J} \int_{\mathbb{H}^n} d((\varphi_j^{-1})^* (\psi_j\omega)) \\
&= \sum_{j\in J} \int_{\partial\mathbb{H}^n} (\varphi_j^{-1} \circ i)^* (\psi_j\omega) \\
&= \sum_{j\in J} \int_{U_j\cap\partial M} i^*(\psi_j\omega) \\
&= \int_{\partial M} i^*\omega.
\end{aligned}$$
□

1.5 Hamiltonian formalism and symplectic geometry

1.5.1 Symplectic manifolds

The fundamental model of a symplectic manifold is the cotangent bundle of a smooth manifold, which occurs as phase space for many mechanical models. Let Q be a smooth manifold with cotangent bundle T^*Q, which is itself again a smooth manifold. Recall that there is a canonical projection map $\pi\colon T^*Q \to Q$. The pull-back π^* maps $\Omega^1(Q)$ to $\Omega^1(T^*Q)$. Moreover, any $\alpha \in \Omega^1(Q)$ is a smooth map from Q to T^*Q, hence the pull-back α^* maps $\Omega^1(T^*Q)$ to $\Omega^1(Q)$.

Proposition 1.5.1. *There exists a unique* 1*-form* $\lambda \in \Omega^1(T^*Q)$ *such that for every* $\alpha \in \Omega^1(Q)$ *we have* $\alpha^*\lambda = \alpha$. *It satisfies*

$$\lambda_{(q,p)}(\xi) = p(D\pi(q,p)[\xi])$$

for all $q \in Q$, $p \in T_qQ$, $\xi \in T_{(q,p)}T^*Q$.

Proof. For every $q \in Q$, $p \in T_qQ$, the map

$$\lambda_{(q,p)}\colon T_{(q,p)}T^*Q \to \mathbb{R},\ \xi \mapsto p(D\pi(q,p)[\xi])$$

is linear. In other words, $\lambda_{(q,p)} \in T^*_{(q,p)}T^*Q$. Since all the maps involved in the definition of λ are smooth, it is not hard to see that the map

$$T^*Q \to T^*(T^*Q),\ (q,p) \mapsto ((q,p), \lambda_{(q,p)})$$

is smooth. Thus $\lambda \in \Omega^1(T^*Q)$.

If $\alpha \in \Omega^1(Q)$ and $v \in T_qQ$, then

$$\begin{aligned}
(\alpha^*\lambda)_q[v] &= \lambda_{\alpha(q)}[D\alpha(q)[v]] \\
&= \alpha_q[D\pi(\alpha(q)[D\alpha(q)[v]]] \\
&= \alpha_q[D(\pi \circ \alpha)(q)[v]] \\
&= \alpha_q[v],
\end{aligned}$$

where we used that $\pi \circ \alpha = \mathrm{id}$. Thus $\alpha^*\lambda = \alpha$. This proves the existence part of the statement.

For uniqueness, let $\mu \in \Omega^1(T^*Q)$ such that $\alpha^*\mu = \alpha$ for all $\alpha \in \Omega^1(Q)$. Hence $0 = \alpha^*(\lambda - \mu)$ for all $\alpha \in \Omega^1(Q)$. For every $\xi \in T_{(q,p}T^*Q \ker D\pi(q,p)$ we find a neighborhood U of q and $\alpha \in \Omega^1(T^*U)$ such that $\alpha(q) = p$ and $p(D\pi(q,p)[\xi]) = 0$. Thus $\lambda = \mu$ on $T_{(q,p)}T^*Q \setminus D\pi(q,p)$. By continuity, we conclude $\lambda = \mu$. □

Definition 1.5.2 (Liouville 1-form)**.** The 1-form λ from the previous proposition is called Liouville 1-form.

Remark 1.5.3. If (U, φ) is a coordinate chart of Q and $\varphi = (x^1, \ldots, x^n)$, then one can define coordinates $(q^1, \ldots, q^n, p_1, \ldots, p_n)$ of T^*U by

$$q^j : T^*U \to \mathbb{R},\ (q, p) = x^j(q)$$
$$p_j : T^*U \to \mathbb{R},\ (q, p) = p\left(\left.\frac{\partial}{\partial x^j}\right|_q\right).$$

In these coordinates, the Liouville 1-form λ can be expressed as

$$\lambda = \sum_{j=1}^{n} p_j dq^j.$$

Definition 1.5.4 (Canonical Symplectic form, symplectic manifold)**.** Let Q be a smooth manifold and $\lambda \in T^*Q$ the Liouville 1-form. The *canonical symplectic form* on T^*Q is $\omega_{\mathrm{Liouv}} = d\lambda \in \Omega^2(T^*Q)$.

Remark 1.5.5. Since $d^2 = 0$, the canonical symplectic form satisfies $d\omega_{\mathrm{Liouv}} = 0$.

Remark 1.5.6. In local coordinates as above, the canonical symplectic form can be expressed as

$$\omega_{\mathrm{Liouv}} = \sum_{j=1}^{n} dp_j \wedge dq^j.$$

Definition 1.5.7 (Symplectic form, symplectic manifold)**.** Let M be a smooth manifold. A *symplectic form* on M is a 2-form $\omega \in \Omega^2(M)$ with the following two properties:

(a) Non-degeneracy: The map

$$T_xM \to T_x^*M,\ v \mapsto \omega_x(v, \cdot\,)$$

is a linear isomorphism for every $x \in M$.

(b) Integrability: $d\omega = 0$.

A pair (M, ω) consisting of a smooth manifold M and a symplectic form ω on M is called a *symplectic manifold*.

Remark 1.5.8. A symplectic form on M can only exist if M has even dimension. The non-degeneracy condition (a) is equivalent to requiring that $\omega^{\wedge n}$ is a nowhere vanishing $2n$-form if the dimension of M is $2n$.

Example 1.5.9. The canonical symplectic form $\omega_{\mathrm{Liouv}} \in \Omega^2(T^*Q)$ is indeed a symplectic form on T^*Q. Thus $(T^*Q, \omega_{\mathrm{Liouv}})$ is a symplectic manifold.

Example 1.5.10. If M is a smooth orientable 2-dimensional manifold and $\omega \in \Omega^2(M)$ is a volume form, then ω is a symplectic form.

Example 1.5.11. The sphere S^{2n} does not admit any symplectic form for $n \geq 2$. Indeed, any closed 2-form ω on S^{2n} is exact, i.e. $\omega = d\alpha$ for some $\alpha \in \Omega^1(S^{2n})$. If $\omega^{\wedge n}$ were a volume form, then

$$0 \neq \int_M \omega^{\wedge n} = \int_M d(\alpha \wedge \omega^{\wedge(n-1)}) = 0$$

by Stokes's theorem, a contradiction.

Theorem 1.5.12 (Darboux)**.** *If (M, ω) is a $2n$-dimensional symplectic manifold and $x \in M$, then there exists a chart (φ, U) with $x \in U$ and $\varphi = (q^1, \dots, q^n, p_1, \dots, p_n)$ such that $\omega|_U = \sum_{j=1}^n dp_j \wedge dq^j$.*

Definition 1.5.13 (Canonical coordinates)**.** Let (M, ω) be a symplectic manifold. If (φ, U) is a chart with $\varphi = (q^1, \dots, q^n, p_1, \dots, p_n$ such that $\omega|_U = \sum_{j=1}^n dp_j \wedge dq^j$, then $(q^1, \dots, q^n, p_1, \dots, p_n)$ are called *canonical coordinates.*

1.5.2 Hamiltonian systems

Lemma 1.5.14. *Let (M, ω) be a symplectic manifold. If $H \in C^\infty(M)$, then there exists a unique vector field $X_H \in \mathcal{X}(M)$ such that*

$$i_{X_H}\omega + dH = 0.$$

Proof. By the non-degeneracy condition, for every $x \in M$ there exists a unique $v_x \in T_xM$ such that

$$\omega(v_x, \cdot\,) = -(dH)_x.$$

A vector field $X \in \mathcal{X}(M)$ satisfies $i_X\omega + dH = 0$ if and only if $X_x = v_x$ for every $x \in M$. $\square$

Definition 1.5.15 (Hamiltonian vector field)**.** Let (M, ω) be a symplectic manifold and $H \in C^\infty(M)$. The triple (M, ω, H) is called a *Hamiltonian system.* The vector field X_H from the previous lemma is called the *Hamiltonian vector field* associated with the Hamiltonian H.

Example 1.5.16. Let $M = T^*\mathbb{R}$ with coordinates (q, p), $\omega = dp \wedge dq$ and $H(q, p) = \frac{1}{2}(q^2 + p^2)$. This is a mathematical model of the harmonic oscillator.

The associated Hamiltonian vector field is $X_H = p\frac{\partial}{\partial q} - q\frac{\partial}{\partial p}$. Indeed,

$$\begin{aligned}
i_{X_H}\omega(Y) &= dp \wedge dq \left(p\frac{\partial}{\partial q} - q\frac{\partial}{\partial p}\right) \\
&= \begin{vmatrix} dp\left(p\frac{\partial}{\partial q} - q\frac{\partial}{\partial p}\right) & dp(Y) \\ dq\left(p\frac{\partial}{\partial q} - q\frac{\partial}{\partial p}\right) & dq(Y) \end{vmatrix} \\
&= \begin{vmatrix} -q & dp(Y) \\ p & dq(Y) \end{vmatrix} \\
&= -qdq(Y) - pdp(Y) \\
&= -dH(Y).
\end{aligned}$$

Proposition 1.5.17. *Let (M, ω, H) be a Hamiltonian system. In canonical coordinates $(q^1, \dots, q^n, p_1, \dots, p_n)$ on U, the Hamiltonian vector field satisfies*

$$X_H|_U = \sum_{j=1}^n \frac{\partial H}{\partial p_j}\frac{\partial}{\partial q^j} - \frac{\partial H}{\partial q^j}\frac{\partial}{\partial p_j}.$$

Proof. If $Y \in \mathcal{X}(U)$, then

$$\begin{aligned}
\omega\left(\sum_{j=1}^n \frac{\partial H}{\partial p_j}\frac{\partial}{\partial q^j} - \frac{\partial H}{\partial q^j}\frac{\partial}{\partial p_j}, Y\right) &= \sum_{j,k=1}^n (dq^k \wedge dp_k)\left(\frac{\partial H}{\partial p_j}\frac{\partial}{\partial q^j} - \frac{\partial H}{\partial q^j}\frac{\partial}{\partial p_j}, Y\right) \\
&= \sum_{j=1}^n -\frac{\partial H}{\partial q^j}dq^j(Y) - \frac{\partial H}{\partial p_j}dp_j(Y) \\
&= -dH(Y).
\end{aligned}$$

From the uniqueness statement in the previous lemma we deduce

$$X_H|_U = \sum_{j=1}^n \frac{\partial H}{\partial p_j}\frac{\partial}{\partial q^j} - \frac{\partial H}{\partial q^j}\frac{\partial}{\partial p_j}. \qquad \square$$

Corollary 1.5.18 (Hamilton equations)**.** *Let (M, ω, H) be a Hamiltonian system. In canonical coordinates, $(q(t), p(t))$ is an integral curve of X_H if and only if*

$$\begin{aligned}
\dot{q}^j(t) &= \frac{\partial H}{\partial p_j} \\
\dot{p}_j(t) &= -\frac{\partial H}{\partial q^j}
\end{aligned}$$

for all $j \in \{1, \dots, n\}$.

Proposition 1.5.19 (Energy conservation)**.** *If (M, ω, H) is a Hamiltonian system, $H \circ \gamma$ is constant for any integral curve γ of X_H.*

Proof. By definition of X_H and integral curves, we have

$$\frac{d}{dt}(H \circ \gamma) = dH(\dot{\gamma}) = dH(X_H) = -i_{X_H}\omega(X_H) = -\omega(X_H, X_H) = 0. \qquad \square$$

Proposition 1.5.20 (Liouville)**.** *Let (M, ω, H) be a Hamiltonian system. The flow Φ of X_H satisfies $\Phi_t^*\omega = \omega$.*

Proof. We have

$$\begin{aligned}
\frac{d}{dt}\Phi_t^*\omega &= \Phi_t^* L_{X_H}\omega \\
&= \Phi_t^*(di_{X_H}\omega + i_{X_H}d\omega) \\
&= \Phi_t^*(d(-dH) + 0) \\
&= 0.
\end{aligned}$$

The first identity uses the fact that Φ is the flow of X_H, the second is Cartan's magic formula and the third used the integrability of ω and the definition of the Hamiltonian vector field. Since $\Phi_0^*\omega = \omega$, we conclude $\Phi_t^*\omega = \omega$ for all t. $\square$

Remark 1.5.21. With the definition of the Lie derivative of differential forms, the previous result can be rewritten as $L_{X_H}\omega = 0$. Moreover, an easy consequence is that $\Phi_t^*\omega^{\wedge n} = \omega^{\wedge n}$. Since ω is a symplectic form, $\omega^{\wedge n}$ is a volume form on M. In this sense, the flow of a Hamiltonian vector field is volume-preserving.

Definition 1.5.22 (Poisson bracket)**.** Let (M, ω) be a symplectic manifold. For $f, g \in C^\infty(M)$, the *Poisson bracket* $\{f, g\}_\omega$ is defined as

$$\{f, g\}_\omega = -\omega(X_f, X_g).$$

Lemma 1.5.23. *Let (M, ω) be a symplectic manifold and $f, g \in C^\infty(M)$. In symplectic coordinates we have*

$$\{f, g\}_\omega = \sum_{j=1}^{n} \frac{\partial f}{\partial q^j}\frac{\partial g}{\partial p^j} - \frac{\partial f}{\partial p_j}\frac{\partial g}{\partial q^j}.$$

Proposition 1.5.24. *Let (M, ω, H) be a Hamiltonian system and Φ the flow of X_H. If $f \in C^\infty(M)$, then*

$$\frac{d}{dt}(f \circ \Phi_t) = \{f, H\}_\omega \circ \Phi_t.$$

Definition 1.5.25 (Constant of motion)**.** Let (M, ω, H) be a Hamiltonian system. A function $f \in C^\infty(M)$ is called a *constant of motion* or *first integral* if $\{f, H\}_\omega = 0$.

Chapter 2

Lie Groups

2.1 Basic definitions and Lie algebras

Definition 2.1.1 (Lie group)**.** A Lie group is a group G with the structure of a smooth manifold such that the multiplication $G \times G \to G$, $(g, h) \mapsto gh$ and the inversion $G \to G$, $g \mapsto g^{-1}$ are smooth.

Example 2.1.2. The general linear group $\mathrm{GL}_n(\mathbb{R}) = \{A \in \mathbb{R}^{n\times n} \mid \det(A) \neq 0\}$ is an open subset of $\mathbb{R}^{n\times n} \cong \mathbb{R}^{n^2}$ since det is a continuous map. As such, $\mathrm{GL}_n(\mathbb{R})$ carries a natural smooth structure. Moreover, matrix multiplication and inversion are polynomials in the entries of the matrices and thus smooth maps. Hence $\mathrm{GL}_n(\mathbb{R})$ is a Lie group.

Example 2.1.3. The special linear group $\mathrm{SL}_n(\mathbb{R}) = \{A \in \mathbb{R}^{n\times n} \mid \det(A) = 1\}$ is a smooth manifold of dimension $n^2 - 1$ by the implicit function theorem.

Example 2.1.4. The orthogonal group $\mathrm{O}_n(\mathbb{R}) = \{A \in \mathbb{R}^{n\times n} \mid A^{\mathrm{T}}A = 1\}$ is a smooth manifold of dimension $n(n-1)/2$.

Example 2.1.5. The unitary group $\mathrm{U}_n(\mathbb{C}) = \{A \in \mathbb{C}^{n\times n} \mid A^{\mathrm{H}}A = 1\}$ is a smooth manifold of dimension n^2. Note that this dimension refers to the dimension over $\mathbb{R}$, not $\mathbb{C}$.

If G is a Lie group, the inversion map $i\colon G \to G$, $g \mapsto g^{-1}$ and the left multiplication map $L_g G \to G$, $h \mapsto gh$ for $g \in G$ are diffeomorphisms.

Definition 2.1.6 (Left-invariant vector field)**.** Let G be a Lie group. A vector field $X \in \mathcal{X}(G)$ is called *left-invariant* if $(L_g)_*X = X$ for all $g \in G$. The space of all left-invariant vector fields on G is denoted by $\mathfrak{g}$.

Lemma 2.1.7. *Let G be a Lie group. If $X, Y \in \mathfrak{g}$, then $[X, Y] \in \mathfrak{g}$.*

Lemma 2.1.8. *Let G be a Lie group. The evaluation map $\mathfrak{g} \to T_eG$, $X \mapsto X_e$ is a linear isomorphism.*

Proof. For $v \in T_e G$ and $g \in G$ let $\chi(v)_g = DL_g(e)[v]$. We have

$$\begin{aligned}((L_h)_*\chi(v))_g &= DL_h(L_h^{-1}(g))[\chi(v)_{h^{-1}g}] \\ &= DL_h(L_h^{-1}(g))DL_{h^{-1}g}(e)[v] \\ &= D(L_h L_{h^{-1}g})(e)[v] \\ &= \chi(v)_g.\end{aligned}$$

Thus χ is an inverse of $X \mapsto X_e$. $\square$

Definition 2.1.9 (Lie algebra of a Lie group)**.** Let G be a Lie group. The space of left-invariant vector fields $\mathfrak{g}$ is called the *Lie algebra of the Lie group* G.

Remark 2.1.10. By the previous result, the Lie algebra of a Lie group can be canonically identified with the tangent space at the unit element.

Another important property of left-invariant vector fields on a Lie group is that they are globally integrable.

Proposition 2.1.11. *Let G be a Lie group. For every $X \in \mathfrak{g}$ and $g \in G$ there exists a unique smooth curve $\gamma\colon \mathbb{R} \to G$ such that*

$$\begin{cases} \dot\gamma(t) = X_{\gamma(t)},\ t \in \mathbb{R} \\ \gamma(0) = g \end{cases}$$

Proof. Suppose there exists $g \in G$ such that the maximal existence interval $(t^-(g), t^+(g))$ is not $\mathbb{R}$, say $t^+(g) < \infty$. For $t_0 > 0$ let $h = \gamma(t_0)g^{-1}$ and

$$\tilde\gamma\colon (t^-(g) + t_0, t^+(g), +t_0) \to G,\ \tilde\gamma(t) = L_h\gamma(t - t_0).$$

By definition, $\tilde\gamma(t_0) = \gamma(t_0)$ and

$$\begin{aligned}\frac{d}{dt}\tilde\gamma(t) &= DL_h(\gamma(t-t_0))\dot\gamma(t-t_0) \\ &= DL_h(\gamma(t-t_0))X_{\gamma(t-t_0)} \\ &= DL_h(h^{-1}\tilde\gamma(t))X_{h^{-1}\tilde\gamma(t)} \\ &= ((L_h)_*X)_{\tilde\gamma(t)} \\ &= X_{\tilde\gamma(t)}.\end{aligned}$$

By the uniqueness of integral curves, we conclude $\tilde\gamma = \gamma$ on $(t^-(g)+t_0, t^+(g))$. Thus one can extend γ to an integral curve on $(t^-(g), t^+(g) + t_0)$ in contradiction to the maximality of $(t^-(g), t^+(g))$. Therefore $(t^-(g), t^+(g)) = \mathbb{R}$. $\square$

As a consequence, every $X \in \mathfrak{g}$ admits a global flow $\Phi_X : \mathbb{R} \times G \to G$ and for every $t \in \mathbb{R}$, the map Φ_X^t is a group isomorphism.

Definition 2.1.12 (Exponential map)**.** Let G be a Lie group. The exponential map exp on G is defined by

$$\exp \colon \mathfrak{g} \to G,\ X \mapsto \Phi_X^1.$$

Example 2.1.13. Let $G = \mathrm{GL}_n(\mathbb{R})$. Since $\mathrm{GL}_n(\mathbb{R})$ is open in $\mathbb{R}^{n\times n}$, the tangent space at 1_n is canonically isomorphic to $\mathbb{R}^{n\times n}$. Under this identification, the exponential map is given by

$$\exp \colon \mathbb{R}^{n\times n} \to \mathrm{GL}_n(\mathbb{R}),\ A \mapsto \exp(A) = \sum_{k=0}^{\infty} \frac{t^k A^k}{k!}.$$

The series on the right side converges absolutely in each entry.

Example 2.1.14. If $G \subset \mathbb{R}^{n\times n}$ is a Lie subgroup, then $\mathfrak{g}$ can be identified with a Lie subalgebra of $\mathbb{R}^{n\times n}$ and the exponential map is again given by

$$\exp(A) = \sum_{k=0}^{\infty} \frac{t^k A^k}{k!}.$$

Proposition 2.1.15. *Let G be a Lie group. There exists an open subset U of T_eG with $0 \in U$ and an open subset V of G with $e \in V$ such that* $\exp|_U$ *is a diffeomorphism onto V.*

Theorem 2.1.16 (Campbell–Baker–Hausdorff formula)**.** *Let G be a Lie group.*

(a) If $X, Y \in \mathfrak{g}$ with $[X, Y] = 0$, then

$$\exp(X + Y) = \exp(X)\exp(Y).$$

In particular, $\exp(X)\exp(Y) = \exp(Y)\exp(X)$.

(b) There exists an open subset U of $\mathfrak{g}$ with $0 \in U$ such that if $X, Y \in U$, then there exists $Z \in \mathfrak{g}$ such that $\exp(X + Y) = \exp(Z)$. *There is an explicit series expansion for Z starting with*

$$Z = X+Y+\frac{1}{2}[X,Y]+\frac{1}{12}[X,[X,Y]]-\frac{1}{12}[Y,[X,Y]]-\frac{1}{24}[Y,[X,[X,Y]]]+\dots,$$

where all the higher order terms are iterated commutators of X and Y.

Definition 2.1.17 (Lie algebra)**.** A (finite-dimensional real) *Lie algebra* is a finite-dimensional vector space $\mathfrak{g}$ over $\mathbb{R}$ with an alternating form $[\,\cdot\,,\cdot\,]\colon \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ that satisfies the *Jacobi identity*

$$[X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] = 0$$

for all $X, Y, Z \in \mathfrak{g}$.

Remark 2.1.18. If M is a smooth manifold, then $\mathcal{X}(M)$ with the Lie bracket of vector fields satisfies all the properties of a Lie algebra except that it is rarely a finite-dimensional vector space.

Example 2.1.19. If G is a Lie group, then the space $\mathfrak{g}$ of left-invariant vector fields on G with the Lie bracket is a Lie algebra, justifying the name "Lie algebra of G".

Definition 2.1.20 (Lie algebra homomorphism)**.** Let $\mathfrak{g}$, $\mathfrak{h}$ be Lie algebras. A *Lie algebra homomorphism* is a linear map $\psi\colon \mathfrak{g} \to \mathfrak{h}$ that satisfies $\psi([X,Y]) = [\psi(X), \psi(Y)]$ for all $X, Y \in \mathfrak{g}$.

Example 2.1.21. Let G, H be Lie groups with Lie algebras $\mathfrak{g}$, $\mathfrak{h}$ and $\varphi\colon G \to H$ a smooth group homomorphism. The pushforward map $\varphi_*\colon \mathfrak{g} \to \mathfrak{h}$ is a Lie algebra homomorphism.

Theorem 2.1.22 (Lie's second theorem)**.** *Let G, H be Lie groups with Lie algebras $\mathfrak{g}$, $\mathfrak{h}$ and let $\psi\colon \mathfrak{g} \to \mathfrak{h}$ be a Lie algebra homomorphism. If G is simply connected, then there exists a unique smooth group homomorphism $\varphi\colon G \to H$ such that $\psi = \varphi_*$.*

Remark 2.1.23. A topological space X is called *path-connected* if for every $x, y \in X$ there exists a continuous map $\gamma\colon [0,1] \to X$ such that $\gamma(0) = x$ and $\gamma(1) = y$. The space X is called *simply connected* if it is path connected and for every $x \in X$ and $\gamma\colon S^1 \to X$ continuous with $\gamma(1,0) = x$ there exists a continuous $H\colon S^1 \times [0,1] \to X$ such that

- $H(\,\cdot\,,0) = \gamma$, $H(p,1) = x$ for all $p \in S^1$,
- $H((1,0),t) = x$ for all $t \in [0,1]$.

Intuitively this means that a space is path-connected if any two points can be joint by a continuous path and simply connected if additionally any loop can be continuously be deformed to a point.

2.2 The examples $\mathrm{SO}_3(\mathbb{R})$ and $\mathrm{SU}_2(\mathbb{C})$

2.2.1 The special orthogonal group $\mathrm{SO}_3(\mathbb{R})$

Recall that $\mathrm{SO}_n(\mathbb{R}) = \{A \in \mathbb{R}^{n\times n} \mid A^{\mathrm{T}}A = 1_n,\ \det(A) = 1\}$.

Lemma 2.2.1. *A matrix $A \in \mathbb{R}^{n\times n}$ satisfies $A^{\mathrm{T}}A = 1_n$ if and only if*

$$\|Ax\|_2 = \|x\|_2$$

for all $x \in \mathbb{R}^n$.

Moreover, $A \in \mathrm{GL}_n(\mathbb{R})$ satisfies $\det(A) > 0$ if and only if $(Av_1, \dots, Av_n)$ has the same orientation as $(v_1, \dots, v_n)$ for every basis $(v_i)_{i=1}^n$ of $\mathbb{R}^n$.

In other words, $\mathrm{SO}_n(\mathbb{R})$ contains exactly those matrices that present orientation-preserving isometries. The description is particularly easy in dimensions 2 and 3: The orientation-preserving isometries of $\mathbb{R}^2$ are the rotations around the origin and the orientation-preserving isometries of $\mathbb{R}^3$ are the rotations around an axis through the origin.

Hence for every $O \in \mathrm{SO}_3(\mathbb{R})$ there exists a unit vector $v \in \mathbb{R}^3$ and $\alpha \in [0, \pi]$ such that O represents the rotation around the axis $\mathbb{R}v$ (oriented in direction of v) with angle α. Note that a rotation around v by an angle $\alpha \in (\pi, 2\pi)$ can be represented by a rotation by the angle $2\pi - \alpha$ around $-v$.

The pairs (v, α) and (v', α') represent the same rotation if and only if $\alpha' = \alpha$ and $v' = v$ or $\alpha' = \alpha = \pi$ and $v' = -v$. This gives an identification of $\mathrm{SO}_3(\mathbb{R})$ with $\bar{B}_\pi(0)/\sim$ with $w \sim w'$ if $w = w'$ or $|w| = \pi$, $w' = -w$, where the rotation around v with angle α is mapped to $[(v, \alpha)]$. It takes some work to see that this identification in fact gives rise to a homeomorphism from $\mathrm{SO}_3(\mathbb{R})$ and $\mathbb{R}P^3$. With the right choice of smooth structure on the real projective space, even more is true:

Proposition 2.2.2. *The Lie group $\mathrm{SO}_3(\mathbb{R})$ is diffeomorphic to $P\mathbb{R}^3$.*

Now let us have short look at the Lie algebra of $\mathrm{SO}_3(\mathbb{R})$. Since $\mathrm{SO}_3(\mathbb{R})$ is embedded in $\mathbb{R}^{3\times 3}$, we can compute the tangent space $T_1\,\mathrm{SO}_3(\mathbb{R})$ as subspace of $\mathbb{R}^{3\times 3}$. If $\gamma\colon (-\varepsilon, \varepsilon) \to \mathbb{R}^{3\times 3}$ is a smooth curve such that $\gamma(t) \in \mathrm{SO}_3(\mathbb{R})$ for all $t \in (-\varepsilon, \varepsilon)$ and $\gamma(0) = I_3$, then

$$0 = \left.\frac{d}{dt}\right|_{t=0} (\gamma(t)^{\mathrm{T}}\gamma(t)) = \dot\gamma(0)^{\mathrm{T}} + \dot\gamma(0).$$

Hence $T_1\,\mathrm{SO}_3(\mathbb{R}) \subset \{A \in \mathbb{R}^{3\times 3} \mid A^{\mathrm{T}} = -A\}$. On the other hand, if $A \in \mathbb{R}^{3\times 3}$ such that $A^{\mathrm{T}} = -A$, let $\gamma(t) = \exp(tA)$. By assumption, A^{T} commutes with A, hence

$$\gamma(t)^{\mathrm{T}}\gamma(t) = \exp(tA^{\mathrm{T}})\exp(tA) = \exp(t(A^{\mathrm{T}} + A)) = 1_3.$$

Moreover,

$$\det \gamma(t) = \exp(\operatorname{tr}(A)) = \exp\left(\frac{1}{2}\operatorname{tr}(A) + \frac{1}{2}\operatorname{tr}(A^{\mathrm{T}})\right) = \exp(0) = 1_3.$$

Thus $\gamma(t) \in \mathrm{SO}_3(\mathbb{R})$ for all $t \in \mathbb{R}$.

We conclude that $T_1\,\mathrm{SO}_3(\mathbb{R}) = \{A \in \mathbb{R}^{3\times 3} \mid A^{\mathrm{T}} = -A\}$. In fact, an analogous result holds in arbitrary dimensions:

Proposition 2.2.3. $T_1\,\mathrm{SO}_n(\mathbb{R}) = \{A \in \mathbb{R}^{n\times n} \mid A^{\mathrm{T}} = -A\}$.

Definition 2.2.4 (Special orthogonal Lie algebra)**.** The Lie algebra $\mathfrak{so}(n) = \{A \in \mathbb{R}^{n\times n} \mid A^{\mathrm{T}} = -A\}$ is called the *(special) orthogonal Lie algebra.* It is the same as the Lie algebra of $\mathrm{O}_n(\mathbb{R})$.

2.2.2 The special unitary group $\mathrm{SU}_2(\mathbb{C})$

Now let us turn to the special unitary group $\mathrm{SU}_n(\mathbb{C})$, which is defined as $\mathrm{SU}_n(\mathbb{C}) = \{U \in \mathbb{C}^{n\times n} \mid U^{\mathrm{H}}U = 1_n, \det U = 1\}$. An $n \times n$ matrix is unitary if and only if its row (or equivalently columns) are orthonormal with respect to the Euclidean inner product. Thus

$$\mathrm{SU}_2(\mathbb{C}) = \left\{\begin{pmatrix} a & b \\ -\bar{b} & \bar{a}\end{pmatrix} : a, b \in \mathbb{C},\ |a|^2 + |b|^2 = 1\right\}.$$

The following result is then not hard to see.

Lemma 2.2.5. *The map*

$$\mathrm{SU}_2(\mathbb{C}) \to S^3,\ \begin{pmatrix} a & b \\ -\bar{b} & \bar{a}\end{pmatrix} \mapsto (\operatorname{Re} a, \operatorname{Im} a, \operatorname{Re} b, \operatorname{Im} b)$$

is a homeomorphism

With this homeomorphism, S^3 also inherits a group structure from $\mathrm{SU}_2(\mathbb{C})$. To describe it, it is most convenient to work with quaternions. Formally, $\mathbb{H}$ is the skew field $\{t + x\mathrm{i} + y\mathrm{j} + z\mathrm{k} \mid t, x, y, z \in \mathbb{R}\}$, where the elements $\mathrm{i}, \mathrm{j}, \mathrm{k}$ satisfy the relations $\mathrm{i}^2 = \mathrm{j}^2 = \mathrm{k}^2 = -1$ and $\mathrm{ij} = \mathrm{k}$, $\mathrm{jk} = \mathrm{i}$, $\mathrm{ki} = \mathrm{j}$.

Quaternions can be represented as complex 2×2 matrices as follows: Let

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0\end{pmatrix},\ \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0\end{pmatrix},\ \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1\end{pmatrix}.$$

These are the *Pauli matrices.* The map

$$\mathbb{H} \to \mathbb{C}^{2\times 2},\ t + x\mathrm{i} + y\mathrm{j} + z\mathrm{k} \mapsto t1_2 - ix\sigma_1 - iy\sigma_2 - iz\sigma_3$$

is an algebra homomorphism (a map that respects addition and multiplication).

We can view S^3 as unit group of the quaternions, and the multiplication on S^3 is the one induced by quaternion multiplication.

If $g \in \mathbb{H} \setminus \{0\}$, then the conjugation map $g \cdot g^{-1}$ leaves $\{x\mathrm{i} + y\mathrm{j} + z\mathrm{k} \mid x, y, z \in \mathbb{R}\}$ invariant. If we identify $\{x\mathrm{i} + y\mathrm{j} + z\mathrm{k} \mid x, y, z \in \mathbb{R}\}$ with $\mathbb{R}^3$, we obtain a linear map $T_g \colon \mathbb{R}^3 \to \mathbb{R}^3$. If $\theta \in [0, 2\pi)$, $u \in S^2 \subset \mathbb{R}^3$ and $g = \cos\theta + \sin\theta(u^1\mathrm{i} + u^2\mathrm{j} + u^3\mathrm{k})$, we can explicitly compute

$$T_g v = u^{\mathrm{T}} v u + \cos 2\theta (u \times v) \times u + \sin 2\theta u \times v.$$

In particular, $T_g \in \mathrm{SO}_3(\mathbb{R})$ if g is a unit quaternion. The map $T \colon g \mapsto T_g$ can be viewed as a surjective 2-to-1 map from $\mathrm{SU}_2(\mathbb{C})$ to $\mathrm{SO}_3(\mathbb{R})$, a so-called *double covering*. On the level of Lie algebras, the map $DT(1)$ is an isomorphism.

Chapter 3

Measure and Integration Theory

3.1 Measures

Definition 3.1.1 (σ-algebra)**.** Let X be a set. A *σ-algebra* on X is a subset $\mathcal{A}$ of $\mathcal{P}(X)$ with the following three properties:

(a) $X \in \mathcal{A}$.

(b) If $A \in \mathcal{A}$, then $X \setminus A \in \mathcal{A}$.

(c) If $A_n \in \mathcal{A}$, $n \in \mathbb{N}$, then $\bigcup_n A_n \in \mathcal{A}$.

An element of $\mathcal{A}$ is called an ($\mathcal{A}$-)measurable set. A pair$(X, \mathcal{A})$ consisting of a set X and a σ-algebra $\mathcal{A}$ on X is called a *measurable space*.

Example 3.1.2. For every set X, $\{\emptyset, X\}$ and $\mathcal{P}(X)$ are σ-algebras.

Example 3.1.3. If X is an infinite set, then $\mathcal{A} = \{A \subset X \mid A \text{ countable or } X \setminus A \text{ countable}\}$ is a σ-algebra.

Lemma 3.1.4. *If X is a set and $(\mathcal{A}_i)_{i\in I}$ is a family of σ-algebras, then $\bigcap_{i\in I} \mathcal{A}_i$ is a σ-algebra.*

Definition 3.1.5 (Generated σ-algebra)**.** If $\mathcal{A} \subset \mathcal{P}(X)$, then the *$\sigma$-algebra generated $\sigma(\mathcal{A})$ by $\mathcal{A}$* is defined as

$$\mathcal{A} = \bigcap_{\mathcal{B}\supset\mathcal{A}\ \sigma\text{-algebra}} \mathcal{B}.$$

If $(X, \mathcal{T})$ is a topological space, the σ-algebra generated by $\mathcal{T}$ is called *Borel σ-algebra* and denoted by $\mathcal{B}(X)$.

Remark 3.1.6. Unless stated otherwise, we always endow $\mathbb{K}^n$ with the Borel σ-algebra induced by the Euclidean topology. Note that there is no explicit description of the elements of the Borel σ-algebra in this case. It is in fact much harder to find maps that are not Borel measurable than sets that are.

Definition 3.1.7 (Trace σ-algebra)**.** Let $(X, \mathcal{A})$ be a measurable space. If $S \in \mathcal{A}$, then

$$\mathcal{A}_S = \{A \cap S \mid A \in \mathcal{A}\} \subset \mathcal{P}(S)$$

is a σ-algebra on S, called the *trace σ-algebra*. If not stated otherwise, subsets of a measurable space are always endowed with the trace σ-algebra.

Definition 3.1.8 (Measurable map)**.** Let $(X, \mathcal{A})$, $(Y, \mathcal{B})$ be measurable spaces. A map $f\colon X \to Y$ is called ($\mathcal{A}$-$\mathcal{B}$-)measurable if $f^{-1}(B) \in \mathcal{A}$ for every $B \in \mathcal{B}$.

Lemma 3.1.9. *(a) The composition of measurable maps is measurable.*

(b) Let $f\colon X \to Y$ be a map, $\mathcal{A}$ a σ-algebra on X and $\mathcal{C} \subset \mathcal{P}(Y)$. If $\mathcal{B} = \sigma(\mathcal{C})$ and $f^{-1}(C) \in \mathcal{A}$ for all $C \in \mathcal{C}$, then f is $\mathcal{A}$-$\mathcal{B}$-measurable.

(c) Every continuous map between topological spaces is Borel measurable.

Lemma 3.1.10. *Let $(X, \mathcal{A})$ be a measurable space and $A \subset X$. The characteristic function $\mathbb{1}_A$ is measurable if and only if A is measurable.*

Definition 3.1.11 (Measure)**.** Let $(X, \mathcal{A})$ be a measurable space. A *measure* on $(X, \mathcal{A})$ is a map $\mu\colon X \to [0, \infty]$ with the following two properties:

- $\mu(\emptyset) = 0$.
- If A_n, $n \in \mathbb{N}$, and $A_n \cap A_m = \emptyset$ for $m \neq n$, then
$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

A *measure space* is a triple $(X, \mathcal{A}, \mu)$ consisting of a measurable space $(X, \mathcal{A})$ and a measure μ on $(X, \mathcal{A})$. A measure μ on $(X, \mathcal{A})$ is called *finite* if $\mu(X) < \infty$ and *σ-finite* if there exists a sequence (A_n) in $\mathcal{A}$ such that $\mu(A_n) < \infty$ for all $n \in \mathbb{N}$ and $X = \bigcup_{n=1}^{\infty} A_n$.

Example 3.1.12. If X is a countable set, then

$$\mu\colon \mathcal{P}(X) \to [0, \infty],\; A \mapsto \#A$$

is a measure, called the *counting measure*. Here $\#A$ denotes the cardinality (number of elements) of A. The counting measure on a countable space is always σ-finite and it is finite if and only if X is finite.

Example 3.1.13. If $(X, \mathcal{A})$ is a measurable space and $x \in X$, then

$$\delta_x \colon \mathcal{A} \to [0,1],\ A \mapsto \begin{cases} 1 & \text{if } x \in A \\ \infty & \text{otherwise} \end{cases}$$

is a finite measure, called the *Dirac measure*.

Lemma 3.1.14. *If $(X, \mathcal{A}, \mu)$ is a measure space and $(A_n)_{n\in\mathbb{N}}$ is an increasing sequence in $\mathcal{A}$, then*

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} \mu(A_n).$$

Proof. Let $B_n = A_n \setminus \bigcup_{k=1}^{n-1} A_k$. The sets B_n, $n \in \mathbb{N}$, are pairwise disjoint and we have

$$\begin{aligned} \mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) &= \mu\left(\bigcup_{n\in\mathbb{N}} B_n\right) \\ &= \sum_{n=1}^{\infty} \mu(B_n) \\ &= \sum_{n=1}^{\infty} (\mu(A_n) - \mu(A_{n-1})) \\ &= \lim_{n\to\infty} \mu(A_n). \end{aligned}$$

□

Definition 3.1.15 (Borel measure)**.** Let X be a topological space. A *Borel measure* on X is a measure μ on $(X, \mathcal{B}(X))$ such that

- $\mu(K) < \infty$ for every compact $K \subset X$,
- $\mu(A) = \sup\{\mu(K) \mid K \subset A \text{ compact}\}$ for every $A \in \mathcal{B}(X)$.

Theorem 3.1.16 (Existence and uniqueness of Haar measure)**.** *Let G be a Lie group. There exists a non-zero Borel measure μ on G such that $\mu(gA) = \mu(A)$ for all $A \in \mathcal{B}(G)$, $g \in G$. Moreover, the measure μ is uniquely determined up to multiplication by a positive constant.*

Definition 3.1.17 (Haar measure)**.** Let G be a Lie group. A non-zero Borel measure μ on G such that $\mu(gA) = \mu(A)$ for all $A \in \mathcal{B}(G)$, $g \in G$ is called a *left Haar measure* on G. If G is abelian, one simply calls μ a Haar measure.

Remark 3.1.18. There is a dual notion of a right Haar measure where the right translates Ag are considered instead of the left translates gA. In general, a right Haar measure need not be a left Haar measure, and vice versa. Lie groups for which a left Haar measure is also a right Haar measure are called *unimodular*. In addition to the obvious example of abelian Lie groups, compact Lie groups form another class of unimodular groups.

Corollary 3.1.19 (Existence and uniqueness of Lebesgue measure)**.** *There exists a unique translation-invariant Borel measure on* $\mathbb{R}^n$ *such that* $[0,1]^n$ *has measure* 1*.*

Definition 3.1.20 (Lebesgue measure)**.** The unique translation-invariant Borel measure $\mathcal{L}^n$ on $\mathbb{R}^n$ such that $\mathcal{L}^n([0,1]^n) = 1$ is called the *Lebesgue measure*.

Remark 3.1.21. There are various approaches to show the existence and uniqueness of the Lebesgue measure, all of which are at least somewhat technically involved. The difficulty is to define $\mathcal{L}^n(A)$ not only for "nice" sets A, but all Borel sets.

Lemma 3.1.22. *The Lebesgue measure is* σ*-finite.*

Proof. By translation invariance, every cube $a+[0,1]^n$ has Lebesgue measure 1. Thus

$$A_n = \bigcup_{k\in\mathbb{Z}^n, |k|\leq n} (k + [0,1]^n)$$

has finite Lebesgue measure. Since $\mathbb{R}^n = \bigcup_{n=1}^\infty A_n$, the Lebesgue measure is σ-finite. □

Definition 3.1.23 (Null set)**.** Let $(X, \mathcal{A}, \mu)$ be a measure space. A subset N of X is called a *null set* of there exists $A \in \mathcal{A}$ with $\mu(A) = 0$ such that $N \subset A$. A property is said to hold μ*-almost everywhere*, abbreviated as μ-a.e., if it holds on the complement of a null set.

Remark 3.1.24. Informally, a null set is a set of measure zero. Note however that a null set need not be measurable so that its measure is not defined. However, one can always extend a measure to a bigger σ-algebra, the *completion*, which contains all the null sets.

Definition 3.1.25 (Semi-finite, localizable measure)**.** Let $(X, \mathcal{A})$ be a measurable space. A measure μ on $(X, \mathcal{A})$ is called *semi-finite* if for every $A \in \mathcal{A}$ with $\mu(A) > 0$ there exists $B \subset A$ with $0 < \mu(B) < \infty$.

The measure μ is called *localizable* if it is semi-finite and for every family $(A_i)_{i\in I}$ in $\mathcal{A}$ there exists $S \in \mathcal{A}$ such that

- $\mu(A_i \setminus S) = 0$ for all $i \in I$,
- if $S' \in \mathcal{A}$ such that $\mu(A_i \setminus S') = 0$ for all $i \in I$, then $\mu(S \setminus S') = 0$.

Such a set S is called an *envelope* of $(A_i)_{i\in I}$.

Remark 3.1.26. If I is countable, then $\bigcup_{i\in I} A_i$ is an envelope of $(A_i)_{i\in I}$. If I is uncountable however, then $\bigcup_{i\in I} A_i \in \mathcal{A}$ is not guaranteed. Localizability of μ provides us with a way to take "unions of uncountable families up to measure zero".

Example 3.1.27. If X is a non-empty set and

$$\mu\colon \mathcal{P}(X) \to \{0,\infty\},\ A \mapsto \begin{cases} 0 & \text{if } A = \emptyset \\ \infty & \text{otherwise} \end{cases},$$

then μ is a measure that is not semi-finite.

Example 3.1.28. Every Borel measure is semi-finite.

Lemma 3.1.29. *Every σ-finite measure is localizable.*

Proof. Let (A_n) be a sequence in $\mathcal{A}$ such that $\mu(A_n) < \infty$ and $X = \bigcup_{n=1}^\infty A_n$. We can assume without loss of generality that $A_n \subset A_{n+1}$ for $n \in \mathbb{N}$. If $A \in \mathcal{A}$ with $\mu(A) > 0$, then

$$\mu(A \cap A_n) = \mu(A_n) - \mu(A_n \setminus A) \leq \mu(A_n) < \infty.$$

Moreover,

$$\mu(A \cap A_n) = \sum_{j=1}^{n-1} \mu(A \cap (A_{j+1} \setminus A_j)) \to \sum_{j=1}^{\infty} \mu(A \cap (A_{j+1} \setminus A_j)) = \mu(A).$$

In particular, $\mu(A \cap A_n) > 0$ for n sufficiently large. Thus μ is semi-finite.

To show that μ is localizable, let $(B_i)_{i\in I}$ be a family in $\mathcal{A}$. For $n \in \mathbb{N}$ let $c_n = \sup\{\mu(\bigcup_{j\in J} B_j \cap A_n) \mid J \subset I \text{ countable}\}$. For $k \in \mathbb{N}$ we can choose $J_{k,n} \subset I$ finite such that $\mu(\bigcup_{j\in J_{k,n}} B_j \cap A_n) \geq c_n - k^{-1}$ and $J_{k+1,n} \supset J_{k,n}$. Then $J_n = \bigcup_{k\in\mathbb{N}} J_k$ is countable and $S_n = \bigcup_{j\in J_n} B_j \cap A_n \in \mathcal{A}$ satisfies

$$c_n \geq \mu(S_n) = \lim_{k\to\infty} \mu\left(\bigcup_{j\in J_{k,n}} B_j \cap A_n\right) \geq \limsup_{k\to\infty}(c_n - k^{-1}) = c_n,$$

hence $\mu(S_n) = c_n$. Let $S = \bigcup_{n\in\mathbb{N}} S_n \in \mathcal{A}$.

To show that S is an envelope of $(B_i)_{i\in I}$, we first show that S_n is an envelope of $(B_i \cap A_n)_{i\in I}$. Note that

$$\mu(S_n) + \mu((B_i \cap A_n) \setminus S_n) = \mu(S_n \cup (B_i \cap A_n)) \leq c_n = \mu(S_n)$$

Thus $\mu((B_i \cap A_n) \setminus S_n) = 0$. Moreover, if $S'_n \in \mathcal{A}$ such that $\mu((B_i \cap A_n) \setminus S'_n) = 0$ for all $i \in I$, then

$$\mu(S_n \setminus S'_n) = \mu\left(\bigcup_{j\in J_n} (B_j \cap A_n) \setminus S'_n\right) \leq \sum_{j\in J_n} \mu((B_j \cap A_n) \setminus S'_n) = 0.$$

Therefore S_n is an envelope of $(B_i \cap A_n)_{i\in I}$. To see that S is an envelope of $(B_i)_{i\in I}$, one uses the monotonicity of the measure. The details are left as an exercise. $\square$

Example 3.1.30. Let J be a set. The counting measure on $(J, \mathcal{P}(J))$ is localizable. It is σ-finite if and only if J is countable. To see that the counting measure is localizable, it suffices to notice that arbitrary unions of subsets of J belong to the σ-algebra $\mathcal{P}(J)$. An envelope of $(A_i)_{i\in I}$ is therefore simply given by $\bigcup_{i\in I} A_i$. To see the statement about σ-finiteness, notice that sets with finite counting measure are exactly the finite subsets, and J is a countable union of finite subsets if and only if it is countable.

3.2 Integration

Definition 3.2.1 (Extended real line)**.** We endow $[0,\infty]$ with the σ-algebra $\mathcal{B}([0,\infty]) = \{A \subset [0,\infty] \mid A \cap [0,\infty) \in \mathcal{B}([0,\infty))\}$. Further, we extend addition and multiplication to $[0,\infty]$ by defining $a+\infty = \infty+a = \infty$ for all $a \in [0,\infty]$ and $a \cdot \infty = \infty \cdot a = \infty$ if $a \neq 0$ and $0 \cdot \infty = \infty \cdot 0 = 0$.

Definition 3.2.2 (Lebesgue integral)**.** Let $(X, \mathcal{A}, \mu)$ be a measure space and $f\colon X \to [0,\infty]$ measurable. The Lebesgue integral of f is defined as

$$\int_X f\, d\mu = \sup\left\{\sum_{j=1}^n \alpha_j \mu(A_j) \mid \alpha_j \in [0,\infty],\ A_j \in \mathcal{A},\ \sum_{j=1}^n \alpha_j \mathbb{1}_{A_j} \leq f\right\}.$$

Example 3.2.3. Let μ be the counting measure on $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$. If $f\colon \mathbb{N} \to [0,\infty]$, then

$$\int_{\mathbb{N}} f\, d\mu = \sum_{n=1}^\infty f(n).$$

Note that since f is nonnegative, the sum on the right side does not depend on the order of summation.

More generally, if μ is the counting measure on $(J, \mathcal{P}(J))$ for an arbitrary set J and $f\colon J \to [0, \infty]$, then

$$\int_J f \, d\mu = \sup_{F \subset J \text{ finite}} \sum_{j \in F} f(j).$$

That is a way to make sense of the sum $\sum_{j \in J} f(j)$ for uncountable sets J.

Proposition 3.2.4. *Let $(X, \mathcal{A}, \mu)$ be a measure space. The Lebesgue integral has the following properties:*

- *If $\alpha_n \in [0, \infty]$ and $A_n \in \mathcal{A}$ for $n \in \mathbb{N}$, then*

$$\int_X \sum_{n=1}^{\infty} \alpha_n \mathbb{1}_{A_n} \, d\mu = \sum_{n=1}^{\infty} \alpha_n \mu(A_n).$$

- *If $f, g\colon X \to [0, \infty]$ are measurable and $f \leq g$, then $\int_X f \, d\mu \leq \int_X g \, d\mu$.*
- *If $f, g\colon X \to [0, \infty]$ and $\alpha, \beta \in [0, \infty]$, then*

$$\int_X (\alpha f + g\mu) \, d\mu = \alpha \int_X f \, d\mu + \beta \int_X g \, d\mu.$$

Proposition 3.2.5. *Let $(X, \mathcal{A}, \mu)$ be a localizable measure space. If $(f_i)_{i \in I}$ is a family of measurable maps from X to $[0, \infty]$, then there exists a measurable function $f\colon X \to [0, \infty]$ with the following two properties:*

- *$f_i \leq f$ μ-a.e. for every $i \in I$.*
- *If $g\colon X \to [0, \infty]$ is measurable and $f_i \leq g$ μ-a.e. for every $i \in I$, then $f \leq g$ μ-a.e.*

Moreover, f is uniquely determined up to equality μ-a.e.

Definition 3.2.6 (Envelope)**.** If $(X, \mathcal{A}, \mu)$ is a measure space and $(f_i)_{i \in I}$ is a family of measurable functions from X to $[0, \infty]$, then a measurable function $f\colon X \to [0, \infty]$ that satisfies the properties from the previous proposition is called an *envelope* of $(f_i)_{i \in I}$.

Example 3.2.7. Let J be a set and μ the counting measure on $(J, \mathcal{P}(J))$. The envelope f of a family $(f_i)_{i \in I}$ of functions from J to $[0, \infty]$ is uniquely determined and given by $f(j) = \sup_{i \in I} f_i(j)$.

Example 3.2.8. Let $(X, \mathcal{A}, \mu)$ be an arbitrary measure space. If $(f_n)_{n\in\mathbb{N}}$ is a sequence of measurable functions from X to $[0, \infty]$, then an envelope f of $(f_n)_{n\in\mathbb{N}}$ is given by $f(x) = \sup_{n\in\mathbb{N}} f_n(x)$ for $x \in X$.

Example 3.2.9. For $x \in \mathbb{R}$ let $f_x = \mathbb{1}_{\{x\}}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{L}^1)$. A measurable function $f\colon \mathbb{R} \to [0, \infty]$ is an envelope of $(f_x)_{x\in\mathbb{R}}$ if and only if $f = 0$ $\mathcal{L}^1$-a.e. In particular, the pointwise supremum, which is the constant function 1, is *not* an envelope of $(f_x)_{x\in\mathbb{R}}$: Clearly, $f_n \leq f$ for all $n \in \mathbb{N}$. Moreover, if $g\colon X \to [0, \infty]$ is measurable and $f_n \leq g$ μ-a.e. for all $n \in \mathbb{N}$, let $A_n = \{x \in X \mid f_n(x) > g(x)\}$. By assumption, $\mu(A_n) = 0$ and $f(x) \leq g(x)$ for all $x \in X \setminus \bigcup_{n\in\mathbb{N}} A_n$. By σ-additivity of μ, we have $\mu(\bigcup_{n\in\mathbb{N}} A_n) = 0$. Hence $f \leq g$ μ-a.e.

Definition 3.2.10 (Directed set, Net)**.** A *directed set* is a pair $(I, \prec)$ consisting of a set I and a relation $\prec$ on I such that

- $i \prec i$ for all $i \in I$,
- $i \prec j$ and $j \prec k$ implies $i \prec k$ for all $i, j, k \in I$,
- for all $i, j \in I$ there exists $k \in I$ such that $i \prec k$ and $j \prec k$.

A net $(x_i)_{i\in I}$ in X is a map from a directed set I to X. In particular, if X is a set of functions, then a net $(f_i)_{i\in I}$ is called *increasing* if $i \prec j$ implies $f_i \leq f_j$.

Example 3.2.11. The natural numbers with their natural order form a directed set. Thus every sequence is a net.

Example 3.2.12. If J is any set, then $\mathcal{P}(J)$ with the preorder $A \prec B$ if $A \subset B$ is directed set. The same holds if one replaces $\mathcal{P}(J)$ by the set of finite subsets of J.

Theorem 3.2.13 (Monotone Convergence Theorem for Nets)**.** *Let* $(X, \mathcal{A}, \mu)$ *be a localizable measure space. If* $(f_i)_{i\in I}$ *is an increasing net of measurable function from* X *to* $[0, \infty]$ *and* f *is an envelope of* $(f_i)_{i\in I}$*, then*

$$\int_X f \, d\mu = \sup_{i\in I} \int_X f_i \, d\mu.$$

Theorem 3.2.14 (Monotone Convergence Theorem for Sequences)**.** *Let* $(X, \mathcal{A}, \mu)$ *be a measure space. If* $(f_n)_{n\in\mathbb{N}}$ *is an increasing sequence of measurable functions from* X *to* $[0, \infty]$ *and* f *is the pointwise limit* $\lim_{n\to\infty} f_n$*, then*

$$\int_X f \, d\mu = \lim_{n\to\infty} \int_X f_n \, d\mu.$$

Proof. By a previous example, the envelope of $(f_n)_{n\in\mathbb{N}}$ exists and coincides with f μ-a.e. However, to apply the previous theorem, we need a localizable measure space.

There are two cases: If $\int_X f_n\,d\mu = \infty$ for some $n \in \mathbb{N}$, then

$$\int_X f\,d\mu \geq \int_X f_n\,d\mu = \infty$$

by monotonicity of the integral. In this case, there is nothing left to show.

Let us assume that $\int_X f_n\,d\mu < \infty$ for all $n \in \mathbb{N}$ and let $A_{k,n} = \{x \in X \mid f_n(x) \geq 2^{-k}\}$, $A = \bigcup_{k,n\in\mathbb{N}} A_{k,n}$. We have

$$\mu(A_{k,n}) = \int_X \mathbb{1}_{A_{k,n}}\,d\mu \leq \int_X 2^k f_n\,d\mu < \infty.$$

Thus A is a countable union of sets with finite measure. As $f_n|_{X\setminus A} = 0$ for all $n \in \mathbb{N}$, we can restrict the integrals in the monotonce convergence theorem to A with the trace σ-algebra. As we have just seen, this measure space is σ-finite. Hence we can apply the monotonce convergence theorem for nets. $\square$

Lemma 3.2.15 (Fatou). *If $(X, \mathcal{A}, \mu)$ be a measure space. If (f_n) is a sequence of measurable functions from X to $[0, \infty]$, then the pointwise limit inferior $\liminf_{n\to\infty} f_n$ is measurable and*

$$\int_X \liminf_{n\to\infty} f_n\,d\mu \leq \liminf_{n\to\infty} \int_X f_n\,d\mu.$$

Proof. Let

$$g_n\colon X \to [0, \infty],\ g_n(x) = \inf_{k\geq n} f_k(x).$$

By definition, (g_n) is an increasing sequence of measurable functions from X to $[0, \infty]$ such that $g_n \leq f_n$ and $\lim_{n\to\infty} g_n(x) = \liminf_{n\to\infty} f_n(x)$ for all $x \in X$. By the monotone convergence theorem for sequences and monotonicity of the integral, we have

$$\int_X \liminf_{n\to\infty} f_n\,d\mu = \int_X \lim_{n\to\infty} g_n\,d\mu = \lim_{n\to\infty} \int_X g_n\,d\mu \leq \liminf_{n\to\infty} \int_X f_n\,d\mu. \quad \square$$

Remark 3.2.16. With an appropriate definition of the limit inferior of nets, there is also a version of Fatou's lemma for nets on localizable measure spaces.

So far, we have only integrated non-negative functions (that possibly take the value ∞). To define the integral of general real- or complex-valued functions, one needs to make sure that no "competing divergences" of the form $\infty - \infty$ arise. This can be done for example by decomposing the function into positive functions and requiring that these functions have finite integrals. This is the content of the next definition.

Definition 3.2.17 (Integrable function)**.** Let $(X, \mathcal{A}, \mu)$ be a measure space. A function $f\colon X \to \mathbb{R}$ is called *integrable* if it is measurable and the integrals

$$\int_X f_+ \, d\mu, \quad \int_X f_- \, d\mu$$

are both finite. In this case we define

$$\int_X f \, d\mu = \int_X f_+ \, d\mu - \int_X f_- \, d\mu.$$

Likewise, a function $f\colon X \to \mathbb{C}$ is called *integrable* if $\operatorname{Re} f$, $\operatorname{Im} f$ are integrable, and in this case we define

$$\int_X f \, d\mu = \int_X \operatorname{Re} f \, d\mu + i \int_X \operatorname{Im} f \, d\mu.$$

Lemma 3.2.18. *Let $(X, \mathcal{A}, \mu)$ be a measure space. The integral has the following properties*

- *If $f, g\colon X \to \mathbb{K}$ are are integrable and $f \leq g$, then $\inf_X f \, d\mu \leq \int_X g \, d\mu$.*
- *If $f, g\colon X \to \mathbb{K}$ are integrable and $\alpha, \beta \in \mathbb{K}$, then $\alpha f + \beta g$ is integrable and*
$$\int_X (\alpha f + \beta g) \, d\mu = \alpha \int_X f \, d\mu + \beta \int_X g \, d\mu.$$

Theorem 3.2.19 (Dominated Convergence Theorem)**.** *Let $(X, \mathcal{A}, \mu)$ be a measure space. If (f_n) is a sequence of measurable functions from X to $\mathbb{K}$ for which the pointwise limit $\lim_{n\to\infty} f_n$ exists and there exists an integrable function $g\colon X \to [0, \infty)$ such that $|f_n| \leq g$ μ-a.e. for every $n \in \mathbb{N}$, then $\lim_{n\to\infty} f_n$ is integrable and*

$$\int_X \lim_{n\to\infty} f_n \, d\mu = \lim_{n\to\infty} \int_X f_n \, d\mu.$$

Proof. We only prove the case $\mathbb{K} = \mathbb{R}$ here. The case $\mathbb{K} = \mathbb{C}$ can be deduced by considering real and imaginary parts of the involved functions.

Write f for $\lim_{n\to\infty} f_n$. Upon changing g on a null set, we may assume that $|f_n(x)| \leq g(x)$ for all $x \in X$. By monotonicity of the integral,

$$\int_X (f_n)_\pm \, d\mu \leq \int_X |f_n| \, d\mu \leq \int_X g \, d\mu < \infty.$$

Thus f_n is integrable. Similarly, $|f| = \lim_{n\to\infty}|f_n| \leq g$ implies that f is integrable.

Let $h_n = g - f_n$, which is non-negative by assumption. By Fatou's lemma,

$$\begin{aligned}\int_X (g - f)\, d\mu &= \int_X \liminf_{n\to\infty} h_n\, d\mu \\ &\leq \liminf_{n\to\infty} \int_X h_n\, d\mu \\ &= \int_X g\, d\mu - \limsup_{n\to\infty} \int_X f_n\, d\mu.\end{aligned}$$

Hence

$$\int_X f\, d\mu \geq \limsup_{n\to\infty} \int_X f_n\, d\mu.$$

If we apply Fatou's lemma to $g + f_n$ instead, we obtain

$$\int_X f\, d\mu \leq \liminf_{n\to\infty} \int_X f_n\, d\mu.$$

These two inequalities combined yield the claim. □

3.3 Lebesgue spaces

Definition 3.3.1 (Lebesgue space)**.** Let $(X, \mathcal{A}, \mu)$ be a measure space and $p \in [1, \infty)$. We write $\mathcal{L}^p(X, \mu)$ for the set of all functions from X to $\mathbb{K}$ for which $|f|^p$ is integrable. We define an equivalence relation on $\mathcal{L}^p(X, \mu)$ by $f \sim g$ if $f = g$ μ-a.e. The *Lebesgue space* $L^p(X, \mu)$ is the set of all equivalence classes of the equivalence relation $\sim$ on $\mathcal{L}^p(X, \mu)$.

Moreover, we write $\mathcal{L}^\infty(X, \mu)$ for the set of all measurable functions f from X to $\mathbb{K}$ for which there exists $C > 0$ such that $|f| \leq C$ μ-a.e. The set of all μ-a.e. equivalence classes in $\mathcal{L}^\infty(X, \mu)$ is denoted by $L^\infty(X, \mu)$.

Definition 3.3.2 (L^p semi-norm)**.** Let $(X, \mathcal{A}, \mu)$ be a measure space and $p \in [1, \infty]$. The $\mathcal{L}^p$ semi-norm on $\mathcal{L}^p(X, \mu)$ is defined by

$$\|f\|_p = \left(\int_X |f|^p\, d\mu\right)^{1/p}.$$

if $p < \infty$ and by

$$\|f\|_\infty = \inf\{C > 0 : |f| \leq C\ \mu\text{-a.e.}\}.$$

if $p = \infty$.

Proposition 3.3.3 (Hölder inequality)**.** *Let $(X, \mathcal{A}, \mu)$ be a measure space and $p, q \in [1, \infty]$ such that $\frac{1}{p} + \frac{1}{q} = 1$. If $f \in \mathcal{L}^p(X, \mu)$ and $g \in \mathcal{L}^q(X, \mu)$, then $fg \in \mathcal{L}^1(X, \mu)$ and*

$$\|fg\|_1 \leq \|f\|_p \|g\|_q.$$

Proof. Let $a, b > 0$. Since log is concave, we have

$$\log(ab) = \frac{1}{p} \log a^p + \frac{1}{q} \log b^q \leq \log\left(\frac{a^p}{p} + \frac{b^q}{q}\right).$$

Thus $ab \leq \frac{a^p}{p} + \frac{b^q}{q}$. It follows that

$$\int_X |fg| \, d\mu \leq \frac{1}{p} \int_X |f|^p \, d\mu + \frac{1}{q} \int_X |g|^q \, d\mu.$$

If we replace f by λf and g by g/λ for $\lambda > 0$, we obtain

$$\|fg\|_1 \leq \frac{\lambda^p}{p} \|f\|_p^p + \frac{1}{q\lambda^q} \|g\|_q^q.$$

If we optimize the right side over $\lambda > 0$, we obtain the desired inequality. $\square$

Proposition 3.3.4 (Minkowski inequality)**.** *Let $(X, \mathcal{A}, \mu)$ be a measure space and $p \in [1, \infty]$. If $f, g \in \mathcal{L}^p(X, \mu)$, then $f + g \in \mathcal{L}^p(X, \mu)$ and $\|f + g\|_p \leq \|f\|_p + \|g\|_p$.*

Let V be a normed space. In the following we write V^* for $\mathcal{L}(V, \mathbb{K})$, which is the set of all *bounded* linear maps from V to $\mathbb{K}$. This is strictly smaller than the set of all not necessarily bounds linear maps from V to $\mathbb{K}$ if V is infinite-dimensional.

Proposition 3.3.5. *Let $(X, \mathcal{A}, \mu)$ be a measure space and $p \in [1, \infty]$.*

(a) The space $\mathcal{L}^p(X, \mu)$ is a vector space.

(b) $\|f\|_p = 0$ if and only if $f = 0$ μ-a.e.

(c) The map

$$\|\cdot\|_p \colon L^p(X, \mu) \to [0, \infty), \; f \mapsto \|[f]_\mu\|_p$$

defines a norm on $L^p(X, \mu)$. Here $[f]_\mu$ denotes the μ-a.e. equivalence class of f.

(d) *If* $p, q \in (1, \infty)$ *with* $\frac{1}{p} + \frac{1}{q} = 1$, *then the map*

$$L^p(X, \mu) \to L^q(X, \mu)^*, \; f \mapsto \left(g \mapsto \int_X fg \, d\mu \right)$$

is an isometric isomorphism. The same is true for $p = \infty$, $q = 1$ *if and only* $(X, \mathcal{A}, \mu)$ *is localizable.*

Remark 3.3.6. One can also define L^p spaces for $p \in (0, 1)$ in the same way. However, the functional

$$f \mapsto \left(\int_X |f|^p \, d\mu \right)^{1/p}$$

fails to be a norm for $p < 1$.

Example 3.3.7. If I is a set and μ the counting measure on $\mathcal{P}(I)$, then $L^p(I, \mu) = \ell^p(I)$, the set of all families $(a_i)_{i \in I}$ with $\sum_{i \in I} |a_i|^p < \infty$.

Proposition 3.3.8. *If* V, W *are normed spaces and* W *is complete, then* $\mathcal{L}(V, W)$ *is complete in the operator norm. In particular,* V^* *is complete.*

Proof. Let (T_n) be a Cauchy sequence in $\mathcal{L}(V, W)$. For every $v \in V$, we have $\|T_n v - T_m v\| \leq \|T_n - T_m\|_{\mathrm{op}} \|v\|$. Since W is complete, there exists $T(v)$ such that $T_n v \to T(v)$ in W. It is not hard to see that the assignment $v \mapsto T(v)$ is linear. Moreover, if $m \in \mathbb{N}$ such that $\|T_n - T_m\| \leq 1$ for $n \geq m$, then

$$\|T(v)\| = \lim_{n \to \infty} \|T_n v\| \leq \liminf_{n \to \infty} \|T_n v - T_m v\| + \|T_m v\| \leq (1 + \|T_m\|_{\mathrm{op}}) \|v\|.$$

Thus T is bounded.

To finish the proof, we have to show that $T_n \to T$ in operator norm. Let $\varepsilon > 0$ and $N \in \mathbb{N}$ such that $\|T_n - T_m\| \leq \varepsilon$ for $m, n \geq N$. If $v \in V$ and $n \geq N$, then

$$\|T_n v - Tv\| = \lim_{m \to \infty} \|T_n v - T_m v\| \leq \liminf_{m \to \infty} \|T_n - T_m\|_{\mathrm{op}} \|v\| \leq \varepsilon \|v\|.$$

Therefore $\|T_n - T\|_{\mathrm{op}} \leq \varepsilon$. $\square$

Corollary 3.3.9. *Let* $(X, \mathcal{A}, \mu)$ *be a measure space. For* $p \in (1, \infty)$, *the Lebesgue space* $L^p(X, \mu)$ *is a Banach space.*

Remark 3.3.10. The same is true for $p = 1$ and $p = \infty$, but one cannot appeal to duality (for $p = 1$) or only in the case of localizable measure spaces (for $p = \infty$).

3.4 Integration with respect to the Lebesgue measure and Fubini's theorem

The definition of the Lebesgue integral suggests possible numerical approaches to the integral. However, it is not obvious how to compute the Lebesgue integral symbolically even for nice functions. One advantage of the Riemann integral is that we can compute it for continuously differentiable functions f by finding a primitive function, that is, a function F such that $F' = f$. Luckily enough, both integral coincide for this class (and a broader class) of functions.

Theorem 3.4.1. *Let $a, b \in \mathbb{R}$ with $a < b$. A $f\colon [a,b] \to \mathbb{K}$ is Riemann integrable if and only if its bounded and continuous at $\mathcal{L}^1$-a.e. point. In this case,*

$$\int_a^b f(x)\,dx = \int_{[a,b]} f\,d\mathcal{L}^1.$$

Remark 3.4.2. The result is no longer true if one considers improper Riemann integrals. For example, the function

$$(0,\infty) \to \mathbb{R},\; x \mapsto \frac{\sin x}{x}$$

has an improper Riemann integral, but is not Lebesgue integrable.

The Lebesgue integral in higher dimensions can be reduced to iterated one-dimensional integrals by Tonelli's and Fubini's theorem. Both have similar statements, just under different conditions to avoid problems with divergences of the form $\infty - \infty$.

Theorem 3.4.3 (Tonelli)**.** *If $f\colon \mathbb{R}^m \times \mathbb{R}^n \to [0,\infty]$ is a Borel-measurable function, then $f(x,\cdot)$ is Borel-measurable for $\mathcal{L}^m$-a.e. $x \in \mathbb{R}^m$, $f(\cdot,y)$ is Borel-measurable for $\mathcal{L}^n$-a.e. $y \in \mathbb{R}^n$ and*

$$\begin{aligned}\int_{\mathbb{R}^m\times\mathbb{R}^n} f\,d\mathcal{L}^{m+n} &= \int_{\mathbb{R}^m}\left(\int_{\mathbb{R}^n} f(x,y)\,d\mathcal{L}^n(y)\right)d\mathcal{L}^m(x)\\ &= \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^m} f(x,y)\,d\mathcal{L}^m(x)\right)d\mathcal{L}^n(y).\end{aligned}$$

Theorem 3.4.4 (Fubini)**.** *If $f\colon \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{K}$ is an $\mathcal{L}^{m+n}$-integrable function, then $f(x,\cdot)$ is $\mathcal{L}^n$-integrable for $\mathcal{L}^m$-a.e. $x \in \mathbb{R}^m$, $f(\cdot,y)$ is $\mathcal{L}^m$-integrable* *for $\mathcal{L}^n$-a.e. $y \in \mathbb{R}^n$ and*

$$\begin{aligned}\int_{\mathbb{R}^m \times \mathbb{R}^n} f \, d\mathcal{L}^{m+n} &= \int_{\mathbb{R}^m} \left(\int_{\mathbb{R}^n} f(x,y) \, d\mathcal{L}^n(y) \right) d\mathcal{L}^m(x) \\ &= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} f(x,y) \, d\mathcal{L}^m(x) \right) d\mathcal{L}^n(y).\end{aligned}$$

Remark 3.4.5. While we only state these results for the Lebesgue measure, they are valid more generally. To formulate them, one needs the notion of product measure of σ-finite measures, which we did not introduce in this course.

Chapter 4

Operators and Spectral Theory

4.1 Hilbert spaces

Definition 4.1.1 (Inner product)**.** Let V be a vector space over $\mathbb{K}$. An *inner product* on V is a map $\langle\,\cdot\,,\cdot\,\rangle\colon V\times V\to\mathbb{K}$ with the following three properties:

(a) Linearity in the second argument:

$$\langle\xi,\lambda\eta+\mu\zeta\rangle=\lambda\langle\xi,\eta\rangle+\mu\langle\xi,\zeta\rangle$$

for all $\xi,\eta,\zeta\in V$ and $\lambda,\mu\in\mathbb{K}$

(b) Conjugate symmetry:

$$\langle\eta,\xi\rangle=\overline{\langle\xi,\eta\rangle}$$

for all $\xi,\eta\in V$.

(c) Positive definiteness:

$$\langle\xi,\xi\rangle>0$$

for all $\xi\in V\setminus\{0\}$.

A vector space with an inner product is called an *inner product space.*

Remark 4.1.2.
- The are two different conventions for inner products. Some authors assume linearity in the first argument, some authors linearity in the second argument. Note that this only makes a difference if $\mathbb{K}=\mathbb{C}$ (see below).
- $\bar{\cdot}$ denotes the complex conjugate of an element. In the case $\mathbb{K}=\mathbb{R}$, (b) reduces to ordinary symmetry $\langle\eta,\xi\rangle=\langle\xi,\eta\rangle$.

- If $\mathbb{K} = \mathbb{C}$, property (b) implies $\langle \xi, \xi \rangle \in \mathbb{R}$ for all $\xi \in V$. Thus the inequality in (c) makes sense.
- Properties (a) and (b) together imply that an inner product is conjugate linear in the first argument:

$$\langle \lambda\xi + \mu\eta, \zeta \rangle = \overline{\lambda}\langle \xi, \zeta \rangle + \overline{\mu}\langle \eta, \zeta \rangle.$$

Example 4.1.3. The Euclidean inner product on $\mathbb{K}^n$:

$$\langle \xi, \eta \rangle = \sum_{j=1}^{n} \overline{\xi_j}\eta_j, \; \xi, \eta \in \mathbb{K}^n$$

Example 4.1.4. Let $(X, \mathcal{A}, \mu)$ be a measure space. The L^2-inner product on $L^2(X, \mu)$ is defined as

$$\langle [f], [g] \rangle = \int_X \overline{f} g \, d\mu, \quad f, g \in \mathcal{L}^2(X, \mathcal{A}).$$

One can check that this definition is indeed independent of the chosen representatives f, g.

In the special case when $X = \{1, \ldots, n\}$, $\mathcal{A} = \mathcal{P}(X)$ and μ is the counting measures, one obtains the Euclidean inner product from the previous example. In this case, one does not have to quotient out almost-everywhere equal functions because the measure has no-nontrivial null sets.

Proposition 4.1.5 (Cauchy–Schwarz inequality)**.** *If V is an inner product space, then*

$$|\langle \xi, \eta \rangle| \leq \langle \xi, \xi \rangle^{1/2} \langle \eta, \eta \rangle^{1/2}$$

for all $\xi, \eta \in V$.

Proof. First note that

$$0 \leq \langle \xi - \eta, \xi - \eta \rangle = \langle \xi, \xi \rangle + \langle \eta, \eta \rangle \underbrace{- \langle \xi, \eta \rangle - \langle \eta, \xi \rangle}_{-2\,\mathrm{Re}\langle \xi, \eta \rangle}.$$

Thus $\mathrm{Re}\langle \xi, \eta \rangle \leq \frac{1}{2}\langle \xi, \xi \rangle + \frac{1}{2}\langle \eta, \eta \rangle$.

Let $\lambda \in \mathbb{K}$ with $|\lambda| = 1$ such that $\lambda\langle \xi, \eta \rangle = |\langle \xi, \eta \rangle|$. If we apply the previous inequality to ξ and $\lambda\eta$, we obtain

$$|\langle \xi, \eta \rangle| = \mathrm{Re}\langle \xi, \lambda\eta \rangle \leq \frac{1}{2}\langle \xi, \xi \rangle + \frac{1}{2}\underbrace{\langle \lambda\eta, \lambda\eta \rangle}_{|\lambda|^2\langle \eta, \eta \rangle}.$$

Hence $|\langle \xi, \eta \rangle| \le \frac{1}{2}\langle \xi, \xi \rangle + \frac{1}{2}\langle \eta, \eta \rangle$.

Let $\mu \in \mathbb{K} \setminus \{0\}$. If we apply the previous inequality to ξ/μ and $\mu\eta$, we obtain

$$|\langle \xi, \eta \rangle| = |\langle \xi/\mu, \mu\eta \rangle| \le \frac{1}{2|\mu|^2}\langle \xi, \xi \rangle + \frac{|\mu|^2}{2}\langle \eta, \eta \rangle.$$

We can assume $\xi, \eta \neq 0$, since otherwise the left side is zero by linearity in the second argument. If we apply the previous inequality with $\mu = \frac{\langle \xi, \xi \rangle^{1/4}}{\langle \eta, \eta \rangle^{1/4}}$, we obtain

$$|\langle \xi, \eta \rangle| \le \frac{1}{2}\langle \xi, \xi \rangle^{1/2}\langle \eta, \eta \rangle^{1/2} + \frac{1}{2}\langle \xi, \xi \rangle^{1/2}\langle \eta, \eta \rangle^{1/2} = \langle \xi, \xi \rangle^{1/2}\langle \eta, \eta \rangle^{1/2}. \qquad \square$$

Lemma 4.1.6. *If V is a vector space over $\mathbb{K}$ and $\langle \cdot\,, \cdot \rangle$ is an inner product on V, then*

$$\|\cdot\| \colon V \to [0, \infty),\ \xi \mapsto \langle \xi, \xi \rangle^{1/2}$$

is a norm.

Proof. The only property of a norm that is not obvious from the properties of an inner product is the triangle inequality. To prove it, we use the Cauchy–Schwarz inequality: If $\xi, \eta \in V$, then

$$\|\xi+\eta\|^2 = \|\xi\|^2+2\operatorname{Re}\langle \xi, \eta \rangle+\|\eta\|^2 \le \|\xi\|^2+2\|\xi\|\|\eta\|+\|\eta\|^2 = (\|\xi\|+\|\eta\|)^2. \ \square$$

Remark 4.1.7. The norm from the previous lemma is called the norm induced by the inner product. The norm in turn gives rise to a metric (and the metric to a topology). As such, metric properties like completeness make sense for inner product spaces.

Proposition 4.1.8. *Let V be an inner product space and $\xi, \eta \in V$.*

(a) Polarization identity: If V is a real inner product space, then

$$\langle \xi, \eta \rangle = \frac{1}{4}(\|\xi + \eta\|^2 - \|\xi - \eta\|^2),$$

and if V is a complex inner product space, then

$$\langle \xi, \eta \rangle = \frac{1}{4}\sum_{k=0}^{3} i^{-k}\|\xi + i^k\eta\|^2.$$

(b) Parallelogram identity: $\|\xi - \eta\|^2 + \|\xi + \eta\|^2 = 2\|\xi\|^2 + 2\|\eta\|^2$.

Remark 4.1.9. Both identities from the previous proposition follow from a direct computation. They are very useful: The polarization identity shows that an inner product is uniquely determined by its induced norm and the parallelogram identity characterizes inner product spaces among all normed spaces – a norm is induced by an inner product if and only if it satisfies the parallelogram identity.

Definition 4.1.10 (Hilbert space)**.** A *Hilbert space* is an inner product space which is complete.

Example 4.1.11. The Euclidean space $\mathbb{K}^n$ with the Euclidean inner product is a Hilbert space.

Example 4.1.12. The Lebesgue space $L^2(X, \mu)$ with the L^2 inner product is a Hilbert space.

Example 4.1.13. Let c_c denote the space of all functions from $\mathbb{N}$ to $\mathbb{K}$ that have finite support, i.e., $\{n \in \mathbb{N} \mid f(n) \neq 0\}$ is finite for every $f \in c_c$. An inner product on c_c can be defined by

$$\langle f, g \rangle = \sum_{n=1}^{\infty} \overline{f(n)} g(n).$$

Note that only finitely many summands are non-zero, so there is no convergence problem. This inner product is not complete.

Definition 4.1.14 (Orthonormal family)**.** Let H be a Hilbert space. A family $(\xi_i)_{i \in I}$ is called *orthogonal* if $\langle \xi_i, \xi_j \rangle = 0$ for $i \neq j$. It is called *orthonormal* if it is orthogonal and $\|\xi_i\| = 1$ for every $i \in I$.

Proposition 4.1.15 (Bessel's inequality)**.** *If H is a Hilbert space and $(\xi_i)_{i \in I}$ is an orthonormal family, then*

$$\sum_{j \in J} |\langle \xi_j, \eta \rangle|^2 \leq \|\eta\|^2$$

for every $\eta \in H$ and every finite subset J of I.

Definition 4.1.16 (Orthonormal basis)**.** Let H be a Hilbert space. An *orthonormal basis* is an orthonormal family $(e_i)_{i \in I}$ in H that satisfies

$$\sup_{J \subset I \text{ finite}} \sum_{j \in J} |\langle e_j, \eta \rangle|^2 = \|\eta\|^2$$

for all $\eta \in H$. A Hilbert space is called *separable* if it admits a countable orthonormal basis.

Remark 4.1.17. Every Hilbert space admits an orthonormal basis, but it may have a large cardinality. For most applications in mathematics and quantum mechanics, it suffices to study separable Hilbert spaces.

Definition 4.1.18 (Unitary operator)**.** Let H, K be Hilbert spaces. A linear operator $U\colon H \to K$ is called an *isometry* if $\|U\xi\| = \|\xi\|$ for all $\xi \in H$. A surjective isometry is called *unitary*.

Lemma 4.1.19. *Let H, K be Hilbert spaces. If $U\colon H \to K$ is an isometry, then $\langle U\xi, U\eta\rangle = \langle \xi, \eta\rangle$ for all $\xi, \eta \in H$.*

Proof. The map

$$\langle\cdot,\cdot\rangle_U\colon H \times H \to \mathbb{K},\ (\xi,\eta) \mapsto \langle U\xi, U\eta\rangle$$

is an inner product on H. Since U is an isometry, we have $\langle \xi, \xi\rangle = \langle U\xi, U\xi\rangle$ for all $\xi \in H$. By the polarization identity, $\langle\cdot,\cdot\rangle_U = \langle\cdot,\cdot\rangle$. □

Theorem 4.1.20. *If H is a Hilbert space, then there exists a set J and a unitary operator $U\colon H \to \ell^2(J)$. If H is separable, then J can be chosen countable.*

Proof. Let $(e_j)_{j\in J}$ be an orthonormal basis of H. For $\xi \in H$ let

$$U\xi\colon J \to \mathbb{K},\ (U\xi)(j) = \langle e_j, \xi\rangle.$$

By the definition of an orthonormal basis, $U\xi \in \ell^2(J)$ and $\|U\xi\| = \|\xi\|$. Thus U is an isometry from H to $\ell^2(J)$.

In particular, U preserves Cauchy sequences. Thus $U(H)$ is complete and thus closed in $\ell^2(J)$. To show that U is surjective, it suffices therefore to show that $\overline{U(H)} = \ell^2(J)$. For $f \in \ell^2(J)$ and $F \subset J$ finite let $\xi_F = \sum_{j\in F} f(j)e_j \in H$. Then

$$\|f - U(\xi_F)\|_2^2 = \sum_{j\notin F} |f(j)|^2.$$

Recall that the counting measure on $(J, \mathcal{P}(J))$ is localizable. By the monotone convergence theorem for nets, we have

$$\|f\|_2^2 = \sup_{F\subset J \text{ finite}} \sum_{j\in F} |f(j)|^2.$$

In particular, for every $\varepsilon > 0$ there exists $F \subset J$ finite such that

$$\sum_{j\notin F} |f(j)|^2 = \|f\|_2^2 - \sum_{j\in F} |f(j)|^2 < \varepsilon.$$

Therefore $f \in \overline{U(H)}$. □

Corollary 4.1.21. *If H is a separable infinite-dimensional Hilbert space and $(e_n)_{n\in\mathbb{N}}$ is an orthonormal basis, then*

$$\lim_{n\to\infty}\sum_{k=1}^{n}\langle e_k,\xi\rangle e_k=\xi$$

for every $\xi\in H$.

Theorem 4.1.22 (Riesz representation theorem)**.** *Let H be a Hilbert space. For every $\varphi\in\mathcal{L}(H,\mathbb{K})$ there exists a unique $\eta\in H$ such that $\varphi(\xi)=\langle\eta,\xi\rangle$ for all $\xi\in H$. Moreover, $\|\eta\|=\|\varphi\|_{\mathrm{op}}$.*

Vice versa, for every $\eta\in H$ the map $\varphi_\eta\colon H\to\mathbb{K}$, $\xi\mapsto\langle\eta,\xi\rangle$ is a bounded linear map.

Proof. Since H is isometrically isomorphic to $\ell^2(J)$ for some set J, this can easily be deduced from the Riesz representation theorem for L^2 spaces. $\square$

Definition 4.1.23 (Orthogonal complement)**.** Let V be an inner product space. Two elements ξ, η are called *orthogonal*, denoted by $\xi\perp\eta$, if $\langle\xi,\eta\rangle=0$. For a subset S of V, the *orthogonal complement* $S^\perp$ is defined as $S^\perp=\{\xi\in V\mid\xi\perp\eta\text{ for all }\eta\in S\}$.

Lemma 4.1.24. *Let V be an inner product space.*

(a) The orthogonal complement of any subset of V is a closed subspace of V.

(b) If V is a Hilbert space, then $K\subset V$ is a closed linear subspace if and only if $K^{\perp\perp}=K$.

(c) The inner product space V is a Hilbert space if and only if $K^{\perp\perp}=K$ for every closed subspace K of V.

Proposition 4.1.25. *Let H be a Hilbert space and $K\subset H$ a closed linear subspace. For every $\xi\in H$ there exists a unique decomposition $\xi=\eta+\zeta$ with $\eta\in K$ and $\zeta\in K^\perp$.*

Proof. We only give the proof in the case when K is separable here. The proof in the general case is similar, one just has to deal properly with uncountable orthonormal bases. Let $(e_j)_{j\in J}$ be an orthonormal basis of K with J finite or $J=\mathbb{N}$. For $\xi\in H$ let $\eta=\sum_{j\in J}\langle e_j,\xi\rangle e_j$ if J finite and

$$\eta=\lim_{n\to\infty}\sum_{j=1}^{n}\langle e_j,\xi\rangle e_j$$

if $J = \mathbb{N}$. To see that the limit exists, let $m > n$. We have

$$\begin{aligned}\left\|\sum_{j=1}^{m}\langle e_j,\xi\rangle e_j - \sum_{j=1}^{n}\langle e_j,\xi\rangle e_j\right\| &= \left\|\sum_{j=n+1}^{m}\langle e_j,\xi\rangle e_j\right\|^2 \\ &= \sum_{j=n+1}^{m}|\langle e_j,\xi\rangle|^2 \\ &\le \sum_{j=n+1}^{\infty}|\langle e_j,\xi\rangle|^2.\end{aligned}$$

Since $\sum_{j=1}^{\infty}|\langle e_j,\xi\rangle|^2 \le \|\xi\|^2$ by Bessel's inequality, the right side of the previous displayed formula goes to zero as $n \to \infty$. This implies that $\left(\sum_{j=1}^{n}\langle e_j,\xi\rangle e_j\right)_n$ is a Cauchy sequence. As H is complete, it converges.

Let $\zeta = \xi - \eta$. If J is finite, then

$$\langle\zeta, e_k\rangle = \langle\xi - \eta, e_k\rangle = \langle\xi, e_k\rangle - \sum_{j\in J}\langle\xi, e_j\rangle\langle e_j, e_k\rangle = 0,$$

that is, $\zeta \perp e_k$. The same is true if $J = \mathbb{N}$ using a limiting argument.Since $(e_j)_{j\in J}$ is an orthonormal basis of K, every element of K is in the closed linear span of $\{e_j \mid j \in J\}$. Thus $\zeta \in K^\perp$. The equality $\xi = \eta + \zeta$ holds by definition. This settles the existence part of the statement.

For uniqueness, let $\eta, \eta' \in K$ and $\zeta, \zeta' \in K^\perp$ such that $\eta + \zeta = \eta' + \zeta'$. Then $\eta - \eta' = \zeta' - \zeta$ and $\langle\eta - \eta', \zeta' - \zeta\rangle = 0$. Thus $\eta = \eta'$ and $\zeta' = \zeta$. □

Definition 4.1.26 (Orthogonal projection)**.** Let H be a Hilbert space and $K \subset H$ a closed linear subspace. The map P_K from H to H that maps $\xi \in H$ to the unique element $\eta \in K$ such that $\xi - \eta \in K^\perp$ is called the *(orthogonal) projection* onto K.

Lemma 4.1.27. *Let H be a Hilbert space.*

(a) If $K \subset H$ is a closed linear subspace, then P_K is a bounded linear map.

(b) For $P \in \mathcal{L}(H)$ there exists a closed linear subspace $K \subset H$ such that $P = P_K$ if and only if $P^2 = P$ and $\langle P\xi, \eta\rangle = \langle\xi, P\eta\rangle$ for all $\xi, \eta \in H$.

Proof. (a) To show linearity of P_K, let $\xi_1, \xi_2 \in H$ and $\alpha_1, \alpha_2 \in \mathbb{K}$. We have

$$\begin{aligned}(\alpha_1\xi_1 + \alpha_2\xi_2) - (\alpha_1 P_K(\xi_1) + \alpha_2 P_K(\xi_2)) &= \alpha_1(\xi_1 - P_K(\xi_1)) + \alpha_2(\xi_2 - P_K(\xi_2)) \\ &\in K^\perp.\end{aligned}$$

Thus $P_K(\alpha_1\xi_1 + \alpha_2\xi_2) = \alpha_1 P_K(\xi_1) + \alpha_2 P_K(\xi_2)$.

To show boundedness, let $\xi \in K$. Since $\xi - P_K\xi \perp \xi$, we have

$$\|\xi\|^2 = \|P_K\xi\|^2 + \|\xi - P_K\xi\|^2 \geq \|P_K\xi\|^2.$$

Thus P_K is bounded (with $\|P_K\|_{\text{op}} \leq 1$).

(b) First assume that $P = P_K$ for some closed linear subspace K of H. Clearly, $P^2 = P$. Let $\xi, \eta \in H$. Since $\xi - P_K\xi, \eta - P_K\eta \perp K$, we have

$$\langle P_K\xi, \eta\rangle = \langle P_K\xi, \eta - P_K\eta + P_K\eta\rangle = \langle P_K\xi, P_K\eta\rangle = \langle \xi, P_K\eta\rangle.$$

Now assume conversely that $P \in \mathcal{L}(H)$ with $P^2 = P$ and $\langle P\xi, \eta\rangle = \langle \xi, P\eta\rangle$. Let $K = (\ker P)^\perp$. If $\xi \in H$ and $\eta \in \ker P$, then

$$\langle P\xi, \eta\rangle = \langle \xi, P\eta\rangle = 0.$$

Thus $P\xi \in K$. Furthermore, $P(\xi - P\xi) = P\xi - P^2\xi = 0$. Hence $\xi - P\xi \in \ker P = K^\perp$. Hence $P = P_K$. □

4.2 Uniform boundedness, open mapping and closed graph theorem

In this section we will prove three of the cornerstone results of functional analysis – the uniform boundedness principle, the open mapping theorem and the closed graph theorem. These three results are intimately related. They are usually presented as consequences of Baire's theorem, which is quite useful on its own. However, in this course, we will take a short cut that avoids Baire's theorem altogether. All we need is the following lemma with a three-line proof.

Lemma 4.2.1. *Let X, Y be normed spaces and $T\colon X \to Y$ a bounded linear operator. For all $x \in X$ and $r > 0$ we have*

$$\sup_{\|y-x\|\leq r} \|Ty\| \geq \|T\|r.$$

Proof. For $\xi \in X$ we have

$$\|T\xi\| \leq \frac{1}{2}(\|T(x-\xi)\| + \|T(x+\xi)\|) \leq \max\{\|T(x+\xi)\|, \|T(x-\xi)\|\}.$$

Taking the supremum over all $\xi \in \bar{B}_r(0)$ yields the claim. □

Theorem 4.2.2 (Uniform boundedness principle)**.** *Let X be a Banach space, Y a normed space, and $(T_i)_{i\in I}$ a family of bounded linear operators from X to Y such that $\sup_{i\in I}\|T_i x\| < \infty$ for all $x \in X$. Then $\sup_{i\in I}\|T_i\| < \infty$.*

Proof. Suppose that $\sup_{i\in I}\|T_i\| = \infty$. Let (i_n) be a sequence in I such that $\|T_{i_n}\| \geq 4^n$. Set $x_0 = 0$ and choose inductively $x_n \in X$ such that $\|x_n - x_{n-1}\| \leq 3^{-n}$ and $\|T_{i_n} x_n\| \geq \frac{2}{3}\cdot 3^{-n}\|T_{i_n}\|$ (this is possible due to the previous lemma).

Then (x_n) is a Cauchy sequence, hence it converges to some $x \in X$ (that's where the completeness of X is needed). Furthermore,

$$\|x - x_n\| = \lim_{m\to\infty}\|x_m - x_n\| \leq \lim_{m\to\infty}\sum_{k=n+1}^{m}\|x_k - x_{k-1}\| \leq \sum_{k=n+1}^{\infty} 3^{-k} \leq \frac{1}{2}3^{-n}.$$

Thus,

$$\|T_{i_n} x\| \geq \|T_{i_n} x_n\| - \|T_{i_n}(x - x_n)\| \geq \frac{2}{3}\cdot 3^{-n}\|T_{i_n}\| - \frac{1}{2}3^{-n}\|T_{i_n}\| \geq \frac{1}{6}\left(\frac{4}{3}\right)^n,$$

contradicting the assumption $\sup_{i\in I}\|T_i x\| < \infty$. □

Definition 4.2.3 (Open map)**.** A map between topological spaces is called *open* if the images of open sets are open.

Theorem 4.2.4 (Open mapping theorem)**.** *Let X, Y be Banach spaces. If the bounded linear operator $T\colon X \to Y$ is surjective, then it is open.*

Proof. Let $U \subset X$ be open. Since translations are homeomorphisms, we may assume that $0 \in U$. Then there is a ball B with center 0 such that $B \subset U$. Since dilations are homeomorphisms, we additionally assume that $B = B_1(0)$. It suffices to show that $T(B)$ contains a neighborhood of 0.

In the first step we show that $\overline{T(B)}$ contains a neighborhood of 0. For each $n \in \mathbb{N}$ define the norm $\|\cdot\|_n$ on Y by

$$\|y\|_n = \inf\{\|u\| + n\|v\| \colon u \in X, v \in Y, Tu + v = y\}.$$

Let Z be the set of all finitely supported sequences in Y with pointwise addition and scalar multiplication and the norm

$$\|\cdot\|_Z \colon Z \to [0,\infty),\ \|f\|_Z = \sup_n \|f(n)\|_n.$$

For $n \in \mathbb{N}$ let $S_n\colon Y \to Z$, $y \mapsto y\delta_n$. Note that $\|S_n y\|_Z = \|y\|_n$.

Taking $u = 0, v = y$ in the definition of $\|\cdot\|_n$ we get $\|y\|_n \leq n\|y\|$, hence S_n is bounded for all $n \in \mathbb{N}$. Taking $u \in T^{-1}(y)$, $v = 0$, we obtain $\|y\|_n \leq \|u\|$, thus $(S_n y)_n$ is bounded for all $y \in Y$. By the uniform boundedness theorem there is a constant $C > 0$ such that $\|S_n\| \le C$ for all $n \in \mathbb{N}$.

Now let $\delta = 1/C$. If $y \in B_\delta(0)$, then $\|y\|_n \le C\|y\| < 1$. Thus for every $n \in \mathbb{N}$ there exist $u_n \in X$, $v_n \in Y$ such that $Tu_n + v_n = y$ and $\|u_n\| + n\|v_n\| < 1$. In particular $u_n \in B_1(0)$ and $v_n \in B_{1/n}(0)$, hence $T(B) \ni Tu_n \to y$. Thus $y \in \overline{T(B)}$.

In the second step we show that $T(B)$ contains $B_{\delta/2}(0)$. If $\|y\| < \delta/2$, then by the first step and scaling there exists $x_1 \in B_{1/2}(0)$ such that $\|y - Tx_1\| < \delta/4$.

This way we get recursively a sequence (x_n) in X with $\|x_n\| < 2^{-n}$ and

$$\left\| y - \sum_{k=1}^{n} Tx_k \right\| < \delta 2^{-(n+1)}.$$

Hence $y = \sum_{k=1}^\infty Tx_k$. On the other hand, the norm estimate for x_n and completeness of X imply that $(\sum_{k=1}^n x_k)_n$ converges to some $x \in B_1(0)$. Thus $y = Tx \in T(B)$. □

Corollary 4.2.5 (Bounded inverse)**.** *Let X, Y be Banach spaces. If $T \in \mathcal{L}(X, Y)$ is bijective, then its inverse is bounded.*

Corollary 4.2.6. *Let X be a vector space and $\|\cdot\|_1$, $\|\cdot\|_2$ complete norms. If there exists a constant $C > 0$ such that $\|\cdot\|_1 \le C\|\cdot\|_2$, then $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent.*

Definition 4.2.7 (Closed operator)**.** Let X, Y be normed spaces and let $T\colon X \to Y$ be a linear operator. The *graph norm* $\|\cdot\|_T$ is defined as

$$\|\cdot\|_T\colon X \to [0, \infty),\ \|x\|_T = \|x\| + \|Tx\|.$$

The operator T is called *closed* if $(X, \|\cdot\|_T)$ is complete.

Proposition 4.2.8. *Let X be a normed space and Y a Banach space. A linear operator $T\colon X \to Y$ is closed if and only if whenever (x_n) is a Cauchy sequence in X and (Tx_n) is a Cauchy sequence in Y, then (x_n) converges and $T(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} Tx_n$.*

Example 4.2.9. Let Ω be a domain. The operator $\Delta\colon H^2(\Omega) \to L^2(\Omega)$ is closed.

Theorem 4.2.10 (Closed graph theorem)**.** *Let X be a normed space, Y a Banach space and $T\colon X \to Y$ be linear. Of the following three properties, every pair of two implies the third.*

(i) T is closed.

(ii) X is complete.

(iii) T is continuous.

Proof. The only nontrivial implication is (i)+(ii) $\implies$ (iii). By definition, T is continuous w.r.t. the graph norm on X. Moreover, $\|\cdot\|_X \leq \|\cdot\|_T$. Since X is complete with respect to both $\|\cdot\|_X$ and $\|\cdot\|_T$, the norms are equivalent. Thus T is continuous with respect to $\|\cdot\|_X$. $\square$

4.3 Spectrum

From now on, we assume that $\mathbb{K} = \mathbb{C}$, i.e., all Hilbert spaces are complex Hilbert spaces.

Definition 4.3.1. Let H, K be Hilbert spaces. A (possibly unbounded) *operator from H to K* is a linear map T defined on a linear subspace of H with values in K. If $H = K$, we also say that T is an operator *in H*. The domain of T is denoted by $\mathrm{dom}(T)$. The operator T is called *densely defined* if $\mathrm{dom}(T)$ is dense in H.

Remark 4.3.2. The domain of an operator is a crucial part of information. It often happens in that two operators act in the same way, but on different domains.

Definition 4.3.3 (Spectrum)**.** Let H be a Hilbert space and T a densely defined operator in H. For $z \in \mathbb{C}$, the operator $T - z$ is defined by $\mathrm{dom}(T - z) = \mathrm{dom}(T)$ and $(T - z)\xi = T\xi - z\xi$. The *resolvent set* $\rho(T)$ is defined as

$$\rho(T) = \{z \in \mathbb{C} \mid T - z \text{ bijective with bounded inverse}\}.$$

For $z \in \rho(T)$, the *resolvent* of T at z is the (bounded) operator $(T - z)^{-1}$.

The *spectrum* $\sigma(T)$ of T is the complement of $\rho(T)$.

Example 4.3.4. If $H = \mathbb{C}^n$, then $T - z$ is bijective if and only if it is injective if and only if it surjective. Thus $z \in \sigma(T)$ if and only if there exists $\xi \neq 0$ such that $T\xi = z\xi$. In other words, the spectrum of T is the set of eigenvalues of T.

In infinite dimensions, one can have spectral values that are not eigenvalues:

Example 4.3.5. Let

$$T\colon L^2([0,1]) \to L^2([0,1]),\ (Tf)(x) = xf(x).$$

For every $z \in [0,1]$, the operator $T - z$ is not surjective: If $Tf - zf = 1$, then $(x-z)f(x) = 1$ for a.e. $x \in [0,1]$, which implies $f(x) = (x-z)^{-1}$ a.e. However, $x \mapsto (x-z)^{-1}$ is not square integrable.

The operator $T - z$ is injective for every $z \in \mathbb{C}$: If $Tf - zf = 0$, then $xf(x) = zf(x)$ for a.e. $x \in [0,1]$, which implies $f = 0$ a.e. Thus T has no eigenvalues.

Example 4.3.6. Let $(X, \mathcal{A}, \mu)$ be a semi-finite measure space and $\varphi\colon X \to \mathbb{C}$ measurable. The operator M_φ of multiplication with φ on $L^2(X,\mu)$ is defined by

$$\begin{aligned}\operatorname{dom}(M_\varphi) &= \{f \in L^2(X,\mu) \mid \varphi f \in L^2(X,\mu)\},\\ M_\varphi f &= \varphi f.\end{aligned}$$

Let us first show that M_φ is densely defined. Let $A_n = \{x \in X : |\varphi(x)| \leq n\}$. Clearly $\mathbb{1}_{A_n} \to 1$ pointwise. By the dominated convergence theorem, $\|f - f\mathbb{1}_{A_n}\|_2 \to 0$ for every $f \in L^2(X,\mu)$. Moreover,

$$\int_X |\varphi f \mathbb{1}_{A_n}|^2\,d\mu \leq n^2 \int_X |f|^2\,d\mu < \infty.$$

Thus $f\mathbb{1}_{A_n} \in \operatorname{dom}(M_\varphi)$. Therefore M_φ is densely defined. The operator M_φ is also closed, but that requires some measure theory tools we have not covered in this course.

We claim that

$$\sigma(M_\varphi) = \{\lambda \in \mathbb{C} \mid \mu(\varphi^{-1}(B_\varepsilon(\lambda))) > 0 \text{ for all } \varepsilon > 0\}.$$

First, if there exists $\varepsilon > 0$ such that $\mu(\varphi^{-1}(B_\varepsilon(\lambda))) = 0$, let $R_\lambda = M_{(\varphi-\lambda)^{-1}}$. Note that the function $(\varphi - \lambda)^{-1}$ is finite μ-a.e., and functions that coincide μ-a.e. define the same multiplication operator.

If $x \notin \varphi^{-1}(B_\varepsilon(0))$, then $|\varphi(x) - \lambda| \geq \varepsilon$. As $\mu(\varphi^{-1}(B_\varepsilon(0)) = 0$, we conclude that $|\varphi - \lambda| \geq \varepsilon$ μ-a.e. and hence

$$\int_X |f(\varphi-\lambda)^{-1}|^2\,d\mu \leq \varepsilon^{-2} \int_X |f|^2\,d\mu.$$

Thus $\operatorname{dom}(R_\lambda) = L^2(X,\mu)$ and R_λ is bounded. The identities $R_\lambda(M_\varphi - \lambda)f = f$ for $f \in \operatorname{dom}(M_\varphi)$ and $(M_\varphi - \lambda)R_\lambda f = f$ for $f \in L^2(X,\mu)$ are clear. Therefore $\lambda \in \rho(M_\varphi)$.

For the converse inclusion, let $\lambda \in \mathbb{C}$ with $\mu(\varphi^{-1}(B_\varepsilon(\lambda))) > 0$ for all $\varepsilon > 0$. Since $(X, \mathcal{A}, \mu)$ is assumed to be semi-finite, there exists for every $\varepsilon > 0$ a set $A_\varepsilon \in \mathcal{A}$ such that $A_\varepsilon \subset \varphi^{-1}(B_\varepsilon(\lambda))$ and $0 < \mu(A_\varepsilon) < \infty$. Let $f_\varepsilon = \mathbb{1}_{A_\varepsilon}/\mu(A_\varepsilon)^{1/2}$. We have $\|f_\varepsilon\|_2 = 1$ and

$$|(M_\varphi - \lambda) f_\varepsilon| < \varepsilon |f_\varepsilon|.$$

Hence $\|(M_\varphi - \lambda) f_\varepsilon\|_2 \leq \varepsilon$, which implies that $M_\varphi - \lambda$ cannot have a bounded inverse. Therefore $\lambda \in \sigma(M_\varphi)$.

Lemma 4.3.7. *Let H be a Hilbert space and T a densely defined operator in H. If T is not closed, then $\rho(T) = \emptyset$.*

Proposition 4.3.8. *Let H be a Hilbert space. For densely defined operator T in H the resolvent set $\rho(T)$ is an open subset of $\mathbb{C}$ and the map*

$$\rho(T) \to \mathbb{C},\ z \mapsto \langle \xi, (T - z)^{-1} \eta \rangle$$

is differentiable for all $\xi, \eta \in H$.

Remark 4.3.9. Since we are dealing with a function on a complex domain, differentiability is much stronger than for functions with real domain. For example, if $U \subset \mathbb{C}$ is open and $f \colon U \to \mathbb{C}$ is differentiable, then for every $w \in U$ and every $r > 0$ such that $B_r(w) \subset U$ there exists a sequence (a_n) in $\mathbb{C}$ such that $\sum_{n=0}^\infty a_n (z - w)^n$ converges absolutely on $B_r(w)$ to $f(z)$. In particular, such a function is necessarily smooth.

Proof. Let $z_0 \in \rho(T)$. If $z \in \mathbb{C}$ with $|z - z_0| < \|(T - z_0)^{-1}\|^{-1}$, let

$$S_n = \sum_{k=0}^{n} (z - z_0)^k (T - z_0)^{-(k+1)}.$$

We want to show that (S_n) is a Cauchy sequence. If $m > n$, then

$$\begin{aligned} \|S_m - S_n\| &\leq \sum_{k=n+1}^{m} |z - z_0|^k \|(T - z_0)^{-1}\|^{k+1} \\ &\leq \|(T - z_0)^{-1}\| \sum_{k=n+1}^{\infty} \left(|z - z_0| \|(T - z_0)^{-1}\|\right)^k. \end{aligned}$$

Since $|z - z_0| < \|(T - z_0)^{-1}\|^{-1}$, the series $\sum_{k=0}^\infty \left(|z - z_0| \|(T - z_0)^{-1}\|\right)^k$ converges. In particular, for every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $\sum_{k=N}^\infty \left(|z - z_0| \|(T - z_0)^{-1}\|\right)^k < \varepsilon$. Thus (S_n) is a Cauchy sequences.

As $\mathcal{L}(H)$ is complete, there exists $S \in \mathcal{L}(H)$ such that $S_n \to S$. We claim that S is an inverse of $T - z$. Indeed, if $\xi \in H$, then $S_n\xi \in \operatorname{dom}(T)$ and

$$
\begin{aligned}
(T - z)S_n\xi &= \sum_{k=0}^{n}(z - z_0)^k(T - z)(T - z_0)^{-(k+1)}\xi \\
&= \sum_{k=0}^{n}(z - z_0)^k(T - z_0 + z_0 - z)(T - z_0)^{-(k+1)}\xi \\
&= \sum_{k=0}^{n}\Big((z - z_0)^k(T - z_0)^{-k} - (z - z_0)^{k+1}(T - z_0)^{-(k+1)}\Big)\xi \\
&= \xi - (z - z_0)^{n+1}(T - z_0)^{-(n+1)}\xi.
\end{aligned}
$$

Since $S_n\xi \to S\xi$ and T is closed, we conclude $S\xi \in \operatorname{dom}(T)$ and

$$(T - z)S\xi = \lim_{n\to\infty}(\xi - (z - z_0)^{n+1}(T - z_0)^{-(n+1)}\xi) = \xi.$$

A similar argument shows that $S(T - z)\xi = \xi$ for every $\xi \in \operatorname{dom}(T)$. Thus $z \in \rho(T)$ and $(T - z)^{-1} = S$. Furthermore, if $\xi, \eta \in H$, then

$$\langle \xi, (T - z)^{-1}\eta\rangle = \lim_{n\to\infty}\langle \xi, S_n\eta\rangle = \sum_{k=0}^{\infty}\langle \xi, (T - z_0)^{-(k+1)}\eta\rangle(z - z_0)^k,$$

which depends smoothly on z. □

Remark 4.3.10. Here is a brief summary of the proof: For $R \in \mathcal{L}(H)$ with $\|R\| < 1$, the series

$$\sum_{k=0}^{\infty} R^k$$

converges in operator norm. This is called the *Neumann series* (named after Carl Neumann, *not* John von Neumann). By a telescoping trick one can show that

$$(1 - R)\sum_{k=0}^{\infty} R^k = \sum_{k=0}^{\infty} R^k(1 - R) = 1.$$

Thus $1 - R$ is invertible with inverse $\sum_{k=0}^{\infty} R^k$. What we used in the proof is that one can write $T - z$ as

$$T - z_0 + z_0 - z = (1 - (z - z_0)(T - z_0)^{-1})(T - z_0).$$

If $|z - z_0| < \|(T - z_0)^{-1}\|^{-1}$, one can then apply the Neumann series to $R = (z - z_0)(T - z_0)^{-1}$ to find an inverse of $T - z$.

4.4 Symmetric and self-adjoint operators

Definition 4.4.1 (Adjoint of an operator)**.** Let H, K be Hilbert spaces and T a densely defined operator from H to K. The adjoint T^* of T is the (possibly unbounded) operator from K to H defined by

$$\begin{aligned} \operatorname{dom}(T^*) &= \{\xi \in K \mid \exists \eta \in H \,\forall \zeta \in \operatorname{dom}(T) \colon \langle \xi, T\zeta \rangle = \langle \eta, \zeta \rangle\}, \\ T^*\xi &= \eta. \end{aligned}$$

An operator T in H is called *self-adjoint* if $T^* = T$.

Remark 4.4.2. The density of $\operatorname{dom}(T)$ in H guarantees that the element η in the definition of the adjoint is uniquely determined, if it exists.

Lemma 4.4.3. *The adjoint of a densely defined operator between Hilbert spaces is closed.*

Proof. Let T be a densely defined operator from H to K, (ξ_n) a sequence in $\operatorname{dom}(T^*)$ such that $\xi_n \to \xi$ in K and $T^*\xi_n \to \eta$ in H. To prove that T^* is close, we have to show that $\xi \in \operatorname{dom}(T^*)$ and $T^*\xi = \eta$.

Let $\zeta \in \operatorname{dom}(T)$. By definition of the adjoint,

$$\langle \xi, T\zeta \rangle = \lim_{n\to\infty} \langle \xi_n, T\zeta \rangle = \lim_{n\to\infty} \langle T^*\xi_n, \zeta \rangle = \langle \eta, \zeta \rangle.$$

Hence $\xi \in \operatorname{dom}(T^*)$ and $T^*\xi = \eta$. □

Proposition 4.4.4. *Let H, K be Hilbert spaces. A densely defined operator T from H to K has an everywhere defined adjoint if and only if T is continuous.*

Proof. First let T be a densely defined continuous operator from H to K. If $\xi \in K$, then

$$H \to \mathbb{C},\ \zeta \mapsto \langle \xi, T\zeta \rangle$$

is a bounded linear functional. By the Riesz representation theorem, there exists $\eta \in H$ such that $\langle \xi, T\zeta \rangle = \langle \eta, \zeta \rangle$ for all $\zeta \in H$. Thus $\operatorname{dom}(T^*) = H$.

Assume conversely that T has an everywhere defined adjoint. Since the adjoint is closed by the previous lemma, it is also bounded by the closed graph theorem. We claim that $T\xi = T^{**}\xi$ for all $\xi \in \operatorname{dom}(T)$. In fact, T^{**} is everywhere defined and bounded and

$$\langle \xi, T^*\eta \rangle = \langle T\xi, \eta \rangle$$

for all $\xi \in \operatorname{dom}(T)$, $\eta \in \operatorname{dom}(T^*)$, hence $\xi \in \operatorname{dom}(T^{**})$ and $T^{**}\xi = T\xi$. Since T^{**} is continuous, so is T. □

Definition 4.4.5 (Extension of operators, symmetric operator)**.** Let H, K be Hilbert spaces and S, T (possibly unbounded) operators from H to K. The operator T is called an *extension* of S, written as $S \subset T$, if $\operatorname{dom}(S) \subset \operatorname{dom}(T)$ and $T\xi = S\xi$ for $\xi \in \operatorname{dom}(S)$.

A densely defined operator T in H is called *symmetric* if $T \subset T^*$.

Remark 4.4.6. Clearly, every self-adjoint operator is symmetric. The converse is not true, as we shall see in the examples.

Example 4.4.7. A bounded everywhere defined operator is symmetric if and only if it self-adjoint since an everywhere defined operator has no non-trivial extensions. In this case, it suffices to check

$$\langle T\xi, \eta\rangle = \langle \xi, T\eta\rangle$$

for all $\xi, \eta \in H$.

Example 4.4.8. Let $(X, \mathcal{A}, \mu)$ be a semi-finite measure space and $\varphi\colon X \to \mathbb{R}$ measurable. If

$$D \subset \{f \in L^2(X, \mu) \mid \varphi f \in L^2(X, \mu)\}$$

is a dense subspace, then $M_\varphi|_D$ is symmetric. However, it is self-adjoint if and only if $D = \operatorname{dom}(M_\varphi)$, the maximal domain:

The symmetry is not hard to see: If $f, g \in D$, then

$$\langle M_\varphi|_D f, g\rangle = \int_X \overline{\varphi f} g \, d\mu = \int_X \bar{f}(\varphi g)\, d\mu = \langle f, M_\varphi|_D g\rangle.$$

In fact, the same computation shows that if $f \in D$ and $g \in \operatorname{dom}(M_\varphi)$, then

$$\langle M_\varphi|_D f, g\rangle = \langle f, M_\varphi g\rangle.$$

Therefore $\operatorname{dom}(M_\varphi) \subset \operatorname{dom}(M_\varphi|_D^*)$ and $M_\varphi f = M_\varphi|_D^* f$ for $f \in \operatorname{dom}(M_\varphi|_D^*)$, which can be summarized as $M_\varphi \subset M_\varphi|_D^*$. Thus a necessary condition for self-adjointness is $D = \operatorname{dom}(M_\varphi)$.

Let us show that it is also sufficient. Let $f \in \operatorname{dom}(M_\varphi^*)$ and $A_n = \{x \in X : |\varphi(x)| \leq n\}$, $g_n = \mathbb{1}_{A_n}\varphi f$. As $|\varphi g_n| \leq n^2|f|$, we have $g_n \in \operatorname{dom}(M_\varphi)$. Therefore

$$\|g_n\|_2^2 = \left|\int_X \bar{g}_n \varphi f \, d\mu\right| = |\langle M_\varphi g_n, f\rangle| = |\langle g_n, M_\varphi^* f\rangle| \leq \|g_n\|_2 \|M_\varphi^* f\|_2.$$

It follows that $\|g_n\|_2 \leq \|M_\varphi^* f\|_2$. By the monotone convergence theorem,

$$\|g_n\|_2^2 \to \int_X |\varphi f|^2 \, d\mu.$$

We conclude that $\varphi f \in L^2(X, \mu)$, which implies $f \in \operatorname{dom}(M_\varphi)$. As we have already shown that $M_\varphi \subset M_\varphi^*$, we arrive at $M_\varphi^* = M_\varphi$.

Remark 4.4.9. The previous example my suggest that symmetric operators differ from self-adjoint operators only in that one has not chosen the maximal domain. It is true that self-adjoint operators do not have a non-trivial symmetric extension. However, there are symmetric operators with several self-adjoint extensions and symmetric operators with no self-adjoint extension, as the next examples show.

Example 4.4.10. The operator

$$T\colon C_c^2((0,\pi)) \to L^2((0,\pi)),\ f \mapsto f''$$

is a symmetric operator in $L^2((0,\pi))$, as integration by parts shows. However, it is not self-adjoint and has several self-adjoint extensions

Two of them are given as follows: Let $a_k = \int_0^\pi \cos^2(kx)\,dx$ for $k \in \mathbb{N}_0$ and $b_k = \int_0^\pi \sin^2(k\pi)\,dx$ for $k \in \mathbb{N}$. Note that $(a_k^{-1/2}\cos(k\cdot))_{k\in\mathbb{N}_0}$ and $(b_k^{-1/2}\sin(k\cdot))_{k\in\mathbb{N}}$ are orthonormal bases of $L^2((0,\pi))$.

The *Laplacian with Dirichlet boundary conditions* on $L^2((0,\pi))$ is defined by

$$\begin{aligned}
\operatorname{dom}(\Delta^{(D)}) &= \left\{ f \in L^2((0,\pi)) : \sum_{k=1}^\infty k^4 b_k^{-1} |\langle f, \sin(k\cdot)\rangle|^2 < \infty \right\},\\
\Delta^{(D)} f &= -\sum_{k=1}^\infty k^2 b_k^{-1} \langle f, \sin(k\cdot)\rangle \sin(k\cdot).
\end{aligned}$$

The *Laplacian with Neumann boundary conditions* on $L^2((0,\pi))$ is defined by

$$\begin{aligned}
\operatorname{dom}(\Delta^{(N)}) &= \left\{ f \in L^2((0,\pi)) : \sum_{k=1}^\infty k^4 a_k^{-1} |\langle f, \cos(k\cdot)\rangle|^2 < \infty \right\},\\
\Delta^{(N)} f &= -\sum_{k=0}^\infty k^2 a_k^{-1} \langle f, \cos(k\cdot)\rangle \cos(k\cdot).
\end{aligned}$$

We will see later that both $\Delta^{(D)}$ and $\Delta^{(N)}$ are self-adjoint. Moreover, if $f = \sum_{k=1}^N b_k^{-1}\langle f, \sin(k\cdot)\rangle \sin(k\cdot)$, then

$$\begin{aligned}
f''(x) &= \sum_{k=1}^N b_k^{-1} \langle f, \sin(k\cdot)\rangle \frac{d^2}{dx^2}\sin(kx)\\
&= -\sum_{k=1}^N k^2 b_k^{-1} \langle f, \sin(k\cdot)\rangle \sin(kx)\\
&= \Delta^{(D)} f(x).
\end{aligned}$$

An analogous result holds for $\Delta^{(N)}$ and finite linear combinations of cosine functions. With a bit more work, one can show that $C_c^2((0,\pi)) \subset \operatorname{dom}(\Delta^{(D)}) \cap \operatorname{dom}(\Delta^{(N)})$ and $f'' = \Delta^{(D)} f = \Delta^{(N)} f$ for $f \in C_c^2((0,\pi))$.

Example 4.4.11. The operator

$$T\colon C_c^1((0,\infty)) \to L^2((0,\infty)),\ f \mapsto if'$$

is a densely defined symmetric operator in $L^2((0,\infty))$ without self-adjoint extensions.

Lemma 4.4.12. *Let H be a Hilbert space. If T is a symmetric operator in H, then $\ker(T-\lambda) = \{0\}$ for $\lambda \in \mathbb{C}\setminus\mathbb{R}$.*

Proof. If $\xi \in \ker(T-\lambda)$, then

$$\lambda\langle \xi, \xi\rangle = \langle \xi, T\xi\rangle = \langle T\xi, \xi\rangle = \bar{\lambda}\langle \xi, \xi\rangle.$$

If $\lambda \neq \bar{\lambda}$, we conclude $\xi = 0$. □

In other words, symmetric operators have only real eigenvalues. Note however, that an operator on an infinite-dimensional Hilbert space can have spectral values that are not eigenvalues. This is in fact always the case for symmetric operators that are not self-adjoint, as the next result shows. This is one of the reasons why one requires the observables in quantum mechanics to be self-adjoint and not only symmetric.

Proposition 4.4.13. *Let H be a Hilbert space and T a densely defined symmetric operator in H. The spectrum of T is either $\mathbb{C}$, $\{\lambda \in \mathbb{C} \mid \operatorname{Im}\lambda \geq 0\}$, $\{\lambda \in \mathbb{C} \mid \operatorname{Im}\lambda \leq 0\}$ or a subset of $\mathbb{R}$, and T is self-adjoint if and only if $\sigma(T) \subset \mathbb{R}$.*

Corollary 4.4.14. *Let H be a Hilbert space. A densely defined closed symmetric operator T in H is self-adjoint if and only if $\operatorname{ran}(T \pm i) = H$.*

Corollary 4.4.15. *Let H be a Hilbert space. If T is a densely defined symmetric operator in H such that $\rho(T) \cap \mathbb{R} \neq \emptyset$, then T is self-adjoint.*

4.5 The spectral theorem

The spectral theorem links self-adjoint operators (observables in quantum mechanics) to measurements. It takes a particularly simple form in finite dimensions.

Proposition 4.5.1 (Spectral theorem in finite dimensions)**.** *Let H be finite-dimensional Hilbert space and $T\colon H \to H$ a symmetric operator. For $\lambda \in \sigma(T)$ let P_λ be the orthogonal projection on* $\ker(T - \lambda)$*. Then*

$$T = \sum_{\lambda \in \sigma(T)} \lambda P_\lambda.$$

Proof. Let $\lambda \in \sigma(T)$. For $\xi \in \ker(T - \lambda)$ we have $T\xi = \lambda\xi$. Moreover, if $\eta \in \ker(T - \lambda)^\perp$, then

$$\langle T\eta, \xi\rangle = \langle \eta, T\xi\rangle = \lambda\langle \eta, \xi\rangle = 0,$$

hence $T\eta \in \ker(T - \lambda)^\perp$. Thus $(T - \lambda P_\lambda)(\ker(T - \lambda)^\perp) \subset \ker(T - \lambda)^\perp$. Hence we can apply the previous step to $(T - \lambda P_\lambda)|_{\ker(T-\lambda)^\perp}$. If we iterate this procedure, we end up with the claimed identity (this iteration terminates because we are in finite dimensions). $\square$

In infinite dimensions, not all spectral values are eigenvalues and in general, one cannot expect a self-adjoint operator to be a linear combination of projections. To deal with cases as in Example 4.3.5 where the spectrum is continuous and $\ker(T - \lambda) = \{0\}$ for all $\lambda \in \mathbb{C}$, we need the following concept that generalizes families of orthogonal projections.

Definition 4.5.2 (Projection-valued measure)**.** Let $(X, \mathcal{A})$ be a measurable space and H a Hilbert space. A map $E\colon \mathcal{A} \to \mathcal{L}(H)$ is called *projection-valued measure (PVM)* if

- $E(A)$ is a projection for all $A \in \mathcal{A}$,
- $E(\emptyset) = 0$, $E(X) = 1_H$,
- $\mathcal{A} \to [0, \infty)$, $A \mapsto \langle \xi, E(A)\xi\rangle$ is a measure for all $\xi \in H$.

Remark 4.5.3. By definition of a projection-valued measure, the map $\mu_\xi\colon A \mapsto \langle \xi, E(A)\xi\rangle$ is a measure. We write $\int_\Omega g(\omega)\, d\langle \xi, E(\omega)\xi\rangle$ for $\int_\Omega g\, d\mu_\xi$.

Example 4.5.4. Let $X = \{1, \dots, n\}$ and $\mathcal{A} = \mathcal{P}(X)$. If $P_1, \dots, P_n \in \mathcal{L}(H)$ are projections such that $P_1 + \dots + P_n = 1$, then

$$E\colon \mathcal{A} \to \mathcal{L}(H),\ E(A) = \sum_{k \in A} P_k$$

is a projection-valued measure.

Example 4.5.5. If $H = L^2(\mathbb{R})$, then the map

$$E\colon \mathcal{A} \to \mathcal{L}(H),\ A \mapsto M_{\mathbb{1}_A}$$

is a projection-valued measure on $(X, \mathcal{A})$.

The following properties of projection-valued measures come in handy.

Lemma 4.5.6. *Let $(X, \mathcal{A})$ be a measurable space, H a Hilbert space and E a projection-valued measure on $(X, \mathcal{A})$ with values in $\mathcal{L}(H)$.*

(a) If $A, B \in \mathcal{A}$, then $E(A)E(B) = E(A \cap B)$. In particular, if $A \cap B = \emptyset$, then $E(A)E(B) = 0$.

(b) If (A_n) is an increasing sequence in $\mathcal{A}$ and $\xi \in H$, then $E(A_n)\xi \to E(\bigcup_{n\in\mathbb{N}} A_n)\xi$.

Lemma 4.5.7. *Let H be a Hilbert space and T an operator in H. If $\langle \xi, T\xi \rangle = 0$ for all $\xi \in \operatorname{dom}(T)$, then $T\xi = 0$ for all $\xi \in \operatorname{dom}(T)$.*

Proposition 4.5.8. *Let H be a Hilbert space, $(\Omega, \mathcal{A})$ a measurable space and $E\colon \mathcal{A} \to \mathcal{L}(H)$ a projection-valued measure. For every measurable function $f\colon \Omega \to \mathbb{C}$ there exists a unique operator T in H with domain*

$$\operatorname{dom}(T) = \left\{ \xi \in H : \int_\Omega |f(\omega)|^2 \, d\langle \xi, E(\omega)\xi \rangle < \infty \right\}$$

that satisfies

$$\langle \xi, T\xi \rangle = \int_\Omega f(\omega) \, d\langle \xi, E(\omega)\xi \rangle$$

for all $\xi \in \operatorname{dom}(T)$.

The operator T is densely defined and it is bounded if f is bounded

Proof. For $\xi \in H$ let μ_ξ denote the measure $A \mapsto \langle \xi, E(A)\xi \rangle$. The domain of T can be rephrased as

$$\operatorname{dom}(T) = \{ \xi \in H \mid f \in L^2(\Omega, \mu_\xi) \}.$$

First note that since $|f(\omega)| \leq 1 + |f(\omega)|^2$ for all $\omega \in \Omega$, we have

$$\int_\Omega |f(\omega)| \, d\mu_\xi(\omega) \leq \mu_\xi(A) + \int_\Omega |f(\omega)|^2 \, d\mu_\xi(\omega).$$

Hence the integral $\int_\Omega f(\omega) \, d\mu_\xi(\omega)$ is well-defined whenever $f \in L^2(\Omega, \mu_\xi)$. By the previous lemma, T is uniquely determined by $\langle \xi, T\xi \rangle$ for $\xi \in \operatorname{dom}(T)$.

To see that $\mathrm{dom}(T)$ is dense, let $A_n = \{\omega \in \Omega : |f(\omega)| \leq n\}$. By a previous lemma, $E(A_n)\xi \to \xi$ for all $\xi \in H$. Moreover, $\mu_{E(A_n)\xi}(\Omega \setminus A_n) = 0$. Thus

$$\begin{aligned}\int_\Omega |f(\omega)|^2 \, d\mu_{E(A_n)\xi}(\omega) &= \int_\Omega \mathbb{1}_{A_n}(\omega)|f(\omega)|^2 \, d\mu_{E(A_n)\xi}(\omega) \\ &\leq n^2 \|E(A_n)\xi\|^2 \\ &< \infty,\end{aligned}$$

which implies $E(A_n)\xi \in \mathrm{dom}(T)$. Therefore $\mathrm{dom}(T)$ is dense.

To show existence of T, let

$$Q_f \colon \mathrm{dom}(T) \times \mathrm{dom}(T) \to \mathbb{C},\ (\xi, \eta) \mapsto \sum_{k=0}^{3} i^{-k} \int_\Omega f \, d\mu_{\xi + i^k \eta}.$$

Note that by the polarization identity, $Q_f(\xi, \xi) = \int_\Omega f \, d\mu_\xi$.

If $f = \sum_{j=1}^n \alpha_j \mathbb{1}_{A_j}$ with disjoint measurable sets A_j, then the polarization identity implies

$$\begin{aligned}|Q_f(\xi, \eta)| &= \left| \sum_{j=1}^n \sum_{k=0}^3 i^{-k} \alpha_j \langle \xi + i^k \eta, E(A_j)(\xi + i^k \eta) \rangle \right| \\ &= \left| \sum_{j=1}^n \alpha_j \langle \xi, E(A_j)\eta \rangle \right| \\ &\leq \left(\sum_{j=1}^n \langle \xi, E(A_j)\xi \rangle \right)^{1/2} \left(\sum_{j=1}^n |\alpha_j|^2 \langle \eta, E(A_j)\eta \rangle \right)^{1/2} \\ &\leq \|\xi\| \left(\int_\Omega |f|^2 \, d\mu_\eta \right)^{1/2}.\end{aligned}$$

For arbitrary measurable $f \colon \Omega \to \mathbb{R}$, the inequality $|Q_f(\xi, \eta)| \leq \|\xi\| \|f\|_{L^2(\Omega, \mu_\eta)}$ can be proven by approximation.

It follows from the Riesz representation theorem that for every $\eta \in \mathrm{dom}(T)$ there exists a unique $T\eta \in H$ such that $\langle \xi, T\eta \rangle = Q_f(\xi, \eta)$ for all $\xi \in \mathrm{dom}(T)$, and $\|T\eta\| \leq \|f\|_{L^2(\Omega, \mu_\eta)}$. In particular, $\langle \xi, T\xi \rangle = \int_\Omega f \, d\mu_\xi$ as observed above.

If f is bounded, then

$$\int_\Omega |f(\omega)|^2 \, d\mu_\xi(\omega) \leq \|f\|_\infty^2 \|\xi\|^2 < \infty$$

for all $\xi \in H$. Thus $\mathrm{dom}(T) = H$ and $\|T\xi\| \leq \|f\|_{L^2(\Omega, \mu_\xi)} \leq \|f\|_\infty \|\xi\|$. Therefore T is bounded. $\square$

Definition 4.5.9 (Integral with respect to a PVM)**.** Let H be a Hilbert space, $(\Omega, \mathcal{A})$ a measurable space, $E\colon \mathcal{A} \to \mathcal{L}(H)$ a projection-valued measure and $f\colon \Omega \to \mathbb{C}$ a measurable function. The unique self-adjoint operator T from the previous result is called the integral of f with respect to E and denoted by $\int f\,dE$.

Proposition 4.5.10. *Let H be a Hilbert space, $(\Omega, \mathcal{A})$ a measurable space, $E\colon \mathcal{A} \to \mathcal{L}(H)$ a projection-valued measure and $f, g\colon \Omega \to \mathbb{C}$ measurable.*

(a) If (f_n) is an increasing sequence of non-negative bounded measurable functions on Ω and $f\colon \Omega \to \mathbb{R}$ is a bounded measurable function such that $f(\omega) = \lim_{n\to\infty} f_n(\omega)$ for all $\omega \in \Omega$, then

$$\left\langle \xi, \int_\Omega f_n\,dE\,\eta \right\rangle \to \left\langle \xi, \int_\Omega f\,dE\,\eta \right\rangle$$

for all $\xi, \eta \in H$.

(b) If f, g are bounded and $\alpha, \beta \in \mathbb{C}$, then

$$\int_\Omega (\alpha f + \beta g)\,dE = \alpha \int_\Omega f\,dE + \beta \int_\Omega g\,dE$$
$$\int_\Omega fg\,dE = \left(\int_\Omega f\,dE\right)\left(\int_\Omega g\,dE\right).$$

(c) $(\int_\Omega f\,dE)^* = \int_\Omega \bar{f}\,dE$. In particular, $\int_\Omega f\,dE$ is self-adjoint if f is real-valued.

(d) $\int_\Omega f\,dE$ is closed.

Proof. (a) By the polarization identity, it suffices to consider the case $\xi = \eta$. In this case, the statement is a direct consequence of the definition of the integral together with the monotone convergence theorem.

(b) Linearity is an easy consequence of the linearity of the integral of scalar-valued functions. Multiplicativity is a bit trickier. First note that if $A \in \mathcal{A}$, then

$$\int_\Omega \mathbb{1}_A(\omega)\,d\langle \xi, E(\omega)\xi\rangle = \langle \xi, E(A)\xi\rangle$$

for all $\xi \in H$, hence $\int_\Omega \mathbb{1}_A\,dE = E(A)$ by uniqueness. Therefore, if $A, B \in \mathcal{A}$, then

$$\begin{aligned}\int_\Omega \mathbb{1}_A \mathbb{1}_B \, dE &= \int_\Omega \mathbb{1}_{A\cap B} \, dE \\ &= E(A \cap B) \\ &= E(A)E(B) \\ &= \left(\int_\Omega \mathbb{1}_A \, dE\right) \left(\int_\Omega \mathbb{1}_B \, dE\right).\end{aligned}$$

By linearity, we obtain $\int_\Omega fg \, dE = (\int_\Omega f \, dE)(\int_\Omega g \, dE)$ whenever f and g are linear combinations of indicator functions of measurable sets. With the help of (a), one can then extend this identity to arbitrary bounded measurable functions. We leave the details as an exercise.

(c) Let $T_f = \int_\Omega f \, dE$. By the polarization identity, if $\xi, \eta \in \operatorname{dom}(T_f)$, then

$$\begin{aligned}\langle \xi, T_f \eta \rangle &= \frac{1}{4} \sum_{k=1}^4 i^{-k} \langle \xi + i^k \eta, T_f(\xi + i^k \eta) \rangle \\ &= \frac{1}{4} \sum_{k=1}^4 i^{-k} \int_\Omega f \, d\mu_{\xi + i^k \eta}(\omega) \\ &= \frac{1}{4} \sum_{k=1}^4 \overline{i^k \int_\Omega \bar{f} \, d\mu_{\xi + i^k \eta}(\omega)} \\ &= \frac{1}{4} \sum_{k=1}^4 \overline{i^k \langle \xi + i^k \eta, T_{\bar{f}}(\xi + i^k \eta) \rangle} \\ &= \frac{1}{4} \sum_{k=1}^4 i^{-k} \langle T_{\bar{f}}(\xi + i^k \eta), \xi + i^k \eta \rangle \\ &= \langle T_{\bar{f}} \xi, \eta \rangle.\end{aligned}$$

Thus $T_{\bar{f}} \subset T_f^*$. In particular, $T_f^* = T_{\bar{f}}$ if f is bounded.

To prove $T_f^* \subset T_{\bar{f}}$, let $A_n = \{\omega \in \Omega : |f(\omega)| \leq n\}$. First note that

$$\mu_{E(A_n)\xi}(B) = \langle E(A_n)\xi, E(B)E(A_n)\xi \rangle = \langle \xi, E(B \cap A_n)\xi \rangle = \mu_\xi(B \cap A_n)$$

for all $B \in \mathcal{A}$. Approximation by simple functions then shows that

$$\int_\Omega |f|^2 \, d\mu_{E(A_n)\xi} = \int_{A_n} |f|^2 \, d\mu_\xi \leq n^2 \|\xi\|^2$$

and

$$\langle E(A_n)\xi, T_f E(A_n)\xi \rangle = \int_\Omega f \, d\mu_{E(A_n)\xi} = \int_{A_n} f \, d\mu(\xi) = \langle \xi, T_{f \mathbb{1}_{A_n}} \xi \rangle$$

for all $\xi \in H$. In particular, $E(A_n)H \subset \mathrm{dom}(T_f)$ and $T_f E(A_n)\xi = T_{f\mathbb{1}_{A_n}}\xi$ for all $\xi \in H$.

If $\xi \in \mathrm{dom}(T_f^*)$ and $\eta \in H$, then the result in the bounded case implies

$$\langle T_{\bar{f}\mathbb{1}_{A_n}}\xi, \eta\rangle = \langle \xi, T_f E(A_n)\eta\rangle = \langle E(A_n)T_f^*\xi, \eta\rangle.$$

Hence $T_{\bar{f}\mathbb{1}_{A_n}}\xi = E(A_n)T_f^*\xi$. By the monotone convergence theorem,

$$\int_\Omega |f|^2\,d\mu_\xi = \lim_{n\to\infty}\int_{A_n}|f|^2\,d\mu = \lim_{n\to\infty}\|T_{f\mathbb{1}_{A_n}}\xi\|^2 = \lim_{n\to\infty}\|E(A_n)T_f^*\xi\|^2 \leq \|T_f^*\xi\|^2.$$

Therefore, $\xi \in \mathrm{dom}(T_{\bar{f}})$.

(d) Since $T_f = T_{\bar{f}}^*$, the operator T_f is closed. □

Remark 4.5.11. Similar to the case of measures, if E is a projection-valued measure, one can define an equivalence relation $\sim_E$ on the space of all measurable functions from Ω to $\mathbb{R}$ by setting $f \sim_E g$ if there exists $N \in \mathcal{A}$ with $E(N) = 0$ such that $\{\omega \in \Omega \mid f(\omega) = g(\omega)\} \subset N$. Let $L^\infty(\Omega, E)$ be the set of all equivalence classes of bounded measurable functions from Ω to $\mathbb{C}$.

The space $L^\infty(\Omega, E)$ has the structure of a *von Neumann algebra*, and the previous result says that the integral with respect to E is a (normal unital) $*$-homomorphism between the von Neumann algebras $L^\infty(\Omega, E)$ and $\mathcal{L}(H)$.

An important application concerns operators that admit an orthonormal basis consisting of eigenvectors:

Example 4.5.12. Let H be a separable Hilbert space with orthonormal basis $(\xi_n)_{n\in\mathbb{N}}$ and $(\lambda_n)_{n\in\mathbb{N}}$ a sequence in $\mathbb{R}$. The map

$$E\colon \mathcal{B}(\mathbb{R}) \to \mathcal{L}(H),\ A \mapsto \sum_{n\colon \lambda_n\in A}\langle \xi_n, \cdot\rangle\xi_n$$

is a spectral measure and the operator given by

$$\mathrm{dom}(T) = \left\{\xi \in H : \sum_{n=1}^\infty \lambda_n^2|\langle \xi_n, \xi\rangle|^2\right\}$$
$$T\xi = \sum_{n=1}^\infty \lambda_n\langle \xi_n, \xi\rangle\xi_n$$

is self-adjoint.

In particular, the Laplacian with Dirichlet and Neumann boundary conditions in $L^2((0,\pi))$ are self-adjoint.

Theorem 4.5.13 (Spectral theorem)**.** *Let H be a Hilbert space. For every self-adjoint operator T in H the there exists a unique projection-valued measure E on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with values in $\mathcal{L}(H)$ such that*

$$T = \int_{\mathbb{R}} \lambda \, dE(\lambda).$$

Theorem 4.5.14 (Spectral theorem in multiplication operator form)**.** *Let H be a Hilbert space. For every self-adjoint operator T in H there exists a localizable measure space $(X, \mathcal{A}, \mu)$, a measurable function $\varphi\colon X \to \mathbb{R}$ and a unitary operator $U\colon H \to L^2(X, \mu)$ such that*

$$T = U^* M_\varphi U.$$

Definition 4.5.15 (Spectral measure, functional calculus)**.** Let H be a Hilbert space. If T is a self-adjoint operator in H, then the unique PVM E on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with values in $\mathcal{L}(H)$ such that

$$T = \int_{\mathbb{R}} \lambda \, dE(\lambda)$$

is called the *spectral measure* of E.

If $f\colon \mathbb{R} \to \mathbb{C}$ is a measurable function, we define

$$f(T) = \int_{\mathbb{R}} f(\lambda) \, dE(\lambda).$$

We will not give a full proof of either version of the spectral theorem here. However, let us at least see how the spectral measure of intervals can be described. To do so, we need to properly define powers of a (possibly unbounded) operator.

Definition 4.5.16 (Powers of an operator)**.** Let H be a Hilbert space and T an operator in H. The powers T^n, $n \in \mathbb{N}$, are inductively defined by $T^1 = T$ and

$$\begin{aligned} \operatorname{dom}(T^{n+1}) &= \{\xi \in \operatorname{dom}(T^n) \mid T^n\xi \in \operatorname{dom}(T)\}, \\ T^{n+1}\xi &= T(T^n\xi). \end{aligned}$$

Lemma 4.5.17. *Let H be a Hilbert space and T a self-adjoint operator in H with spectral measure E. For $a, b \in \mathbb{R}$ with $a < b$, the spectral projection $E([a, b])$ is the orthogonal projection onto*

$$\left\{ \xi \in \bigcap_{n=1}^{\infty} \operatorname{dom}(T^n) : \|(T - \frac{1}{2}(a+b))^n \xi\| \leq \frac{(a-b)^n}{2^n} \|\xi\| \text{ for all } n \in \mathbb{N} \right\}.$$

In particular, $E(\{a\})$ is the orthogonal projection onto $\ker(T - a)$.

Remark 4.5.18. If $R \geq 0$, then the previous lemma implies that $E([-R,R])$ is the orthogonal projection onto

$$\left\{\xi \in \bigcap_{n=1}^{\infty} \operatorname{dom}(T^n) : \|T^n\xi\| \leq R^n\|\xi\| \text{ for all } n \in \mathbb{N}\right\},$$

which can be interpreted as the maximal closed subspace on which T acts like a bounded operator with norm less or equal to R.

Example 4.5.19. Let T be a self-adjoint operator in H that admits an orthonormal basis $(\xi_i)_{i\in I}$ of eigenfunctions with associated eigenvalues λ_i, $i \in I$. Since $(\xi_i)_{i\colon \lambda_i=\lambda}$ is an orthonormal basis of $\ker(T-\lambda)$, we have

$$E(\{\lambda\})\xi = \sum_{i\colon \lambda_i=\lambda} \langle \xi_i, \xi\rangle \xi_i$$

for all $\xi \in H$. As $\sum_{\lambda\colon \ker(T-\lambda)\neq\{0\}} E(\{\lambda\}) = 1$, we conclude

$$E(\{\lambda \in \mathbb{R}\colon \ker(T-\lambda) = \{0\}\}) = 0.$$

This completely determines the spectral measure E and we have

$$\begin{aligned} \operatorname{dom}(T) &= \{\xi \in H : \sum_{i\in I} \lambda_i^2 |\langle \xi_i, \xi\rangle|^2 < \infty\}, \\ T\xi &= \sum_{i\in I} \lambda_i \langle \xi_i, \xi\rangle \xi_i. \end{aligned}$$

Example 4.5.20. Let $(X, \mathcal{A}, \mu)$ be a semi-finite measure space and $\varphi\colon X \to \mathbb{R}$ measurable. Recall that the multiplication operator M_φ is defined by

$$\begin{aligned} \operatorname{dom}(M_\varphi) &= \{f \in L^2(X,\mu) \mid \varphi f \in L^2(X,\mu)\}, \\ M_\varphi f &= \varphi f. \end{aligned}$$

As we have seen before, this operator is self-adjoint.

If $f = 0$ μ-a.e. on the complement of $\varphi^{-1}([a,b])$, then

$$\int_X |(\varphi - \frac{1}{2}(a+b))^n f|^2 \, d\mu \leq \frac{(a-b)^{2n}}{2^{2n}} \|f\|_2^2.$$

If on the other hand $\mu(\{x \in X \mid f(x) \neq 0\} \setminus \varphi^{-1}([a,b])) > 0$, then there exists $\varepsilon > 0$ such that

$$\int_{X\setminus\varphi^{-1}([a-\varepsilon,b+\varepsilon])} |f|^2 \, d\mu > 0.$$

Then

$$\begin{aligned}\int_X |(\varphi - \frac{1}{2}(a+b))^n f|^2 \, d\mu &\geq \int_{X\setminus\varphi^{-1}([a-\varepsilon,b+\varepsilon])} |(\varphi - \frac{1}{2}(a+b))^n f|^2 \, d\mu \\ &\geq \left(\frac{a-b}{2} + \delta\right)^{2n} \int_{X\setminus\varphi^{-1}([a-\varepsilon,b+\varepsilon])} |f|^2 \, d\mu\end{aligned}$$

This shows that $\|(M_\varphi - \frac{1}{2}(a+b))^n f\|_2 \leq (\frac{a-b}{2})^n \|f\|_2$ cannot hold. Therefore $E([a,b])$ is the projection onto

$$\left\{f \in L^2(X,\mu) \mid f = 0 \ \mu\text{-a.e. outside } \varphi^{-1}([a,b])\right\}$$

This shows that $E(A) = M_{\mathbb{1}_A \circ \varphi}$ for every closed interval A in $\mathbb{R}$. In fact, the same is true for arbitrary Borel sets, but one needs more measure theory to show this.

As the name and the previous two examples suggest, the spectral theorem is also related to the spectrum.

Proposition 4.5.21. *Let H be a Hilbert space and T a self-adjoint operator in H with spectral measure E. The spectrum of T satisfies*

$$\sigma(T) = \{\lambda \in \mathbb{R} \mid E((\lambda - \varepsilon, \lambda + \varepsilon)) \neq 0 \textit{ for all } \varepsilon > 0\}.$$

In particular, $E(\rho(T)) = 0$.

Going further, the spectral theorem allows for a finer distinction between parts of the spectrum.

Proposition 4.5.22. *Let H be a Hilbert space and T a self-adjoint operator in H with spectral measure E. The sets*

$$\begin{aligned}H_{\mathrm{ac}} &= \{\xi \in H \mid E(A)\xi = 0 \textit{ for all } A \in \mathcal{B}(\mathbb{R}) \textit{ s.t. } \mathcal{L}^1(A) = 0\}, \\ H_{\mathrm{pp}} &= \{\xi \in H \mid \exists \lambda_k \in \mathbb{R}, \, \alpha_k \geq 0 \, \forall A \in \mathcal{B}(\mathbb{R})\colon \|E(A)\xi\|^2 = \sum_{k=1}^\infty \alpha_k \delta_{\lambda_k}(A)\}\end{aligned}$$

are closed orthogonal subspaces of H. Moreover, $T(H_\bullet \cap \operatorname{dom} T) \subset H_\bullet$ and $\sigma(T|_{H_\bullet \cap \operatorname{dom} T}) \subset \sigma(T)$ for $\bullet \in \{\mathrm{ac}, \mathrm{pp}\}$.

Definition 4.5.23 (Absolutely continuous, singularly continuous and pure point spectrum)**.** Let H be a Hilbert space and T a self-adjoint operator in H. The subsets

$$\begin{aligned}\sigma_{\mathrm{ac}}(T) &= \sigma(T|_{H_{\mathrm{ac}} \cap \operatorname{dom} T}) \\ \sigma_{\mathrm{pp}}(T) &= \sigma(T|_{H_{\mathrm{pp}} \cap \operatorname{dom} T}) \\ \sigma_{\mathrm{sc}}(T) &= \sigma(T) \setminus (\sigma_{\mathrm{ac}}(T) \cup \sigma_{\mathrm{pp}}(T))\end{aligned}$$

are called the *absolutely continuous spectrum*, *pure point spectrum* and *singular continuous spectrum* of T.

Remark 4.5.24. If μ is a finite measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then the function

$$F_\mu \colon \mathbb{R} \to [0, \infty),\ t \mapsto \mu((-\infty, t])$$

is called its distribution function. The measure μ is called a *pure point measure* if F_μ is constant except for (countably many) jumps and continuous otherwise.

A continuous measure μ is called *absolutely continuous* if there exists $f \in L^1(\mathbb{R})$ such that

$$F_\mu(t) = \int_{(-\infty,t]} f \, d\mathcal{L}^1$$

for all $t \in \mathbb{R}$.

A continuous measure μ is called *singular continuous* if it is continuous and there exists $N \in \mathcal{B}(\mathbb{R})$ such that $\mathcal{L}^1(N) = 0$ and $\mu(\mathbb{R} \setminus N) = 0$.

The Lebesgue decomposition theorem states that every finite measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ can be uniquely decomposed as $\mu = \mu_{\mathrm{pp}} + \mu_{\mathrm{ac}} + \mu_{\mathrm{sc}}$ with μ_{pp} a pure point measure etc.

4.6 Stone's theorem

Proposition 4.6.1. *Let H be a Hilbert space, T a self-adjoint operator in H and let $U_t = e^{itT}$ for $t \in \mathbb{R}$.*

(a) $U_t^* U_t = U_t U_t^* = 1$ for all $t \in \mathbb{R}$,

(b) $U_0 = 1$,

(c) $U_s U_t = U_{s+t}$ for all $s, t \in \mathbb{R}$,

(d) $\mathbb{R} \to H$, $t \mapsto U_t \xi$ is continuous for all $\xi \in H$.

Proof. (a), (b) and (c) follow immediately from the algebraic properties of functional calculus. To show (d), let E be the spectral measure of T and let (t_n) be a sequence in $\mathbb{R}$ such that $t_n \to t$. By the definition of functional calculus,

$$\|U_{t_n}\xi - U_t\xi\|^2 = \langle \xi, (e^{it_nT} - e^{itT})^*(e^{it_nT} - e^{itT})\xi \rangle = \int_{\mathbb{R}} |e^{it_n\lambda} - e^{it\lambda}|^2 \, d\langle \xi, E(\lambda)\xi \rangle.$$

The last integral converges to zero by the dominated convergence theorem. $\square$

Definition 4.6.2 (Strongly continuous unitary group)**.** Let H be a Hilbert space. A *strongly continuous unitary group* on H is a family $(U_t)_{t\in\mathbb{R}}$ of bounded operators on H such that

(a) $U_t^*U_t = U_tU_t^* = 1$ for all $t \in \mathbb{R}$,

(b) $U_0 = 1$,

(c) $U_sU_t = U_{s+t}$ for all $s, t \in \mathbb{R}$,

(d) $\mathbb{R} \to H$, $t \mapsto U_t\xi$ is continuous for all $\xi \in H$.

The previous proposition shows that if T is a self-adjoint operator, then $(e^{itT})_{t\in\mathbb{R}}$ is a strongly continuous unitary group. Stone's theorem asserts that the converse is also true.

Theorem 4.6.3 (Stone)**.** *Let H be a Hilbert space. If $(U_t)_{t\in\mathbb{R}}$ is a strongly continuous unitary group, then*

$$D = \{\xi \in H \mid \exists \eta \in H\, \forall \zeta \in H \colon \left.\frac{d}{dt}\right|_{t=0} \langle U_t\xi, \zeta\rangle = \langle \eta, \zeta\rangle\}$$

is a dense subspace of H and the operator

$$T\colon D \to H,\ \xi \mapsto -i\eta$$

is a self-adjoint operator in H. Moreover, $U_t = e^{itT}$ for all $t \in \mathbb{R}$.

Proof. Clearly, D is a subspace. For $\xi \in H$ and $\delta > 0$ let

$$f_\delta \colon H \to \mathbb{C},\ \eta \mapsto \frac{1}{\delta}\int_0^\delta \langle U_t\xi, \eta\rangle\, dt.$$

Since

$$|f_\delta(\eta)| \leq \frac{1}{\delta}\int_0^\delta |\langle U_t\xi, \eta\rangle|\, dt \leq \|\xi\|\|\eta\|$$

for all $\eta \in H$, by the Riesz representation theorem there exists a unique $\xi_\delta \in H$ such that $f_\delta = \langle \xi_\delta, \cdot\,\rangle$. Moreover,

$$|f_\delta(\eta) - \langle \xi, \eta\rangle| \leq \frac{1}{\delta}\int_0^\delta |\langle U_t\xi - \xi, \eta\rangle|\, dt \leq \sup_{t\in[0,\delta]} \|U_t\xi - \xi\|\|\eta\|.$$

In particular, for $\eta = \xi - \xi_\delta$, we obtain

$$\|\xi_\delta - \xi\| \leq \sup_{t\in[0,\delta]} \|U_t\xi - \xi\| \to 0, \delta \to 0.$$

We want to show that $\xi_\delta \in D$. Formally,

$$\frac{d}{dt}\bigg|_{t=0} U_t\xi_\delta = \frac{1}{\delta}\int_0^\delta \frac{d}{dt}\bigg|_{t=0} U_{t+s}\xi\, ds = \frac{1}{\delta}\int_0^\delta \frac{d}{ds}U_s\xi\, ds = \frac{1}{\delta}(U_\delta\xi - \xi).$$

Let us make this rigorous:

$$\begin{aligned}
\frac{\delta}{t}\langle U_t\xi_\delta - \xi_\delta, \eta\rangle &= \frac{1}{t}\langle \xi_\delta, U_{-t}\eta - \eta\rangle \\
&= \frac{1}{t}\int_0^\delta \langle U_s\xi, U_{-t}\eta - \eta\rangle\, ds \\
&= \frac{1}{t}\int_t^{t+\delta} \langle U_s\xi, \eta\rangle\, ds - \frac{1}{t}\int_0^\delta \langle U_s\xi, \eta\rangle\, ds \\
&= \frac{1}{t}\int_\delta^{t+\delta} \langle U_s\xi, \eta\rangle\, ds - \frac{1}{t}\int_0^t \langle U_s\xi, \eta\rangle\, ds \\
&\to \langle U_\delta\xi, \eta\rangle - \langle \xi, \eta\rangle.
\end{aligned}$$

Therefore $\xi_\delta \in D$ and $T\xi_\delta = -\frac{i}{\delta}(U_\delta\xi - \xi)$. It follows that D is dense in H.

Next we show that T is self-adjoint. An element $\eta \in H$ belongs to $\mathrm{dom}(T^*)$ if and only if there exists $\zeta \in H$ such that for all $\xi \in \mathrm{dom}(T)$ we have

$$\langle \xi, \zeta\rangle = \langle T\xi, \eta\rangle = i\lim_{t\to 0}\frac{1}{t}\langle U_t\xi - \xi, \eta\rangle = i\lim_{t\to 0}\frac{1}{t}\langle \xi, U_{-t}\eta - \eta\rangle,$$

which holds if and only if $\eta \in \mathrm{dom}(T)$, and in this case $T^*\eta = \zeta = T\eta$.

We next show that $U_t(D) \subset D$ and $TU_t\xi = U_tT\xi$ for $\xi \in D$. In fact, if $\xi \in D$ and $\zeta \in H$, then

$$\frac{1}{h}\langle U_hU_t\xi - U_t\xi, \zeta\rangle = \frac{1}{h}\langle U_h\xi - \xi, U_{-t}\zeta\rangle \overset{h\to 0}{\to} \langle iT\xi, U_{-t}\zeta\rangle = \langle iU_tT\xi, \zeta\rangle.$$

Thus $U_t\xi \in D$ and $TU_t\xi = U_tT\xi$.

It remains to show that $U_t = e^{itT}$ for all $t \in \mathbb{R}$. To do so, we will show that for all $\xi \in D$, the function

$$w\colon \mathbb{R} \to \mathbb{R},\ t \mapsto U_t\xi - e^{itT}\xi$$

is constant.

First note that by the spectral theorem,

$$\begin{aligned}
\left|\langle \xi, \frac{1}{h}(e^{i(t+h)T}\xi - e^{itT}\xi) - ie^{itT}T\xi\rangle\right| &\le \int_{\mathbb{R}} \left|\frac{1}{h}(e^{i(t+h)\lambda} - e^{it\lambda}) - i\lambda e^{it\lambda}\right|\, d\mu_\xi(\lambda) \\
&= \int_{\mathbb{R}} \left|\frac{1}{h}(e^{ih\lambda} - 1) - i\lambda\right|^2\, d\mu_\xi(\lambda).
\end{aligned}$$

Since $\frac{1}{h}(e^{ih\lambda}-1) \to i\lambda$ as $h \to 0$ *uniformly in* λ (exercise), the integral on the right side converges to 0 as $h \to 0$. By the polarization identity,

$$\lim_{h\to 0} \frac{1}{h}\langle \eta, e^{i(t+h)T}\xi - e^{itT}\xi\rangle = \langle \eta, ie^{itT}T\xi\rangle$$

for all $\xi, \eta \in D$.

Combined with the result for U_t, this implies that

$$\lim_{h\to 0} \frac{1}{h}\langle w(t+h) - w(t), \eta\rangle = \langle iTw(t), \eta\rangle.$$

Therefore,

$$\begin{aligned}\frac{1}{h}\|w(t+h)-w(t)\|^2 &= \frac{1}{h}\langle w(t+h)-w(t), w(t+h)\rangle + \frac{1}{h}\langle w(t), w(t+h)-w(t)\rangle \\ &\overset{h\to 0}{\to} \langle iTw(t), w(t)\rangle + \langle w(t), iTw(t)\rangle \\ &= \langle w(t), -iTw(t) + iTw(t)\rangle \\ &= 0.\end{aligned}$$

Hence w is constant. As $w(0) = 0$, we conclude that $U_t\xi = e^{itT}\xi$ for all $t \in \mathbb{R}$ and $\xi \in D$. Finally, since D is dense in H and U_t and e^{itT} are bounded, the equality extends to all $\xi \in H$ by continuity. $\square$

Remark 4.6.4. In the last step of the proof we tacitly used that if (ξ_n) and (η_n) are sequences in H such that $\xi_n \to \xi$ and $\langle \eta_n, \zeta\rangle \to \langle \eta, \zeta\rangle$ for all $\zeta \in H$, then $\langle \xi_n, \eta_n\rangle \to \langle \xi, \eta\rangle$. The proof of this fact is left as an exercise.

Remark 4.6.5. If (η_n) is a sequence in H such that $\langle \eta_n, \zeta\rangle \to \langle \eta, \zeta\rangle$ for all $\zeta \in H$, then one says that (η_n) converges *weakly* to η. In general, weak convergence does not imply convergence in the norm of H. For example, if (e_n) is an orthonormal basis of H, then e_n converges weakly to zero, but $\|e_n\| = 1$ for all $n \in \mathbb{N}$, so e_n cannot converge to zero in norm. In the situation of the proof of Stone's theorem however it is true that $\frac{1}{h}(U_h\xi - \xi) \to iT\xi$ holds in norm for all $\xi \in D$.

Remark 4.6.6. Informally, Stone's theorem states that for any self-adjoint operator T, the solution operator U_t that maps $\xi \in H$ to the solution $\xi(t)$ of the initial-value problem

$$\begin{cases} \frac{d}{dt}\xi(t) &= iT\xi(t) \\ \xi(0) &= \xi \end{cases}$$

In the case when T is the Hamiltonian of a physical system, this initial-value problem describes the time evolution of the state of the system with initial state ξ (time-dependent Schrödinger equation).

Note however that the strongly continuous unitary group U_t gives a mathematically well-defined time evolution for all initial states whereas the Schrödinger equation only makes sense if $\xi(t) \in \mathrm{dom}(T)$, which is guaranteed for $\xi \in \mathrm{dom}(T)$, but not for general initial states.

This is an instance of the general phenomenon for ordinary and partail differential equations that it is often easier to rigorously define a general notion of *solution* of the equation even when the equation itself is not well-defined in this generality.