ANALYSIS 4 (MATH 206) SPRING 2020

MATTHIAS RUF

Acknowledgments. The author would like to thank Kenneth Jannasch, Antoine Maier, Salomon Guinchard and Baptiste Claudon for pointing out some mistakes/typos and asking interesting questions which lead to an improvement of these lecture notes. If you find further mistakes or have other suggestions, you can send an email to matthias.ruf@epfl.ch.

1. Complex analysis in one variable

Recall that $\mathbb{C} = \{z = x + iy : x, y \in \mathbb{R}\}$ and the complex unit i satisfies $i^2 = -1$. As such $\mathbb{C} \approx \mathbb{R}^2$ via $x + iy \mapsto (x, y)$. Moreover, for $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ we define

- $z_1 + z_2 := (x_1 + x_2) + i(y_1 + y_2)$;
- $z_1 \cdot z_2 := (x_1x_2 - y_1y_2) + i(x_1y_2 + x_2y_1)$;
- $\overline{z_1} = x_1 - iy_1$;
- $|z_1| = \sqrt{x_1^2 + y_1^2}$;
- $z_1^{-1} = \frac{\overline{z_1}}{|z_1|^2}$ provided $z_1 \neq 0$.

We recall the following properties that follow directly from the definitions:

- $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$;
- $\overline{z_1 \cdot z_2} = \overline{z_1} \cdot \overline{z_2}$;
- $\overline{\overline{z_1}} = z_1$;
- $\overline{z_1^{-1}} = (\overline{z_1})^{-1}$;
- $|z_1 \cdot z_2| = |z_1||z_2|$;
- $(z_1 \cdot z_2)^{-1} = z_1^{-1} \cdot z_2^{-1}$;

Note that $|\cdot|$ defines a norm on $\mathbb{C}$ which makes it a Hilbert space with the scalar product $(z_1, z_2) = z_1 \cdot \overline{z_2}$. Then we can define open sets in $\mathbb{C}$ as usual and we can define the convergence of sequences $(z_n)_{n\in\mathbb{N}}$ to some $z \in \mathbb{C}$ by requiring that $\lim\limits_{n\to+\infty} |z_n - z| = 0$. We will use the following notation:

- $\mathbb{C}$: complex numbers
- U: an open subset of $\mathbb{C}$
- D: a domain (open and path-connected subset of $\mathbb{C}$)
- $B_r(z_0) = \{z \in \mathbb{C} : |z - z_0| < r\}$: open ball with radius $r > 0$ and center $z_0 \in \mathbb{C}$
- $\partial B_r(z_0) = \{z \in \mathbb{C} : |z - z_0| = r\}$: the boundary of $B_r(z_0)$

Recall that a set S is path-connected if for any two points $z_1, z_2 \in S$ there exists a continuous curve $\gamma : [0, 1] \to S$ such that $\gamma(0) = z_1$ and $\gamma(1) = z_2$.

1.1. Holomorphic functions.

Now we come to the central object of this chapter.

Definition 1.1. A function $f : U \to \mathbb{C}$ is called complex differentiable in $z_0 \in U$ if there exists the limit

$$f'(z_0) = \lim_{\substack{h\to 0\\ h\neq 0}} \frac{f(z_0 + h) - f(z_0)}{h} \in \mathbb{C}.$$

It is called holomorphic on U if it is complex differentiable in every $z_0 \in U$. If $U = \mathbb{C}$ we say that f is entire.

Remark 1.2. The notion $\lim_{\substack{h\to 0\\ h\neq 0}}$ means that the limit exists for every sequence $h_n \to 0$ with $h_n \neq 0$ for all $n \in \mathbb{N}$. In particular, the sequence can approach 0 from different directions.

As we shall see the complex differentiability of a function implies a lot of nice properties. For instance, f is then automatically infinitely many times differentiable and can be represented locally as a power series.

But first let us stress the difference to differentiability when $f = u + iv$ is identified with a function $\widetilde{f} : \widetilde{U} \subset \mathbb{R}^2 \to \mathbb{R}^2$ via $\widetilde{f}(x, y) = (u(x + iy), v(x + iy))$. To this end, note that for fixed $w \in \mathbb{C}$ the multiplication $z \mapsto w \cdot z$ can be viewed as a linear mapping from $\mathbb{R}^2$ to $\mathbb{R}^2$. On the one hand, writing $w = a + ib$ and $z = x + iy$ we derive the matrix representation

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} ax - by \\ ay + bx \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

On the other hand, writing $w = |w|e^{i\varphi}$ in polar coordinates, we see that $a = |w| \cos(\varphi)$ and $b = |w| \sin(\varphi)$, so that

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} = |w| \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi). \end{pmatrix}$$

Hence the multiplication in $\mathbb{C}$ represents a rotation followed by a dilation. In particular, it preserves angles. Since Definition 1.1 implies that

$$f(z) = f(z_0) + f'(z_0)(z - z_0) + o(|z - z_0|),$$

we conclude that $f(z)$ is locally approximated by a dilation combined with a rotation. Recall that real differentiability of $\widetilde{f}$ requires that there exists a matrix $D\widetilde{f}(z_0) \in \mathbb{R}^{2\times 2}$ such that

$$\widetilde{f}(z) = \widetilde{f}(z_0) + D\widetilde{f}(z_0)(z - z_0) + o(|z - z_0|).$$

The matrix $D\widetilde{f}(z_0)$ contains the partial derivatives and is of the form

$$D\widetilde{f}(z_0) = \begin{pmatrix} \partial_x u(z_0) & \partial_y u(z_0) \\ \partial_x v(z_0) & \partial_y v(z_0) \end{pmatrix}.$$

The main difference to real differentiability is that we require a complex product as linear approximation instead of a matrix-vector product. As we have seen this corresponds to a special structure of the matrix-vector product. Using again the identification $\mathbb{C} \approx \mathbb{R}^2$ this leads us to the following proposition.

Proposition 1.3 (Cauchy-Riemann equations). *A function $f = u + iv : U \to \mathbb{C}$ is complex differentiable in $z_0 \in U$ if and only if $\widetilde{f} : U \to \mathbb{R}^2$ defined above is real differentiable in z_0 and satisfies*

$$\partial_x u(z_0) = \partial_y v(z_0),$$
$$\partial_x v(z_0) = -\partial_y u(z_0).$$

Remark 1.4. As for the standard derivative one can prove the following properties:

(i) If $f, g : U \to \mathbb{C}$ are complex-differentiable in $z_0 \in \mathbb{C}$, then so are $f+g$, $f \cdot g$ and, provided $g(z_0) \neq 0$, also $1/g$. Moreover, $(f + g)'(z_0) = f'(z_0) + g'(z_0)$ and $(f \cdot g)'(z_0) = f(z_0)g'(z_0) + f'(z_0)g(z_0)$, while $(1/g)'(z_0) = -g'(z_0)/g(z_0)^2$.

(ii) If $f : U \to \mathbb{C}$ is holomorphic and $g : V \to U$ is holomorphic, then $f \circ g : V \to \mathbb{C}$ is holomorphic and $(f \circ g)'(z_0) = f'(g(z_0))g'(z_0)$.

Example 1.5. (i) Constant functions $f(z) = c$ are holomorphic with $f'(z) = 0$ for all $z \in \mathbb{C}$.

(ii) The linear function $f(z) = z$ is holomorphic on $\mathbb{C}$. By the previous remark it follows that every polynomial $p(z) = a_0 + a_1 z + \ldots + a_n z^n$ is holomorphic with $p'(z) = a_1 + \ldots + n a_n z^{n-1}$.

(iii) $f(z) = z^{-n}$ with $n \in \mathbb{N}$ is holomorphic on $\mathbb{C} \setminus \{0\}$ with $f'(z) = -nz^{-n-1}$.

(iv) convergent power series $f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k$ are holomorphic on the open disc $B_R(z_0)$, where R denotes the radius of convergence. It holds that $f'(z) = \sum_{k=1}^{\infty} k a_k (z - z_0)^{k-1}$ (see exercise H 1.3). In particular, the complex exponential function $\exp(z) = \sum_{k=0}^{\infty} \frac{z^k}{k!}$ is an entire function and $\exp'(z) = \exp(z)$.

(v) the complex conjugation $z \mapsto \overline{z}$ fails to be complex differentiable at any point $z \in \mathbb{C}$ (see exercise H 1.2).

1.2. **Complex curve integrals.** The first important result about holomorphic functions is Cauchy's integral formula. Before we state it let us recall the notion of curve integrals. In what follows we will always assume that $\gamma : [a,b] \subset \mathbb{R} \to U$ is a piecewise continuously differentiable curve. We say that a curve is closed if $\gamma(a) = \gamma(b)$.

Definition 1.6. Let $\gamma : [a,b] \to U$ be a piecewise C^1-curve and $f : U \to \mathbb{C}$ be holomorphic. The complex curve integral of f along γ is defined as

$$\int_\gamma f \,\mathrm{d}z := \int_a^b f(\gamma(t))\gamma'(t)\,\mathrm{d}t.$$

In the above formula the integrand on the right hand side is complex-valued. Integration has to be understood by integrating separately the real-and imaginary part.

Let us collect some elementary properties of the complex curve integral.

Lemma 1.7. *Let $\gamma : [a,b] \to U$ be a piecewise C^1-path and $f : U \to \mathbb{C}$ be holomorphic.*

(i) The curve integral is invariant under orientation preserving reparametrizations, i.e., if $\varphi : [a_1, b_1] \to [a,b]$ is a bijective piecewise C^1-function such that $\varphi'(s) > 0$ for all $s \in [a_1, b_1]$ then for the curve $\gamma_1 : [a_1, b_1] \to U$ defined by $\gamma_1 = \gamma \circ \varphi$ we have

$$\int_{\gamma_1} f \,\mathrm{d}z = \int_\gamma f \,\mathrm{d}z.$$

If instead $\varphi'(s) < 0$ for all $s \in [a_1, b_1]$ then the integral changes its sign.

(ii) (fundamental estimate)

$$\left| \int_\gamma f \,\mathrm{d}z \right| \le \sup_{s \in [a,b]} |f(\gamma(s))| \underbrace{\int_a^b |\gamma'(t)|\,\mathrm{d}t}_{=:L(\gamma):\ \textit{length of } \gamma} .$$

(iii) Suppose there exists a primitive $F : U \to \mathbb{C}$ of f, i.e., F is holomorphic on U with $F' = f$. Then

$$\int_\gamma f \,\mathrm{d}z = F(\gamma(b)) - F(\gamma(a)).$$

In particular, the curve integral equals 0 if γ is closed.

Proof. (i) We perform the change of variables $s = \varphi(t)$, so that by definition

$$\int_{\gamma_1} f \,\mathrm{d}z = \int_{a_1}^{b_1} f(\gamma(\varphi(t)))\gamma'(\varphi(t))\varphi'(t)\,\mathrm{d}t = \int_a^b f(\gamma(s))\gamma'(s)\,\mathrm{d}s = \int_\gamma f \,\mathrm{d}z,$$

where we used that due to the monotonicity of φ we have that $\varphi(a_1) = a$ and $\varphi(b_1) = b$. Here we assumed that the function is C^1. The general case can be treated by partitioning the interval into finitely many smaller ones where φ is C^1. If φ is monotone decreasing, then the change of variables comes with a sign change.

(ii) By the triangle inequality for integrals we have

$$\left| \int_a^b f(\gamma(t))\gamma'(t)\,\mathrm{d}t \right| \le \int_a^b |f(\gamma(t))||\gamma'(t)|\,\mathrm{d}t \le \sup_{s \in [a,b]} |f(\gamma(s))| \int_a^b |\gamma'(t)|\,\mathrm{d}t$$

as claimed.

(iii) By the chain rule it holds that

$$\frac{\mathrm{d}}{\mathrm{d}t} F(\gamma(t)) = f(\gamma(t))\gamma'(t).$$

Integrating this equality over the interval $[a,b]$ yields the claim by the fundamental theorem of calculus. □

The last point raises the question if every holomorphic function possesses a primitive. As we shall see this strongly depends on the geometry of the set U where the function f is defined. Let us start with an illustrative example.

Example 1.8. Let $f : \mathbb{C} \setminus \{0\} \to \mathbb{C}$ be defined by $f(z) = z^{-n}$ for $n \in \mathbb{N}$. Consider the closed curve $\partial B_1(0) = \{z \in \mathbb{C} : |z| = 1\}$ oriented counter-clockwise, so that we can choose the parametrization $\gamma : [0, 2\pi] \to \mathbb{C}$ given by $\gamma(t) = e^{it}$. Then the complex curve integral equals

$$\int_\gamma f \,\mathrm{d}z = \int_0^{2\pi} e^{-int} i e^{it} \,\mathrm{d}t = i \int_0^{2\pi} e^{(1-n)it} \,\mathrm{d}t = \begin{cases} 2\pi i & \text{if } n = 1, \\ \frac{1}{(1-n)} \left(e^{2\pi(1-n)i} - 1\right) & \text{if } n \geq 2 \end{cases}$$
$$= \begin{cases} 2\pi i & \text{if } n = 1, \\ 0 & \text{if } n \geq 2. \end{cases}$$

Hence there cannot exist a primitive of the function $z \mapsto \dfrac{1}{z}$ on $\mathbb{C} \setminus \{0\}$. In particular, there exists no 'logarithm' on $\mathbb{C} \setminus \{0\}$.

Next we study which conditions on the set U ensure the existence of primitives for all holomorphic functions $f : U \to \mathbb{C}$. We introduce the following geometric properties:

Definition 1.9. A set $M \subset \mathbb{C}$ is called star-shaped with respect to $z_0 \in M$ if for all $z \in M$ the straight line between z_0 and z, denoted by $[z_0, z]$, is contained in M.

Remark 1.10.
- An open star-shaped set is automatically path-connected, so it is a domain.
- Convex sets M are star-shaped with respect to every point $z_0 \in M$.
- the sliced complex plane $\mathbb{C}^- := \mathbb{C} \setminus \{x \in \mathbb{R} : x \leq 0\}$ is star-shaped.
- the punctured complex plane $\mathbb{C} \setminus \{0\}$ is not star-shaped.

We will now prove an auxiliary result on the existence of primitives.

Lemma 1.11. *Let $U \subset \mathbb{C}$ be a star-shaped domain with respect to some $z_0 \in U$ and $f : U \to \mathbb{C}$ be continuous such that*

$$\int_{\partial\Delta} f(z) \,\mathrm{d}z = 0$$

for all closed triangles $\Delta \subset U$. Then the curve integral

$$F(z) = \int_{[z_0,z]} f(\xi) \,\mathrm{d}\xi$$

defines a primitive of f.

Proof. Let $z \in U$ be arbitrary. We have to show that F is complex differentiable in z with $F'(z) = f(z)$. Fix $r > 0$ such that $B_r(z) \subset U$ and fix $w \in B_r(z) \setminus \{z\}$. Choose a triangle Δ with corners z_0, z and w. Then according to the assumption

$$0 = \int_{\partial\Delta} f(\xi) \,\mathrm{d}\xi = \underbrace{\int_{[z_0,z]} f(\xi) \,\mathrm{d}\xi}_{=F(z)} + \int_{[z,w]} f(\xi) \,\mathrm{d}\xi + \underbrace{\int_{[w,z_0]} f(\xi) \,\mathrm{d}\xi}_{-F(w)}.$$

Hence we deduce that

$$F(w) - F(z) = \int_{[z,w]} f(\xi) \,\mathrm{d}\xi.$$

Since the length of $\int_{[z,w]} 1 \,\mathrm{d}\xi = w - z$ it follows that

$$\frac{F(w) - F(z)}{w - z} - f(z) = \frac{1}{w - z} \int_{[z,w]} f(\xi) - f(z) \,\mathrm{d}\xi$$

We are done if we show that the right hand side term tends to 0 when $w \to z$. To this end, we apply the fundamental estimate which yields that

$$\left| \frac{1}{w-z} \int_{[z,w]} f(\xi) - f(z)\,\mathrm{d}\xi \right| \le \frac{L([z,w])}{|w-z|} \sup_{t\in[z,w]} |f(t)-f(z)| = \sup_{t\in[z,w]} |f(t)-f(z)|.$$

Since f is continuous in z, the last term vanishes when $w \to z$. □

The next lemma ensures that holomorphic functions satisfy the triangle condition of the previous lemma.

Lemma 1.12 (Goursat's lemma)**.** *Let $U \subset \mathbb{C}$ be open and $f : U \to \mathbb{C}$ be holomorphic. If $\Delta \subset U$ is a closed triangle , then*

$$\int_{\partial\Delta} f(z)\,\mathrm{d}z = 0.$$

Proof. Fix such a triangle Δ. We subdivide it into four smaller triangles $(\Delta_1^j)_{j=1}^4$ by joining the midpoints of each side of Δ and orient all triangle boundaries counterclockwise (see Figure 1). Then

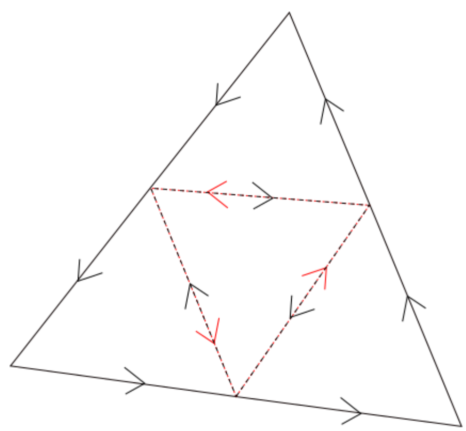

FIGURE 1. Partitioning of the initial triangle Δ.

$$\int_{\partial\Delta} f(z)\,\mathrm{d}z = \sum_{j=1}^{4} \int_{\partial\Delta_1^j} f(z)\,\mathrm{d}z$$

since the contributions from the inner triangle are canceled. Choose an index $j_1 \in \{1,\dots,4\}$ such that $\left| \int_{\Delta_1^{j_1}} f(z)\,\mathrm{d}z \right|$ is maximal. Set then $\Delta_1 = \Delta_1^{j_1}$, so that

$$\left| \int_{\partial\Delta} f(z)\,\mathrm{d}z \right| \le 4 \left| \int_{\partial\Delta_1} f(z)\,\mathrm{d}z \right|.$$

Now we iterate this procedure to obtain a sequence of triangles $(\Delta_n)_{n\in\mathbb{N}}$ such that

$$\left| \int_{\partial\Delta} f(z)\,\mathrm{d}z \right| \le 4^n \left| \int_{\partial\Delta_n} f(z)\,\mathrm{d}z \right|.$$

Moreover, note that in each step the side lengths decrease by a factor of 2, so that

$$L(\Delta_n) = 2^{-n} L(\Delta).$$

The sequence of triangles $(\Delta_n)_{n\in\mathbb{N}}$ is nested and each Δ_n is a compact set. Hence the intersection is non-empty[1] and there exists

$$z_0 \in \bigcap_{n=1}^{\infty} \Delta_n.$$

Next we expand the holomorphic function f in z_0 via $f(z) = f(z_0) + f'(z_0)(z - z_0) + o(|z - z_0|)$. Then

$$\int_{\partial\Delta_n} f(z)\,\mathrm{d}z = \underbrace{\int_{\partial\Delta_n} f(z_0) + f'(z_0)(z - z_0)\,\mathrm{d}z}_{=0 \text{ since there exists a primitive}} + \int_{\partial\Delta_n} o(|z - z_0|)\,\mathrm{d}z.$$

We conclude from the fundamental estimate that

$$\left|\int_{\partial\Delta} f(z)\,\mathrm{d}z\right| \leq 4^n \left|\int_{\partial\Delta_n} o(|z - z_0|)\,\mathrm{d}z\right| \leq 4^n L(\Delta_n) o(L(\Delta_n)) = 2^n o(L(\partial\Delta_n)) \to 0$$

as $n \to +\infty$. This proves the claim. □

Now we are in a position to prove the famous Cauchy integral theorem on star-shaped domains.

Theorem 1.13 (Cauchy's integral theorem, version I)**.** *Let $U \subset \mathbb{C}$ be a star-shaped domain with respect to $z_0 \in \mathbb{C}$ and $f : U \to \mathbb{C}$ be holomorphic. Then*

$$F(z) = \int_{[z_0,z]} f(\xi)\,\mathrm{d}\xi$$

defines a holomorphic primitive of f. In particular, the curve integral of f over curves in U depends only on the start and endpoints.

Proof. Applying Goursat's lemma this is a direct consequence of Lemma 1.11 and Lemma 1.7 . □

We now apply this result to prove a non-trivial invariance property of the complex curve integral for holomorphic functions.

Corollary 1.14. *Let $U \subset \mathbb{C}$ be open and $z_0 \in U$. Suppose that $f : U \setminus \{z_0\} \to \mathbb{C}$ is holomorphic. Let $r, R > 0$ and $c \in \mathbb{C}$ be such that $B_r(z_0) \subset B_R(c)$ and $\overline{B_R(c)} \subset U$. Then*

$$\int_{\partial B_r(z_0)} f\,\mathrm{d}z = \int_{\partial B_R(c)} f\,\mathrm{d}z,$$

where both circles $\partial B_r(z_0), \partial B_R(c)$ are oriented counterclockwise.

Proof. Since $\overline{B_R(c)} \subset U$ there exists a slightly larger open ball $B^* \subset U$ such that $\overline{B_R(c)} \subset B^*$ and f is holomorphic on B^*. Note that due to convexity the infinite line trough c and z_0 (take any line though z_0 if $z_0 = c$) intersects both $\partial B_r(z_0)$ and $\partial B_R(c)$ in exactly two points. Let z_R^1, z_R^2 be the intersection points in $\partial B_R(c)$ and z_r^1, z_r^2 be the intersection points in $\partial B_r(z_0)$, respectively. Then we define the following paths:

- α a straight line from z_R^1 to z_r^1, where we assume without loss of generality that z_r^1 is closer to z_R^1 than z_r^2;
- γ_1 a half circle on $\partial B_r(z_0)$ from z_r^1 to z_r^2 oriented clockwise;
- γ_2 a half circle on $\partial B_r(z_0)$ from z_r^2 to z_r^1 oriented clockwise;
- β a straight line from z_r^2 to z_R^2;
- γ_3 a half circle on $\partial B_R(c)$ from z_R^1 to z_R^2 oriented counterclockwise;
- γ_4 a half circle on $\partial B_R(c)$ from z_R^2 to z_R^1 oriented counterclockwise;

see also Figure 2. Then consider the points $a_1, a_2 \in \partial B^*$ that are the two intersection points of the line

[1]Construct a sequence z_n by picking any $z_n \in \Delta_n$. Since Δ_1 is compact there exists a subsequence z_{n_j} such that $z_{n_j} \to z_0 \in \Delta_1$. Since $z_n \in \Delta_2$ for all $n \geq 2$ the limit belongs also to Δ_2. Continuing this reasoning we find that $z_0 \in \Delta_n$ for all $n \in \mathbb{N}$.

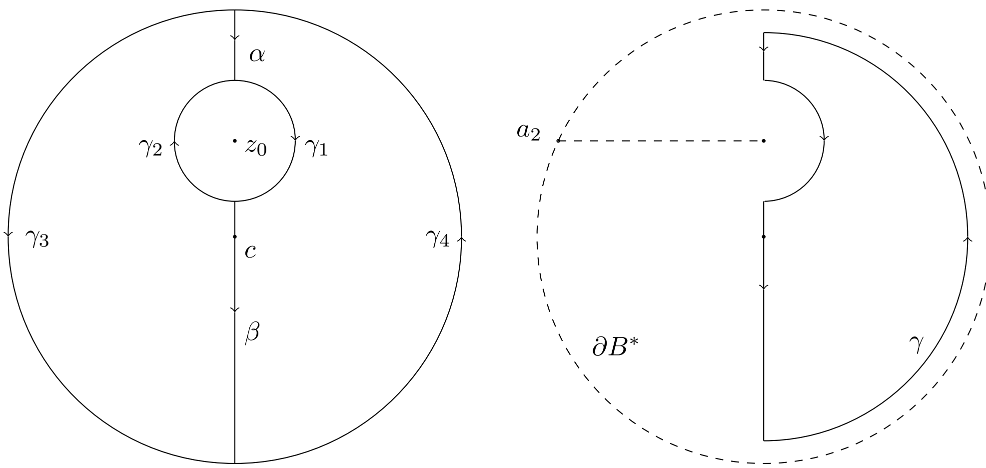

FIGURE 2. A sketch of the paths defined above and the sliced ball (created by Gianluca Orlando).

through z_0 that is orthogonal to the previously chosen line and ∂B^*. We may assume that $[z_0, a_i] \cap \gamma_i \neq 0$ for $i = 1, 2$. Define the two closed paths

$$\gamma = \alpha + \gamma_1 + \beta + \gamma_4, \qquad \gamma' = \gamma_3 - \beta + \gamma_2 - \alpha,$$

where with a slight abuse of notation we write $+$ if we concatenate the paths from left to right and $-$ if we concatenate the paths from left to right, but the one on the right hand side of the $-$ sign in the reverse direction. Note that the set $B^* \setminus [z_0, a_i]$ is star-shaped for $i = 1, 2$ and $\gamma([0,1]) \subset B^* \setminus [z_0, a_2]$ and $\gamma'([0,1]) \subset B^* \setminus [z_0, a_1]$. Hence from Theorem 1.13 we conclude that

$$\int_\alpha f \,\mathrm{d}z + \int_{\gamma_1} f \,\mathrm{d}z + \int_\beta f \,\mathrm{d}z + \int_{\gamma_4} f \,\mathrm{d}z = \int_\gamma f \,\mathrm{d}z = 0,$$
$$\int_{\gamma_3} f \,\mathrm{d}z - \int_\beta f \,\mathrm{d}z + \int_{\gamma_2} f \,\mathrm{d}z - \int_\alpha f \,\mathrm{d}z = \int_{\gamma'} f \,\mathrm{d}z = 0.$$

Adding these two equations the integrals over α and β cancel and by definition of the remaining paths we obtain the claim. □

The previous result is a special case of a more general invariance property involving more topological notions that we introduce for the sake of completeness:

Definition 1.15 (Homotopy and simply connected sets)**.** Let $U \subset \mathbb{C}$ be an open set.

(i) two curves[2] $\gamma_1, \gamma_2 : [0,1] \to U$ with same start and endpoints are called homotopic in U if there exists a continuous function $H : [0,1] \times [0,1] \to U$ such that

$$H(0,t) = \gamma_1(t), \quad H(1,t) = \gamma_2(t), \quad H(s,0) = \gamma_i(0), \quad H(s,1) = \gamma_i(1) \text{ for } i = 1,2;$$

(i.e., they can be continuously deformed into each other keeping the start and endpoint fixed)

(ii) two closed curves $\gamma_1, \gamma_2 : [0,1] \to U$ are called freely homotopic in U if there exists a continuous function $H : [0,1] \times [0,1] \to U$ such that

$$H(0,t) = \gamma_1(t), \quad H(1,t) = \gamma_2(t), \quad H(s,0) = H(s,1);$$

(i.e., they can be continuously deformed into each other preserving closedness)

(iii) a closed curve $\gamma : [0,1] \to U$ is called null-homotopic in U if it is freely homotopic in U to a constant curve.

(iv) a domain $U \subset \mathbb{C}$ is called simply connected if every closed curve $\gamma : [0,1] \to U$ is null-homotopic in U.

[2]Here we always assume that curves are defined on $[0,1]$ which can always be achieved by reparametrization.

In the exercises we will see that star-shaped domains are always simply connected, but that the converse is not true. Hence the result below is a strict generalization of Theorem 1.13. However, a proof of this theorem goes beyond the scope of this course.

Theorem 1.16 (Cauchy's integral theorem, version II). *Let $U \subset \mathbb{C}$ be open and $f : U \to \mathbb{C}$ be holomorphic.*

(i) If $\gamma_1, \gamma_2 : [0,1] \to U$ have the same start and endpoint and are homotopic in U or are closed and freely homotopic in U then

$$\int_{\gamma_1} f \,\mathrm{d}z = \int_{\gamma_2} f \,\mathrm{d}z;$$

(ii) If $\gamma : [0,1] \to U$ is closed and null-homotopic in U then

$$\int_{\gamma} f \,\mathrm{d}z = 0;$$

(iii) If U is simply connected then there exists a primitive $F : U \to \mathbb{C}$ of f. Fixing any $z_0 \in U$ it can be taken of the form

$$F(z) = \int_{\gamma_z} f \,\mathrm{d}\xi,$$

where $\gamma_z : [0,1] \to U$ is any curve that connects z_0 and z.

Remark 1.17. (i) The definition of $F(z)$ is independent of the curve γ_z due to property (ii) when U is simply connected.

(ii) (complex logarithm) The formula for $F(z)$ can be used to derive an expression for the complex logarithm as a primitive on $1/z$ (see also H 2.4). Writing any $z \in \mathbb{C}^- = \mathbb{C} \setminus (-\infty, 0]$ as $z = |z|e^{i\varphi}$ with $\varphi \in (-\pi, \pi)$ we construct γ_z as the straight line from 1 to $|z|$ composed with an arc of circle on $\partial B_{|z|}(0)$. Then

$$\log(z) = \int_0^1 \frac{|z|-1}{1+t(|z|-1)} \,\mathrm{d}t + \int_0^{\varphi} |z|^{-1}e^{-it} i|z|e^{it} \,\mathrm{d}t = \ln(|z|) + i\varphi.$$

1.3. **Consequences of the Cauchy integral theorem.** In this section we derive strong properties of holomorphic functions which follow from Cauchy's integral theorem. We start with another theorem due to Cauchy which expresses a holomorphic function by a curve integral. In particular, it implies that in order to know a holomorphic function in a ball it suffices to know its values on a curve.

Theorem 1.18 (Cauchy's integral formula on discs). *Let $U \subset \mathbb{C}$ be open and $f : U \to \mathbb{C}$ be holomorphic. Let $z_0 \in U$ and $r > 0$ be such that $\overline{B_r(z_0)} \subset U$. Then for all $z \in B_r(z_0)$ we have*

$$f(z) = \frac{1}{2\pi i} \int_{\partial B_r(z_0)} \frac{f(\zeta)}{\zeta - z} \,\mathrm{d}\zeta.$$

Proof. Fix $z \in B_r(z_0)$ and define the function

$$g(\zeta) = \begin{cases} \frac{f(\zeta)-f(z)}{\zeta - z} & \text{if } \zeta \in U \setminus \{z\}, \\ f'(z) & \text{if } \zeta = z. \end{cases}$$

Since f is complex differentiable in z the function g is continuous on U. We show that $\int_{\partial B_r(z_0)} g \,\mathrm{d}\zeta = 0$. Indeed, since g is continuous and $\overline{B_r(z_0)}$ is a compact set g is bounded on $\overline{B_r(z_0)}$. From Corollary 1.14 we infer that for any $\varepsilon > 0$ small enough such that $B_\varepsilon(z) \subset B_r(z_0)$ we have by the fundamental estimate

$$\left| \int_{\partial B_r(z_0)} g \,\mathrm{d}\zeta \right| = \left| \int_{\partial B_\varepsilon(z)} g \,\mathrm{d}\zeta \right| \leq 2\pi\varepsilon \sup_{\zeta \in \partial B_\varepsilon(z)} |g(\zeta)| \leq 2\pi\varepsilon \sup_{\zeta \in \overline{B_r(z_0)}} |g(\zeta)|.$$

Letting $\varepsilon \to 0$ we conclude that indeed $\int_{\partial B_r(z_0)} g \,\mathrm{d}\zeta = 0$. Next we apply again Corollary 1.14 to the function $h(\zeta) = (\zeta - z)^{-1}$, which is holomorphic on $\mathbb{C} \setminus \{z\}$. Again for $\varepsilon > 0$ small enough we know that

$B_\varepsilon(z) \subset B_r(z_0)$ and therefore

$$\int_{\partial B_r(z_0)} \frac{1}{\zeta - z}\,\mathrm{d}\zeta = \int_{\partial B_\varepsilon(z)} \frac{1}{\zeta - z}\,\mathrm{d}\zeta = 2\pi i,$$

where the last equality can be proven as in Example 1.8. Gathering what we proved so far we deduce that

$$0 = \int_{\partial B_r(z_0)} g\,\mathrm{d}\zeta = \int_{\partial B_r(z_0)} \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}\zeta - f(z)\int_{\partial B_r(z_0)} \frac{1}{(\zeta - z)}\,\mathrm{d}\zeta = \int_{\partial B_r(z_0)} \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}\zeta - f(z)2\pi i.$$

Rearranging terms yields the claim. □

Corollary 1.19 (Mean value property)**.** *Under the assumptions of Theorem 1.18 it holds that*

$$f(z) = \frac{1}{2\pi}\int_0^{2\pi} f(z + re^{it})\,\mathrm{d}t.$$

Proof. Insert the parametrization $t \mapsto z + re^{it}$ for $t \in [0, 2\pi]$ in Theorem 1.18 and evaluate the integral with $z_0 = z$. □

Next we prove the announced analyticity of holomorphic functions.

Theorem 1.20. *Let $U \subset \mathbb{C}$ be open and $f : U \to \mathbb{C}$ be holomorphic. Then for each $z_0 \in \mathbb{C}$ and $r > 0$ with $\overline{B_r(z_0)} \subset U$ we can write*

$$f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k \qquad \forall z \in B_r(z_0)$$

with $a_k = \dfrac{1}{2\pi i}\displaystyle\int_{\partial B_r(z_0)} \frac{f(z)}{(z - z_0)^{k+1}}\,\mathrm{d}z.$

Proof. Without loss of generality we set $z_0 = 0$ (otherwise consider $z \mapsto f(z_0 + z)$). Given $r > 0$ as in the theorem we can apply Cauchy's integral formula to deduce

$$f(z) = \frac{1}{2\pi i}\int_{\partial B_r(z_0)} \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}\zeta = \frac{1}{2\pi i}\int_{\partial B_r(z_0)} \frac{f(\zeta)}{\zeta}\underbrace{\frac{1}{1 - \frac{z}{\zeta}}}_{= \sum_{k=0}^{\infty} \left(\frac{z}{\zeta}\right)^k}\,\mathrm{d}\zeta,$$

where the geometric series formula holds since $|z| < r = |\zeta|$. Moreover, since z is fixed the geometric series converges uniformly in $\xi \in \partial B_r(z_0)$, so that we can exchange the sum with the integral which yields

$$f(z) = \sum_{k=0}^{\infty}\left(\frac{1}{2\pi i}\int_{\partial B_r(z_0)} \frac{f(\zeta)}{\zeta^{k+1}}\mathrm{d}\zeta\right) z^k = \sum_{k=0}^{\infty} a_k z^k.$$

□

Corollary 1.21. *Let $U \subset \mathbb{C}$ be open and $f : U \to \mathbb{C}$ be holomorphic. Then f is infinitely complex differentiable in each $z_0 \in U$ and each derivative $f^{(n)}$ is holomorphic on U. Moreover, if $\overline{B_r(z_0)} \subset U$ the derivatives can be calculated by*

$$f^{(n)}(z_0) = \frac{n!}{2\pi i}\int_{\partial B_r(z_0)} \frac{f(z)}{(z - z_0)^{n+1}}\,\mathrm{d}z.$$

Proof. We have proven in exercise H 1.3 that power series are holomorphic within their radius of convergence and that the derivative has again a power series representation with the same radius of convergence. Hence by Theorem 1.20 and induction f is infinitely complex differentiable and every derivative is again complex differentiable in every $z_0 \in U$. Moreover, given the representation $f(z) = \sum_{k=0}^{\infty} a_k(z-z_0)^k$, again by H 1.3 it follows by induction that

$$f^{(n)}(z) = \sum_{k=n}^{\infty} \prod_{i=0}^{n-1} (k-i) a_k (z-z_0)^{k-n},$$

so that $f^{(n)}(z_0) = n! a_n$. Hence also the structure of the derivatives is a consequence of Theorem 1.20. □

Let us derive further consequences of Cauchy's integral formula. The first one is the converse statement to Goursat's lemma.

Theorem 1.22 (Morera's theorem)**.** *Let $U \subset \mathbb{C}$ be open and $f : U \to \mathbb{C}$ be continuous. If for each closed triangle $\Delta \subset U$ we have*

$$\int_{\partial\Delta} f \,\mathrm{d}z = 0,$$

then f is holomorphic on U.

Proof. Fix $z_0 \in U$. By Lemma 1.11 the function f has a primitive F on each ball $B_r(z_0) \subset U$ (note that balls are star-shaped). In particular, by Corollary 1.21 also $f = F'$ is holomorphic on $B_r(z_0)$. Since $z_0 \in U$ was arbitrary this concludes the proof. □

While Morera's theorem seems to be a rather complicated method to prove holomorphy of a given function it implies a strong property for sequences of holomorphic functions.

Theorem 1.23 (Weierstrass convergence theorem)**.** *Let $U \subset \mathbb{C}$ be open and $(f_n)_{n\in\mathbb{N}} : U \to \mathbb{C}$ be a sequence of holomorphic functions. Assume that there exists a function $f : U \to \mathbb{C}$ such that $f_n \to f$ locally uniformly, i.e., for each compact set $K \subset U$ it holds that*

$$\sup_{z\in K} |f_n(z) - f(z)| \to 0 \qquad \text{as } n \to +\infty.$$

Then f is holomorphic on U.

Remark 1.24. On $\mathbb{R}$ differentiability is not preserved under local uniform convergence. For instance, the sequence of smooth functions $f_n(x) = \sqrt{\frac{1}{n} + x^2}$ converges uniformly locally uniformly on $\mathbb{R}$ to the non-differentiable function $f(x) = |x|$.

Proof of Theorem 1.23. Let $\Delta \subset U$ be a closed triangle. Then due to Goursat's lemma (see Lemma 1.12) it holds that $\int_{\partial\Delta} f_n \,\mathrm{d}z = 0$ for all $n \in \mathbb{N}$. According to exercise H 2.3 we have

$$0 = \int_{\partial\Delta} f_n \,\mathrm{d}z \to \int_{\partial\Delta} f \,\mathrm{d}z.$$

Recall that the local uniform limit of continuous functions is always continuous (see Analysis 2), so that Morera's theorem yields that f is holomorphic. □

The next result shows that entire functions (i.e. holomorphic on $\mathbb{C}$) cannot be bounded except when they are constant.

Theorem 1.25 (Liouville's theorem)**.** *Let $f : \mathbb{C} \to \mathbb{C}$ be holomorphic and such that $\sup_{z\in\mathbb{C}} |f(z)| < +\infty$. Then f is constant.*

Proof. It follows from Theorem 1.20 that we can write

$$f(z) = \sum_{k=0}^{\infty} a_k z^k$$

for all $z \in \mathbb{C}$. We show that $a_k = 0$ for all $k \geq 1$, so that $f(z) = a_0$ for all $z \in \mathbb{C}$ as claimed. Using the formula from Theorem 1.20 for any $r > 0$ we deduce from the fundamental estimate that

$$|a_k| = \frac{1}{2\pi} \left| \int_{\partial B_r(0)} \frac{f(z)}{z^{k+1}} \,\mathrm{d}z \right| \leq \frac{1}{2\pi} \sup_{z\in\partial B_r(0)} \frac{|f(z)|}{|z|^{k+1}} L(\partial B_r(0)) \leq r^{-k} \sup_{z\in\mathbb{C}} |f(z)|.$$

By our assumption the supremum is finite, so that letting $r \to +\infty$ yields the claim. □

Recall that Cauchy's integral formula implies that a holomorphic function on a ball is determined by its values on the boundary. The next result is a far reaching generalization.

Theorem 1.26 (Identity theorem)**.** *Let $D \subset \mathbb{C}$ be a domain and $f, g : D \to \mathbb{C}$ be holomorphic functions. Then the following conditions are equivalent:*

(i) $f = g$ *on* D*;*
(ii) the set $\{z \in D : f(z) = g(z)\}$ *has an accumulation point*[3] *in* D*;*
(iii) there exists $z_0 \in D$ *such that* $f^{(n)}(z_0) = g^{(n)}(z_0)$ *for all* $n \in \mathbb{N} \cup \{0\}$*.*

Proof. (i) $\Longrightarrow$ (ii) is clear. Next we show that (ii) $\Longrightarrow$ (iii) by contradiction. Set $h = f - g$ and let $z_0 \in D$ be an accumulation point of the set $\{h = 0\}$. Assume that there exists a minimal $m \in \mathbb{N} \cup \{0\}$ such that $h^{(m)}(z_0) \neq 0$. According to Theorem 1.20 we can write h locally as a power series and, according to Corollary 1.21, the coefficients a_k satisfy $a_k = 0$ for all $0 \le k < m$ and $a_m \neq 0$. In particular, we can write

$$h(z) = \sum_{k=m} a_k (z - z_0)^k = (z - z_0)^m \underbrace{\sum_{k=0}^{\infty} a_{m+k} (z - z_0)^k}_{=: h_m(z)} .$$

Note that h_m is holomorphic in a neighborhood of z_0 and that $h_m(z_0) = a_m \neq 0$. By continuity there exists $\varepsilon > 0$ such that $h_m(z) \neq 0$ for all $z \in B_\varepsilon(z_0)$. Hence also $h(z) \neq 0$ for all $z \in B_\varepsilon(z_0) \setminus \{z_0\}$. This means that z_0 cannot be an accumulation point of $\{h = 0\}$, which yields a contradiction.
Finally we show that (iii) $\Longrightarrow$ (i). Here we have to use that D is a domain. Set again $h = f - g$. By the continuity of each derivative $h^{(k)}$ (see Corollary 1.21) the set $S_k = \{z \in D : h^{(k)} = 0\}$ is closed in D[4]. Hence also the intersection $S = \bigcap_{k=0}^{\infty} S_k$ is closed in D. We claim that S is also open in D. Indeed, if $z_0 \in S$ then the power series of h around z_0 equals zero, so that $h \equiv 0$ on a small ball $B_\varepsilon(z_0)$. Hence $B_\varepsilon(z_0) \subset S$ which proves that S is open. Since D is a domain this implies that $S = D$ since by assumption $S \neq \emptyset$ as $z_0 \in S$[5]. Then the power series representation of h equals 0 in every point $z \in D$. This concludes the proof. □

As a consequence of the identity theorem we deduce that zeros of non-constant holomorphic functions are isolated.

Corollary 1.27. *Let $D \subset \mathbb{C}$ be a domain and $f : D \to \mathbb{C}$ be holomorphic. Then either $f \equiv 0$ or*

$$f(z_0) = 0 \quad \Longrightarrow \quad \exists \varepsilon > 0 \textit{ such that } f(z) \neq 0 \quad \forall z \in B_\varepsilon(z_0) \setminus \{z_0\}.$$

Proof. Assume that f is not constantly 0 and that $f(z_0) = 0$ for some $z_0 \in D$. If for all $\varepsilon > 0$ there exists $z_\varepsilon \in B_\varepsilon(z_0) \setminus \{z_0\}$ such that $f(z_\varepsilon) = 0$, then z_0 is an accumulation point of $\{f = 0\}$ in D. This yields a contradiction to the identity theorem. □

We next prove that a holomorphic function that is not constant on a domain maps open sets to open sets. Note that this property is wrong for smooth real-valued functions. For instance, the function $x \mapsto \sin(x)$ maps the open interval $(0, 2\pi)$ to the closed set $[-1, 1]$. The property that open sets are mapped to open sets is very important from a topological point of view since it implies that the inverse function is also continuous (provided it exists). In the proof we will use the following auxiliary result that is interesting on its own.

Lemma 1.28. *Let $U \subset \mathbb{C}$ be open and $f : U \to \mathbb{C}$ be holomorphic. Assume that $\overline{B_r(z_0)} \subset U$ is a closed ball such that $\min_{z \in \partial B_r(z_0)} |f(z)| > |f(z_0)|$. Then f has a zero in $B_r(z_0)$.*

[3] $z_0 \in D$ is an accumulation point if there exists a sequence $(z_n)_{n \in \mathbb{N}} \subset \{z \in D : f(z) = g(z)\}$ such that $z_n \neq z_0$ for all $n \in \mathbb{N}$ and $z_n \to z_0$.

[4] A set S is closed in D if for all sequences $(z_n)_{n \in \mathbb{N}} \subset S$ with $z_n \to z \in D$ it holds that $z \in S$. The important point is that we assume that the limit z belongs to D. In that sense D is closed in D even when it is not closed in $\mathbb{C}$

[5] Path connectedness implies that the only subsets of D that are open and closed are D and the empty set. This is a general fact which you find in many analysis/topology text books. For open sets in $\mathbb{C}$ also the converse is true.

Proof. Assume that $f(z) \neq 0$ for all $z \in B_r(z_0)$. Then by assumption also $f(z) \neq 0$ for all $z \in \overline{B_r(z_0)}$. By continuity it follows that $f(z) \neq 0$ for all z in a slightly larger open set $V \subset U$ with $\overline{B_r(z_0)} \subset V$. Then the function $z \mapsto 1/f(z)$ is holomorphic on V and by the mean value property (see Corollary 1.19) we deduce that

$$\frac{1}{|f(z_0)|} \leq \sup_{z\in\partial B_r(z_0)} \frac{1}{|f(z)|} = (\min_{z\in\partial B_r(z_0)} |f(z)|)^{-1},$$

which contradicts the assumption. □

Theorem 1.29 (Open mapping theorem)**.** *Let $D \subset \mathbb{C}$ be a domain and $f : D \to \mathbb{C}$ be a non-constant, holomorphic function. Then $f(D)$ is again a domain.*

Proof. We first show that $f(D)$ is path-connected. Let $w_1, w_2 \in f(D)$. Then there exist $z_1, z_2 \in D$ such that $f(z_i) = w_i$ for $i = 1, 2$. Since D is path-connected there exists a continuous curve $\gamma : [0, 1] \to D$ such that $\gamma(0) = z_1$ and $\gamma(1) = z_2$. Then $f \circ \gamma : [0, 1] \to f(D)$ is continuous and connects w_1 with w_2. Hence $f(D)$ is path-connected. It remains to prove that $f(D)$ is open. Let $w_0 \in f(D)$ and $z_0 \in D$ be such that $f(z_0) = w_0$. Up to considering $f - f(z_0)$ we may assume that $w_0 = 0$. Since f is not constant Corollary 1.27 implies that there exists $r > 0$ such that $f(z) \neq 0$ for all $z \in \overline{B_r(z_0)} \setminus \{z_0\}$. Set $\varepsilon = \frac{1}{2} \min_{\zeta\in\partial B_r(z_0)} |f(\zeta)| > 0$ and fix $w \in B_\varepsilon(0)$. It suffices to show that there exists $z \in D$ such that $f(z) = w$. To see this, note that for all $z \in \partial B_r(z_0)$ we have

$$|f(z) - w| \geq \min_{\zeta\in\partial B_r(z_0)} |f(\zeta)| - |w| = 2\varepsilon - |w| > \varepsilon > |w| = |f(z_0) - w|.$$

Hence we can apply Lemma 1.28 to deduce that there exists a zero of $z \mapsto f(z) - w$ in $B_r(z_0)$. This proves the claim. □

The open mapping theorem can be extended to general open sets $U \subset \mathbb{C}$ under the assumption that f is not constant on any open subset $V \subset U$. The conclusion has to be adapted in the sense that $f(U)$ is an open set, but not necessarily path-connected.

From the open mapping theorem we deduce the maximum principle for holomorphic functions.

Theorem 1.30 (Maximum principle)**.** *Let $D \subset \mathbb{C}$ be a domain and $f : D \to \mathbb{C}$ be holomorphic. If $|f|$ attains its maximum on D, then f is constant.*

Proof. Assume that there exists $z_0 \in D$ such that $|f(z_0)| \geq |f(z)|$ for all $z \in D$. If f is not constant, then by the open mapping theorem there exists $\varepsilon > 0$ such that $B_\varepsilon(f(z_0)) \subset f(D)$. But the ball $B_\varepsilon(f(z_0))$ contains elements w with $|w| > |f(z_0)|$, which contradicts the maximality. Hence f is constant. □

Corollary 1.31. *Let $D \subset \mathbb{C}$ be a bounded domain and $f : \overline{D} \to \mathbb{C}$ be continuous and holomorphic on D. Then $|f|$ attains its maximum on ∂D.*

Proof. Since $\overline{D}$ is compact the continuous function $|f|$ attains its maximum in some $z_0 \in \overline{D}$. If $z_0 \in D$ then by the maximum principle f is constant, so the maximum is also attained on ∂D (it is attained in every point $z \in \overline{D}$). This concludes the proof. □

Similar to the open mapping theorem the maximum principle holds for general open sets with the conclusion that f is constant on the connected component containing the maximum of $|f|$. Moreover, also in the case of local maxima the maximum principle allows to conclude that f is constant on the connected component containing the local maximum (see exercise H 4.3).

1.4. **Isolated singularities and Laurent series.** So far we considered only holomorphic functions, that is to say very regular functions. In this section we analyze the behavior of holomorphic functions near isolated singularities as defined below.

Definition 1.32. Let $U \subset \mathbb{C}$ be open and $f : U \to \mathbb{C}$ be holomorphic. $z_0 \in \mathbb{C} \setminus U$ is called an isolated singularity of f if there exists $\varepsilon > 0$ such that $B_\varepsilon(z_0) \setminus \{z_0\} \subset U$.

We classify isolated singularities as follows:

Definition 1.33 (Different types of isolated singularities)**.** Let $U \subset \mathbb{C}$ be open and $f : U \to \mathbb{C}$ be holomorphic. An isolated singularity z_0 of f is called

(i) removable if there exists a holomorphic extension $\widetilde{f} : U \cup \{z_0\} \to \mathbb{C}$ of f;
(ii) a pole of f if there exists $m \in \mathbb{N}$ such that the function $z \mapsto (z - z_0)^m f(z)$ has a removable singularity in z_0. The smallest number m with this property is called the order of the pole;
(iii) an essential singularity if z_0 is neither a removable singularity nor a pole of f.

Here some examples of isolated singularities:

a) $z \mapsto \frac{\sin(z)}{z}$ has a removable singularity in $z = 0$. This can be seen using the power series representation of $\sin(z)$ which is of the form $z - \frac{1}{6}z^3 + \ldots$, so that the denominator is canceled.
b) $z \mapsto (z - z_0)^{-m}$ has a pole of order m in $z_0 \in \mathbb{C}$.
c) $z \mapsto e^{\frac{1}{z}}$ has an essential singularity in $z = 0$ (see exercise H 4.5).

The next result characterizes removable singularities.

Lemma 1.34 (Riemann extension lemma)**.** *Let $U \subset \mathbb{C}$ be open and $f : U \to \mathbb{C}$ be holomorphic. Assume that $z_0 \in \mathbb{C} \setminus U$ is an isolated singularity of f. Then the following conditions are equivalent:*

(i) z_0 is a removable singularity of f;
(ii) f can be extended to a continuous function $\widetilde{f} : U \cup \{z_0\} \to \mathbb{C}$;
(iii) f is bounded in $B_\varepsilon(z_0) \setminus \{z_0\}$ for some $\varepsilon > 0$;
(iv) $\lim\limits_{z \to z_0} (z - z_0) f(z) = 0$.

Proof. See exercise H 3.3. □

Next we want to derive a generalized series representation around isolated singularities. For the sake of generality we consider not only isolated singularities but functions that are holomorphic on an annulus

$$A_{r,R}(z_0) = \{z \in \mathbb{C} : \ r < |z - z_0| < R\}.$$

Isolated singularities form a special case setting $r = 0$. We first prove a generalization of Corollary 1.14.

Lemma 1.35. *Let $f : A_{r,R}(z_0) \to \mathbb{C}$ be holomorphic. Then for every $r < s < \rho < R$ we have*

$$\int_{\partial B_s(z_0)} f \,\mathrm{d}z = \int_{\partial B_\rho(z_0)} f \,\mathrm{d}z.$$

Remark 1.36. This statement is an easy consequence of the second version of Cauchy's integral theorem since the two boundaries are freely homotopic in $A_{r,R}(z_0)$. However, we give a different proof not based on that theorem.

Proof of Lemma 1.35. Given any holomorphic function $g : A_{r,R}(z_0) \to \mathbb{C}$ define the function $J : (r, R) \to \mathbb{C}$ by

$$J(x) = \int_{\partial B_x(z_0)} \frac{g(\zeta)}{\zeta - z_0} \,\mathrm{d}\zeta = i \int_0^{2\pi} g(z_0 + xe^{it}) \,\mathrm{d}t.$$

Since g is smooth on $A_{r,R}(z_0)$ it follows that J is differentiable and

$$J'(x) = i \int_0^{2\pi} \frac{\mathrm{d}}{\mathrm{d}x} g(z_0 + xe^{it}) \,\mathrm{d}t = i \int_0^{2\pi} g'(z_0 + xe^{it}) e^{it} \,\mathrm{d}t = \frac{1}{x}(g(z_0 + x) - g(z_0 + x)) = 0.$$

Thus J is constant on (r, R). Taking $g(z) = f(z)(z - z_0)$ we conclude the proof. □

As for the power series representation of holomorphic functions we need a suitable version of Cauchy's integral formula on annuli.

Proposition 1.37 (Cauchy's integral formula for annuli)**.** *Let $U \subset \mathbb{C}$ be open and $f : U \to \mathbb{C}$ be holomorphic. Assume that $\overline{A_{r,R}(z_0)} \subset U$. Then for any $z \in A_{r,R}(z_0)$ it holds that*

$$f(z) = \frac{1}{2\pi i} \int_{\partial B_R(z_0)} \frac{f(\zeta)}{\zeta - z} \,\mathrm{d}\zeta - \frac{1}{2\pi i} \int_{\partial B_r(z_0)} \frac{f(\zeta)}{\zeta - z} \,\mathrm{d}\zeta.$$

Proof. Fix $z \in A_{r,R}(z_0)$ and define the function $g : U \to \mathbb{C}$ as

$$g(\zeta) = \begin{cases} \frac{f(\zeta)-f(z)}{\zeta - z} & \text{if } \zeta \in U \setminus \{z\}, \\ f'(z) & \text{if } \zeta = z. \end{cases}$$

Since f is complex differentiable in z it follows from Riemann's extension lemma that g is holomorphic

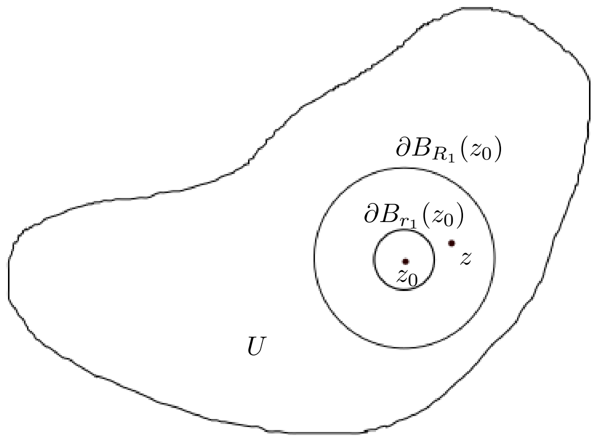

FIGURE 3. The geometric situation of Proposition 1.37. Around z_0 the set U has a hole.

on U. Since $\overline{A_{r,R}(z_0)} \in U$ there exist $r_1 < r$ and $R_1 > R$ such that $A_{r_1,R_1}(z_0) \subset U$ (see also Figure 3). Hence Lemma 1.35 implies that

$$\int_{\partial B_R(z_0)} g \, \mathrm{d}z = \int_{\partial B_r(z_0)} g \, \mathrm{d}z.$$

Inserting the definition of g we obtain that

$$\int_{\partial B_R(z_0)} \frac{f(\zeta)}{\zeta - z} \, \mathrm{d}\zeta - \int_{\partial B_r(z_0)} \frac{f(\zeta)}{\zeta - z} \, \mathrm{d}\zeta = \int_{\partial B_R(z_0)} \frac{f(z)}{\zeta - z} \, \mathrm{d}\zeta - \underbrace{\int_{\partial B_r(z_0)} \frac{f(z)}{\zeta - z} \, \mathrm{d}\zeta}_{=0 \text{ since there exists a primitive}}$$

$$= f(z) \underbrace{\int_{\partial B_R(z_0)} (\zeta - z)^{-1} \, \mathrm{d}z}_{=2\pi i \text{ (apply Thm. 1.18 to } f=1)} = 2\pi i f(z).$$

Rearranging terms yields the claim. □

Now we can derive the so-called Laurent-series representation on annuli.

Theorem 1.38 (Laurent series expansion)**.** *Let $U \subset \mathbb{C}$ be open and $f : U \to \mathbb{C}$ be holomorphic. Assume that $A_{r,R}(z_0) \subset U$. Then for every $z \in A_{r,R}(z_0)$ we can write*

$$f(z) = \sum_{n \in \mathbb{Z}} c_n (z - z_0)^n,$$

where the coefficients are given by

$$c_n = \frac{1}{2\pi i} \int_{\partial B_s(z_0)} \frac{f(z)}{(z - z_0)^{n+1}} \, \mathrm{d}z,$$

which is independent of $s \in (r, R)$.

Proof. Fix $z \in A_{r,R}(z_0)$. Then there exist $r < r_1 < R_1 < R$ such that $z \in A_{r_1,R_1}(z_0)$ and $\overline{A_{r_1,R_1}(z_0)} \subset U$. Therefore Proposition 1.37 yields that

$$f(z) = \frac{1}{2\pi i}\int_{\partial B_{R_1}(z_0)} \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}\zeta - \frac{1}{2\pi i}\int_{\partial B_{r_1}(z_0)} \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}\zeta. \tag{1}$$

Similar to the proof of Theorem 1.18 we have that

$$\begin{aligned}\frac{1}{2\pi i}\int_{\partial B_{R_1}(z_0)} \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}\zeta &= \frac{1}{2\pi i}\int_{\partial B_{R_1}(z_0)} \frac{f(\zeta)}{\zeta - z_0}\frac{1}{1-\frac{z-z_0}{\zeta - z_0}}\,\mathrm{d}\zeta \overset{\text{geometric series}}{=} \frac{1}{2\pi i}\int_{\partial B_{R_1}(z_0)} \sum_{n=0}^{\infty}\frac{f(\zeta)(z-z_0)^n}{(\zeta - z_0)^{n+1}}\,\mathrm{d}\zeta \\ &\overset{\text{uniform convergence}}{=} \sum_{n=0}^{\infty}\frac{1}{2\pi i}\int_{\partial B_{R_1}(z_0)} \frac{f(\zeta)}{(\zeta - z_0)^{n+1}}\,\mathrm{d}\zeta (z-z_0)^n.\end{aligned}$$

Writing

$$-\frac{1}{\zeta - z} = \frac{1}{z-\zeta} = \frac{1}{z - z_0}\frac{1}{1-\frac{\zeta - z_0}{z-z_0}} = \sum_{n=0}^{\infty}\frac{(\zeta - z_0)^n}{(z-z_0)^{n+1}} = \sum_{n=-1}^{-\infty}\frac{(z-z_0)^n}{(\zeta - z_0)^{n+1}}$$

for $\zeta \in \partial B_{r_1}(z_0)$ we conclude by the same reasoning as above that

$$-\frac{1}{2\pi i}\int_{\partial B_{r_1}(z_0)} \frac{f(\zeta)}{\zeta - z}\,\mathrm{d}\zeta = \sum_{n=-1}^{-\infty}\frac{1}{2\pi i}\int_{\partial B_{R_1}(z_0)} \frac{f(\zeta)}{(\zeta - z_0)^{n+1}}\,\mathrm{d}\zeta (z-z_0)^n.$$

Moreover, since $r_1, R_1 \in (r, R)$ we can apply Lemma 1.35 to infer that for any $s \in (r, R)$

$$\frac{1}{2\pi i}\int_{\partial B_{R_1}(z_0)} \frac{f(\zeta)}{(\zeta - z_0)^{n+1}}\,\mathrm{d}\zeta = \frac{1}{2\pi i}\int_{\partial B_{r_1}(z_0)} \frac{f(\zeta)}{(\zeta - z_0)^{n+1}}\,\mathrm{d}\zeta = \frac{1}{2\pi i}\int_{\partial B_{s}(z_0)} \frac{f(\zeta)}{(\zeta - z_0)^{n+1}}\,\mathrm{d}\zeta = c_n.$$

Hence the claim follows from (1). □

Having derived the local Laurent series representation of holomorphic functions let us study such series a little bit more in detail. Given a series of the form

$$\sum_{n\in\mathbb{Z}} c_n(z-z_0)^n = \underbrace{\sum_{n=-\infty}^{-1} c_n(z-z_0)^n}_{=:h(z)} + \underbrace{\sum_{n=0}^{\infty} c_n(z-z_0)^n}_{=:g(z)}$$

we call it a Laurent series and h the principal part of the Laurent series and g the regular part of the Laurent series. We call a Laurent series (uniformly, pointwise, etc.) convergent if both principal and regular part converge (uniformly, pointwise, etc.). In the lemma below we collect some properties of the principal part.

Lemma 1.39. *Let $z_0 \in \mathbb{C}$ and $(c_{-k})_{k=1}^{\infty} \subset \mathbb{C}$. Define*

$$r = \limsup_{k\to+\infty} \sqrt[k]{|c_{-k}|} \in [0, +\infty].$$

Then the series

$$h(z) = \sum_{k=1}^{\infty} c_{-k}(z-z_0)^{-k}$$

converges for all $z \in A_{r,+\infty}(z_0) := \{z \in \mathbb{C} : |z - z_0| > r\}$ and diverges for $z \in B_r(z_0)$. Moreover, it converges uniformly on $A_{\rho,+\infty}(z_0)$ for every $\rho > r$. In particular, it is holomorphic on $A_{r,+\infty}(z_0)$ and $\lim_{|z|\to+\infty} h(z) = 0$.

Proof. It suffices to note that h is a power series in the variable $\zeta = (z - z_0)^{-1}$, r is the inverse of the convergence radius of this power series and $|\zeta| < \frac{1}{r}$ if and only if $|z - z_0| > r$ with the convention that $1/0 = +\infty$ and $1/+\infty = 0$. The claim then follows from the corresponding statements for power series, noting that $|z| \to +\infty$ if and only if $\zeta \to 0$. □

From the previous lemma and the fact that a power series converges on $B_R(z_0)$, where R denotes its radius of convergence while it diverges outside $\overline{B_R(z_0)}$, it follows that the natural domain of Laurent series is given by the annulus $A_{r,R}(z_0)$. Next we prove that the coefficients of the Laurent series representation of a holomorphic function are unique.

Corollary 1.40. *Assume that $f : A_{r,R}(z_0) \to \mathbb{C}$ is holomorphic and given by a Laurent series of the form*

$$f(z) = \sum_{n\in\mathbb{Z}} c_n (z-z_0)^n.$$

Then c_n is given by

$$c_n = \frac{1}{2\pi i}\int_{\partial B_s(z_0)} \frac{f(z)}{(z-z_0)^{n+1}}\,\mathrm{d}z,$$

which is independent of $s \in (r, R)$.

Proof. Due to Lemma 1.39 and the corresponding property for power series we know that the Laurent series representing f converges uniformly on every annulus $A_{r_r,R_1}(z_0)$ with $r < r_1 < R_1 < R$. Hence for $s \in (r, R)$ we may exchange integration and summation, so that

$$\int_{\partial B_s(z_0)} \frac{f(z)}{(z-z_0)^{n+1}}\,\mathrm{d}z = \int_{\partial B_s(z_0)} \sum_{k\in\mathbb{Z}} c_k (z-z_0)^{k-n-1}\,\mathrm{d}z = \sum_{k\in\mathbb{Z}} c_k \int_{\partial B_s(z_0)} (z-z_0)^{k-n-1}\,\mathrm{d}z.$$

For $k-n-1 \neq -1$ the integrand has a primitive on $\mathbb{C}\setminus\{z_0\}$, so that the integral vanishes. For $k = n$ the integral equals $2\pi i$ (see also Example 1.8), so that the right hand side equals $2\pi i c_n$ which proves the claim. □

Next we study the connection of the coefficients $(c_n)_{n\in\mathbb{N}}$ with the type of singularities when f is defined on $A_{0,R}(z_0)$ for some $R > 0$. The following proposition allows to easily determine the type of singularities of many functions.

Proposition 1.41. *Let $U \subset \mathbb{C}$ be open and $f : U \to \mathbb{C}$ be holomorphic. Assume that $z_0 \in \mathbb{C}\setminus U$ is an isolated singularity of f and let $(c_n)_{n\in\mathbb{Z}}$ be the coefficients of the Laurent series expansion of f on $A_{0,R}(z_0)$ ($R > 0$ small enough such that $A_{0,R}(z_0) \subset U$). Then z_0 is a*

(i) removable singularity if and only if $c_k = 0$ for all $k \leq -1$;
(ii) a pole of order $m \in \mathbb{N}$ if and only if $c_{-m} \neq 0$ and $c_k = 0$ for all $k < -m$.
(iii) an essential singularity if and only if $c_k \neq 0$ for infinitely many $k \leq -1$.

Proof. (i) If $c_k = 0$ for $k \leq -1$ then the Laurent series reduces to a power series with a positive radius of convergence. Hence z_0 is a removable singularity by exercise H 1.3. For the converse statement note that, denoting the holomorphic extension still by f, we have for $k \leq -1$ and $s \in (0, r)$ that

$$c_k = \frac{1}{2\pi i}\int_{\partial B_s(z_0)} \underbrace{\frac{f(z)}{(z-z_0)^{k+1}}}_{\text{holomorphic on } B_r(z_0)}\,\mathrm{d}z = 0,$$

where we used Cauchy's integral theorem on the star-shaped set $B_r(z_0)$.

(ii) If z_0 is a pole of order $m \in \mathbb{N}$, then the Laurent series of $(z-z_0)^m f(z)$ reads

$$(z-z_0)^m f(z) = \sum_{k\in\mathbb{Z}} c_k (z-z_0)^{k+m} = \sum_{j\in\mathbb{Z}} c_{j-m}(z-z_0)^j.$$

Applying (i) it follows that $c_{j-m} = 0$ for all $j \leq -1$ or equivalently $c_k = 0$ for $k < -m$. Moreover, since m is minimal it follows again from (i) that $c_{j-m+1} \neq 0$ for some $j \leq -1$ or equivalently $c_k \neq 0$ for some $k \leq -m$. Hence the only possibility is $c_{-m} \neq 0$. On the other hand, if $c_{-m} \neq 0$ and $c_k = 0$ for all $k < -m$, then the Laurent series of $(z-z_0)^m f(z)$ reads

$$(z-z_0)^m f(z) = \sum_{k=-m}^{\infty} c_k (z-z_0)^{k+m} = \sum_{j=0}^{\infty} c_{j-m}(z-z_0)^j,$$

so that by (i) z_0 is a removable singularity of $(z-z_0)^m f(z)$. Moreover, for $\ell < m$ it follows by the same argument that the Laurent series representation of $(z-z_0)^\ell f(z)$ contains a non-vanishing principal part, so that by (i) z_0 cannot be a removable singularity of $(z-z_0)^\ell f(z)$. Thus z_0 is a pole of order m.

(iii) z_0 is an essential singularity if and only if neither (i) nor (ii) are true. This is the case if and only if $c_k \neq 0$ for infinitely many $k \leq -1$. □

Using the previous result it follows easily that $\sin(1/z)$, $\cos(1/z)$ or $e^{1/z}$ have an essential singularity in the origin.

Remark 1.42. The behavior near an essential singularity is quite chaotic. Indeed, Picard's great theorem says that if f has an essential singularity in z_0 then there exists at most one value $w_0 \in \mathbb{C}$ that for each $\varepsilon > 0$ the function $f : B_\varepsilon(z_0) \setminus \{z_0\} \to \mathbb{C}$ is surjective onto $\mathbb{C} \setminus \{w_0\}$, i.e. f omits at most one value no matter how close to the singularity.

1.5. **The residue theorem and applications.** In this final subsection on complex analysis we will see how the theory on the Laurent series representation allows to evaluate some integrals quite easily.

Definition 1.43 (Winding number of a curve)**.** Let $\gamma : [a,b] \to \mathbb{C}$ be a closed curve (as usual we assume that it is piecewise C^1). Let $c \in \mathbb{C}$ be such that $c \neq \gamma(t)$ for all $t \in [a,b]$. The winding number of γ around c is defined by

$$\nu(\gamma, c) := \frac{1}{2\pi i} \int_\gamma (z-c)^{-1}\, \mathrm{d}z.$$

The winding number describes how often the curve γ winds around c counterclockwise (it counts clockwise winding with a minus sign)[6]. We first establish some elementary properties of the winding number (not relying on the previous footnote, but try to first convince yourself about the statements with the informal definition of the winding number).

Lemma 1.44. *Let $\gamma : [a,b] \to \mathbb{C}$ be a closed curve. Set $\Gamma = \gamma([a,b])$ as the trace of γ.*

(i) $\nu(\gamma, c) \in \mathbb{Z}$ for all $c \in \mathbb{C} \setminus \Gamma$;
(ii) If $D \subset \mathbb{C}$ is a domain with $D \cap \Gamma = \emptyset$, then $\nu(\gamma, c)$ is independent of $c \in D$;
(iii) If $c \in \mathbb{C}$ is such that $\lambda c \notin \Gamma$ for all $\lambda \geq 1$, then $\nu(\gamma, c) = 0$. In particular, $\lim_{|z| \to +\infty} \nu(\gamma, z) = 0$.

Proof. (i) Define the functions $f, F : [a,b] \to \mathbb{C}$ by

$$f(t) = \int_a^t \frac{\gamma'(s)}{\gamma(s) - c}\, \mathrm{d}s, \qquad F(t) = e^{-f(t)}(\gamma(t) - c).$$

Then f and F are piecewise C^1-functions with $f(a) = 0$ and $f(b) = 2\pi i \nu(\gamma, c)$. Computing the derivative of F in every point where f and γ are differentiable we obtain that

$$F'(t) = -e^{-f(t)} f'(t)(\gamma(t) - c) + e^{-f(t)}\gamma'(t) = -e^{-f(t)} \frac{\gamma'(t)}{\gamma(t) - c}(\gamma(t) - c) + e^{-f(t)}\gamma'(t) = 0.$$

Since F is continuous and piecewise continuously differentiable this implies that F is constant as it cannot jump at the points where it is not differentiable. Since $F(t) \neq 0$ for all $t \in [a,b]$ this constant is different from zero. Since $\gamma(a) = \gamma(b)$ this implies that

$$1 = \frac{F(b)}{F(a)} = e^{f(a) - f(b)} = e^{-2\pi i \nu(\gamma, c)}.$$

[6]Here is a helpful argument using the following non-trivial result. Let $\gamma : [a,b] \to \mathbb{C}$ be a piecewise C^1-curve such that $c \notin \gamma([a,b])$. Then there exist piecewise C^1-polar coordinates r, θ such that $\gamma(t) = c + r(t)e^{i\theta(t)}$. The winding number should be defined by the increment of the angle along the curve, i.e., $\nu(\gamma, c) = \frac{\theta(b) - \theta(a)}{2\pi}$. Inserting the polar coordinate representation of γ we find indeed that

$$\int_\gamma (z-c)^{-1}\, \mathrm{d}z = \int_a^b \frac{r'(t)e^{i\theta(t)} + r(t) i e^{i\theta(t)}\theta'(t)}{r(t)e^{i\theta(t)}}\, \mathrm{d}t = \int_a^b \frac{r'(t)}{r(t)} + i\theta'(t)\, \mathrm{d}t = \underbrace{\log(r(b)) - \log(r(b))}_{=0 \text{ since } \gamma \text{ is closed}} + i(\theta(b) - \theta(a)).$$

This is equivalent to $\nu(\gamma, c) \in \mathbb{Z}$.

(ii) The function $z \mapsto \nu(\gamma, z)$ is continuous on $\mathbb{C} \setminus \Gamma$. Since $D \subset \mathbb{C} \setminus \Gamma$ is path-connected, the image $\nu(\gamma, D)$ has to be path-connected, too. Since $\mathbb{Z}$ is discrete the only path-connected sets of $\mathbb{Z}$ are points. Hence $\nu(\gamma, \cdot)$ is constant on D.

(iii) Note that the set $\mathbb{C} \setminus \{\lambda c : \lambda \geq 1\}$ is star-shaped. Hence there exists a primitive of $z \mapsto (z - c)^{-1}$ on this set. Thus the integral defining the winding number vanishes since γ is closed. The second claim follows from the fact that Γ is a bounded set, so that for $|z|$ large enough $\lambda z \notin \Gamma$ for all $\lambda \geq 1$. □

The winding number appears naturally in the integration of Laurent series. Here also the importance of the coefficient c_{-1} (which is called residue) becomes evident.

Lemma 1.45. *Let $f : A_{r,R}(z_0) \to \mathbb{C}$ be holomorphic with Laurent series representation*

$$f(z) = \sum_{k \in \mathbb{Z}} c_k (z - z_0)^k.$$

Let $\gamma : [a, b] \to A_{r,R}(z_0)$ be a closed, piecewise C^1-curve. Then

$$\int_\gamma f \, \mathrm{d}z = 2\pi i \nu(\gamma, z_0) c_{-1}.$$

Proof. According to Lemma 1.39 and the general properties of power series the Laurent series converges uniformly on $A_{r_1,R_1}(z_0)$ for every $r < r_1 < R_1 < R$. We chose r_1 and R_1 such that

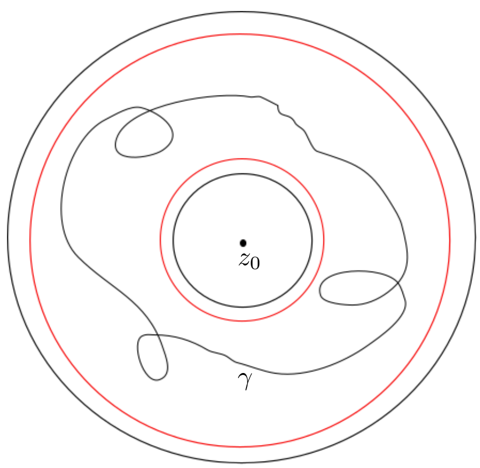

FIGURE 4. Black: the annulus $A_{r,R}(z_0)$; red: the annulus $A_{r_1,R_1}(z_0)$.

$$r < r_1 < \min_{t \in [a,b]} |\gamma(t) - z_0| \leq \max_{t \in [a,b]} |\gamma(t) - z_0| < R_1 < R.$$

See also Figure 4. Since $\gamma([a, b]) \subset A_{r_1,R_1}(z_0)$ the uniform convergence of the Laurent series allows to interchange summation and integration and we obtain

$$\int_\gamma f \, \mathrm{d}z = \int_\gamma \sum_{k \in \mathbb{Z}} c_k (z - z_0)^k \, \mathrm{d}z = \sum_{k \in \mathbb{Z} \setminus \{-1\}} c_k \underbrace{\int_\gamma (z - z_0)^k \, \mathrm{d}z}_{=0 \text{ (primitive!)}} + c_{-1} \int_\gamma (z - z_0)^{-1} \, \mathrm{d}z = 2\pi i c_{-1} \nu(\gamma, z_0).$$

□

Definition 1.46. Let $f : A_{r,R}(z_0) \to \mathbb{C}$ be holomorphic with Laurent series representation

$$f(z) = \sum_{k \in \mathbb{Z}} c_k (z - z_0)^k.$$

Then the residue of f in z_0 is defined by $\operatorname{Res}(f, z_0) = c_{-1}$. Note that by Theorem 1.38

$$\operatorname{Res}(f, z_0) = \frac{1}{2\pi i} \int_{\partial B_s(z_0)} f \, \mathrm{d}z$$

for any $s \in (r, R)$. You will see further properties of the residue in the exercises.

The next result is one of the main theorems in complex analysis.

Theorem 1.47 (The residue theorem)**.** *Let $U \subset \mathbb{C}$ be a star-shaped domain and $S \subset U$ be a finite set. Assume that $f : U \setminus S \to \mathbb{C}$ be holomorphic and let $\gamma : [a, b] \to U \setminus S$ be a closed, piecewise C^1-curve. Then*

$$\int_\gamma f \, \mathrm{d}z = 2\pi i \sum_{c \in S} \operatorname{Res}(f, c) \, \nu(\gamma, c).$$

Remark 1.48. The residue theorem holds under the sole assumption that γ is null-homotopic in U and U is open. This will be clear from the proof since we could use the second version of Cauchy's integral theorem instead of Theorem 1.13.

Proof of Theorem 1.47. f is not necessarily holomorphic on U since the singularities in $s \in S$ might be non-removable. The idea is to subtract all principal parts so that the modified function has only removable singularities. Then the relevant contributions to the curve integral come from these modifications for which we can apply Lemma 1.45. Here come the details. For $s \in S$ let $f_s = \sum_{k=-1}^{-\infty} c_k (z - s)^k$ be the principal part of the Laurent series expansion of f around s (thus the coefficients depend on the point s). According to Lemma 1.39 the function f_s is holomorphic on $\mathbb{C} \setminus \{s\}$. Define the holomorphic function $g : U \setminus S \to \mathbb{C}$ by

$$g(z) = f(z) - \sum_{s \in S} f_s(z).$$

Then by construction each $s \in S$ is a removable singularity since locally around any fixed $s_0 \in S$ we have

$$g(z) = \sum_{k \in \mathbb{Z}} c_k (z - s_0)^k - \underbrace{\sum_{k=-1}^{-\infty} c_k (z - s_0)^k}_{= f_{s_0}(z)} - \sum_{s \in S \setminus \{s_0\}} f_s(z) = \sum_{k=0}^{\infty} c_k (z - s_0)^k - \underbrace{\sum_{s \in S \setminus \{s_0\}} f_s(z)}_{\substack{\text{holomorphic in a} \\ \text{neighborhood of } s_0}} .$$

Both functions on the right hand side can be extended holomorphically to s_0. Call the holomorphic extension of g to U still g. Then due to Cauchy's integral theorem (combine Theorem 1.13 with Lemma 1.7) it follows that

$$0 = \int_\gamma g \, \mathrm{d}z = \int_\gamma f \, \mathrm{d}z - \sum_{c \in S} \int_\gamma f_c \, \mathrm{d}z \overset{\text{Lemma 1.45}}{=} \int_\gamma f \, \mathrm{d}z - 2\pi i \sum_{c \in S} \operatorname{Res}(f, c) \nu(\gamma, c),$$

where Lemma 1.45 can be applied since f_s is holomorphic on $A_{0,+\infty}(s)$ which contains $\gamma([a, b])$. This proves the claim. □

The residue theorem has several interesting applications which we will now discuss as a final part of our introduction to complex analysis.

Corollary 1.49 (Integration of rational functions of sin and cos)**.** *Let $R : \mathbb{R}^2 \to \mathbb{C}$ be a rational function in two variables that is finite on $\partial B_1(0)$. Then*

$$\int_0^{2\pi} R(\cos(t), \sin(t)) \, \mathrm{d}t = 2\pi \sum_{z \in B_1(0)} \operatorname{Res}(\widetilde{R}, z),$$

where

$$\widetilde{R}(z) = \frac{1}{z} R\Big(\tfrac{1}{2} \left(z + \tfrac{1}{z} \right), \tfrac{1}{2i} \left(z - \tfrac{1}{z} \right) \Big).$$

Proof. See exercises including some examples. □

The next result provides a powerful tool to evaluate some improper real integrals where the primitive might not be known explicitly. We shall assume that the improper integrals exist. Conditions for this assumption are taught in the first analysis courses.

Corollary 1.50 (Evaluation of improper integrals)**.** *Let $S \subset \mathbb{C} \setminus \mathbb{R}$ be a finite set and $f : \mathbb{C} \setminus S \to \mathbb{C}$ be holomorphic. Assume that the integral $\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x$ exists and that $\lim\limits_{|z|\to+\infty} zf(z) = 0$. Then*

$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 2\pi i \sum_{s\in S\cap\{\mathrm{Im}(z)>0\}} \mathrm{Res}(f,s).$$

Proof. In what follows we apply a general strategy to evaluate integrals via the residue theorem. It consists of integrating over certain curves in $\mathbb{C}$ and showing that some contributions of the integral (usually those in $\mathbb{C}\setminus\mathbb{R}$) become negligible when some parameter varies. The non-vanishing part is then characterized by the residue theorem. Here we use growing half circles. For $R \gg 1$ define the closed curve $\gamma_R : [0,2] \to \mathbb{C}$ by

$$\gamma_R(t) = \begin{cases} -R(1-t)+tR & \text{if } t\in[0,1], \\ Re^{i\pi(t-1)} & \text{if } t\in(1,2]. \end{cases}$$

See also Figure 5. Let us write $\gamma_R = \gamma_{R,1}+\gamma_{R,2}$, where $\gamma_{R,1}$ denote straight line from $-R$ to R and $\gamma_{R,2}$ the

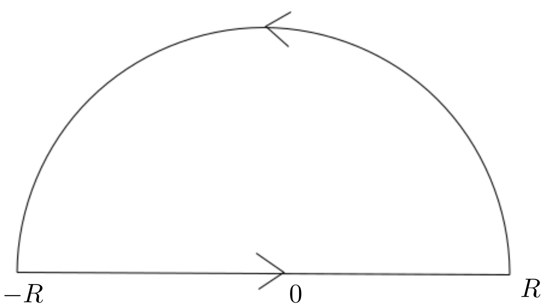

Figure 5. The curve γ_R representing a half-circle with radius R around the origin.

arc of circle on $\partial B_R(0)$ from R to $-R$. Note that for all $s\in S\cap B_R(0)\cap\{\mathrm{Im}(z)>0\}$ we have $\nu(\gamma_R,s)=1$. Moreover, since S is finite it follows that for R large enough we have $S\cap\{\mathrm{Im}(z)>0\}\subset B_R(0)$, so that from the residue theorem we deduce that

$$2\pi i \sum_{s\in S\cap\{\mathrm{Im(z)}>0\}} \mathrm{Res}(f,s) = \int_{-R}^{R} f(x)\,\mathrm{d}x + \int_{\gamma_{R,2}} f\,\mathrm{d}z.$$

We argue that the curve integral over $\gamma_{R,2}$ vanishes when $R\to+\infty$. Indeed, by the fundamental estimate we have

$$\left|\int_{\gamma_{R,2}} f\,\mathrm{d}z\right| \le L(\gamma_{R,2}) \sup_{z\in\gamma_{R,2}} |f(z)| \le \pi R \sup_{|z|=R} |f(z)| = \pi \sup_{|z|=R} |zf(z)| \to 0 \quad \text{as } R\to+\infty,$$

where we used the assumption on $zf(z)$. The other integral converges to the improper integral. Hence we proved the claim. $\square$

As a final application we discuss Fourier transformations in one dimension. At this point of the course it is not important to know about the Fourier transformation. We will discuss it more in detail later.

Corollary 1.51 (Fourier transformation for certain functions in 1D)**.** *Let $S \subset \mathbb{C} \setminus \mathbb{R}$ be a finite set and $f : \mathbb{C} \setminus S \to \mathbb{C}$ be holomorphic. Assume that* $\lim_{|z| \to +\infty} f(z) = 0$. *Then*

$$\int_{-\infty}^{\infty} f(x) e^{i\omega x} \, \mathrm{d}x = \begin{cases} 2\pi i \displaystyle\sum_{s \in S \cap \{\mathrm{Im}(z) > 0\}} \mathrm{Res}(f(z)e^{i\omega z}, s) & \text{if } \omega > 0, \\ -2\pi i \displaystyle\sum_{s \in S \cap \{\mathrm{Im}(z) < 0\}} \mathrm{Res}(f(z)e^{i\omega z}, s) & \text{if } \omega < 0. \end{cases}$$

If $\omega = 0$ the improper integral might not be defined.

Proof. Similar to the previous proof we integrate over a certain growing curve and apply the residue theorem. For the moment assume that $\omega > 0$. Consider the square Q and the curves $\gamma_1, \gamma_2, \gamma_3$ in Figure

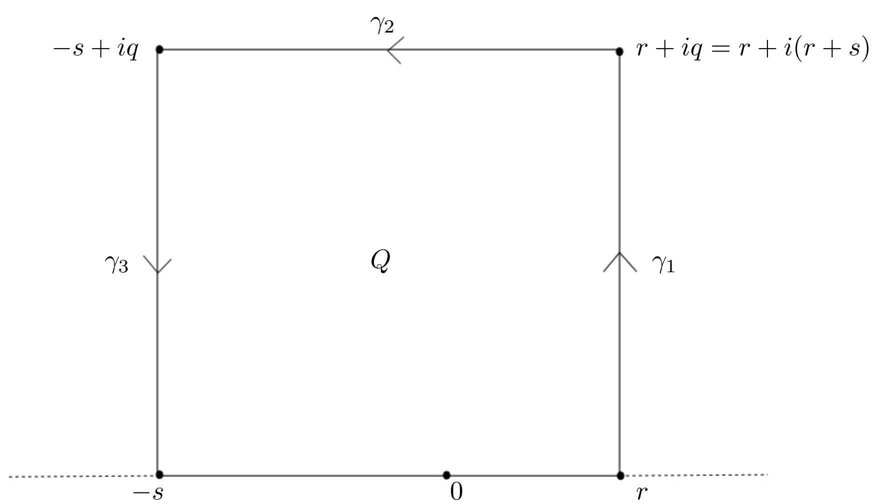

FIGURE 6. The square used to apply the residue theorem.

6 with $r, s > 0$ and $q = r + s$. (Since we do not assume here that the improper integrals exist we need the two parameters r, s to show its existence.) Note that the curves γ_i depend on r and s. Moreover, for r, s large enough it holds that $S \cap \{\mathrm{Im}(z) > 0\} \subset Q$. Orienting ∂Q counterclockwise we have by the residue theorem that

$$2\pi i \sum_{s \in S \cap \{\mathrm{Im}(z) > 0\}} \mathrm{Res}(f(z)e^{i\omega z}, s) = \int_{\partial Q} f(z) e^{i\omega z} \, \mathrm{d}z = \int_{-s}^{r} f(x) e^{i\omega x} \, \mathrm{d}x + \sum_{i=1}^{3} \int_{\gamma_i} f(z) e^{i\omega z} \, \mathrm{d}z. \tag{2}$$

We show that for $i = 1, 2, 3$ it holds that

$$\lim_{r,s \to +\infty} \int_{\gamma_i} f(z) e^{i\omega z} \, \mathrm{d}z = 0,$$

which then proves that the improper integral exists and equals the claimed value on the left hand side in (2). We start with the curve γ_2. Note for $z \in \gamma_2$ it holds that $|e^{i\omega z}| = e^{-\omega \mathrm{Im}(z)} = e^{-\omega q}$ since the imaginary part equals q on the whole curve γ_2. The fundamental estimate then implies

$$\left| \int_{\gamma_2} f(z) e^{i\omega z} \, \mathrm{d}z \right| \le \sup_{z \in \gamma_2} |f(z)| e^{-\omega q} L(\gamma_2) = \sup_{z \in \gamma_2} |f(z)| e^{-\omega q} q \le \sup_{z \in \gamma_2} |f(z)|, \tag{3}$$

where we used that $e^{-\omega q}q \leq 1$ for $q = r+s$ large enough. Next note that we can parameterize γ_1 via $\gamma_1(t) = r + it$ with $t \in [0,q]$. The fundamental estimate does help here, so we use a stronger bound via

$$\left|\int_{\gamma_1} f(z)e^{i\omega z}\,\mathrm{d}z\right| = \left|\int_0^q f(r+it)e^{i\omega(r+it)}i\,\mathrm{d}t\right| \leq \int_0^q |f(r+it)|e^{-\omega t}\,\mathrm{d}t \leq \sup_{z\in\gamma_1}|f(z)|\int_0^q e^{-\omega t}\,\mathrm{d}t$$
$$= \sup_{z\in\gamma_1}|f(z)|\frac{1}{\omega}(1-e^{-q\omega}) \leq \sup_{z\in\gamma_1}|f(z)|\frac{1}{\omega}. \tag{4}$$

Using the same strategy one obtains that also

$$\left|\int_{\gamma_3} f(z)e^{i\omega z}\,\mathrm{d}z\right| \leq \sup_{z\in\gamma_3}|f(z)|\frac{1}{\omega}. \tag{5}$$

Since for all curves γ_i with $i = 1,2,3$ the modulus of all points converges uniformly to $+\infty$ when $r,s \to +\infty$ it follows by the assumption on f that

$$\lim_{r,s\to+\infty}\sup_{z\in\gamma_i}|f(z)| = 0$$

for $i = 1,2,3$. Combined with (3),(4) and (5) this proves the claim in the case $\omega > 0$.

If $\omega < 0$ then one uses the same strategy but with a square in the lower half plane. In this case $q < 0$, so that still $\omega q > 0$. The details are left to the interested reader. □

This was the last topic about complex analysis. Next we will start with a brief summary of abstract Lebesgue integration.

2. The Lebesgue integral: an introduction mostly without proofs

Recall that the Riemann integral of a continuous function f is defined via an approximation with Riemann sums discretizing the interval (say $[a,b]$) where f is defined. The idea of Lebesgue was to discretize not the domain, but the image of f and approximate also discontinuous functions f by a sequence of functions taking only finitely many values. The preimages of those values might be quite irregular sets, so that we need to measure their length in order to define an integral. More precisely, let $f : [a,b] \to \mathbb{R}$ be a function that attains only finitely many values $a_1,\dots,a_n$. Then we would like to define its integral by

$$\int f\,\mathrm{d}x = \sum_{i=1}^n a_i\,\mathrm{length}(f^{-1}(a_i)).$$

So in order to construct an integral following Lebesgue's idea we need to define a notion of length for a large class of sets (the larger the class, the more functions we can integrate). On the one hand the mathematician Felix Hausdorff noted already in 1914 that the usual intuition of length, area or volume cannot be extended to all subsets of $\mathbb{R}^d$ in a consistent way.

On the other hand, in case we can define a suitable notion of length, area or volume, we can perform integration on more abstract spaces. The correct term is the measure of a set. For instance, in probability theory we can consider the set $\{1,2,3,4,5,6\}$ as the outcome of rolling a dice and we associate the measure $1/6$ to each event (if the dice is fair). In this way we could integrate functions $f : \{1,2,3,4,5,6,\} \to \mathbb{R}$.

In what follows we start from a very abstract point of view. First we define the objects which we want to measure.

Definition 2.1 (σ-algebra)**.** Let Ω be a non-empty set and denote by $\mathcal{P}(\Omega)$ the set of all subsets of Ω. A subset $\mathcal{F} \subset \mathcal{P}(\Omega)$ is called a σ-algebra (on Ω) if

(i) $\Omega \in \mathcal{F}$;
(ii) $A \in \mathcal{F} \implies A^c := \Omega \setminus A \in \mathcal{F}$;
(iii) $(A_n)_{n\in\mathbb{N}} \subset \mathcal{F} \implies \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$.

The sets $A \in \mathcal{F}$ are called ($\mathcal{F}$-)measurable sets.

So far this definition is not very enlightening. For every set Ω there are the trivial σ-algebras $\{\emptyset, \Omega\}$ and $\mathcal{P}(\Omega)$. Note that by (i) and (ii) we always have $\emptyset \in \mathcal{F}$. Then also all finite unions are contained in $\mathcal{F}$ and since

$$\left(\bigcup_{n\in\mathbb{N}} A_n\right)^c = \bigcap_{n\in\mathbb{N}} A_n^c$$

also all finite and countable intersections.

Definition 2.2. Let Ω be a non-empty set.

(i) If $B \subset \mathcal{P}(\Omega)$ then there exists a smallest σ-algebra containing B (cf. exercise H 6.5). It is denoted by $\sigma(B)$ and is called the σ-algebra generated by B;

(ii) if $\Omega = \mathbb{R}^d$ and B denotes the set of all open subsets of $\mathbb{R}^d$ we write $\sigma(B) = \mathcal{B}(\mathbb{R}^d)$. This is called the Borel σ-algebra on $\mathbb{R}^d$. It also contains/is generated by the set of all closed subsets;

(iii) if $\mathcal{F} \subset \mathcal{P}(\Omega)$ is a σ-algebra and $C \subset \Omega$ the trace σ-algebra on C is defined by $\mathcal{F}_C = \{A \cap C : A \in \mathcal{F}\}$. One can show that $\mathcal{F}_C$ is a σ-algebra on C.

Next we define what it means to measure sets in a σ-algebra. Again the definition is abstract and just states the minimal requirements.

Definition 2.3. Let $\mathcal{F}$ be a σ-algebra on a non-empty set Ω. A non-negative function $\mu : \mathcal{F} \to [0, +\infty]$ is called a measure if

(i) $\mu(\emptyset) = 0$;

(ii) If $(A_n)_{n\in\mathbb{N}} \subset \mathcal{F}$ is a sequence of pairwise disjoint subsets, then μ is countably additive in the sense that

$$\mu(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} \mu(A_n).$$

The triplet $(\Omega, \mathcal{F}, \mu)$ is called measure space.

Note that a measure is also finitely additive since $A_1 \cup \ldots \cup A_k = A_1 \cup \ldots \cup A_k \cup \emptyset \cup \emptyset \cup \ldots$. On any set Ω we can define the counting measure on $\mathcal{P}(\Omega)$ by

$$\mu(A) = \begin{cases} \text{cardinality of } A & \text{if } A \text{ contains finitely many elements,} \\ +\infty & \text{otherwise.} \end{cases}$$

Other measures are the so-called Dirac measure δ_ω (cf. exercise H 6.5) or the trivial measure $\mu \equiv 0$. In order to define an integral we need the notion of measurable functions.

Definition 2.4. Let Ω_1, Ω_2 be two non-empty sets with σ-algebras $\mathcal{F}_1$ and $\mathcal{F}_2$ on Ω_1 and Ω_2, respectively. A function $f : \Omega_1 \to \Omega_2$ is called ($\mathcal{F}_1$-$\mathcal{F}_2$)-measurable if $f^{-1}(A) \in \mathcal{F}_1$ for all $A \in \mathcal{F}_2$.

Let us collect some abstract results about measurability that turn out to be useful later on.

Lemma 2.5. *Let $\Omega_1, \Omega_2, \Omega_3$ be non-empty sets with σ-algebras $\mathcal{F}_1, \mathcal{F}_2$ and $\mathcal{F}_3$, respectively.*

a) If $f : \Omega_1 \to \Omega_2$ and $g : \Omega_2 \to \Omega_3$ are measurable, then the composition $g \circ f : \Omega_1 \to \Omega_3$ is also measurable.

b) If $\mathcal{F}_2 = \sigma(B)$ for some $B \subset \mathcal{P}(\Omega_2)$, then $f : \Omega_1 \to \Omega_2$ is already measurable if $f^{-1}(A) \in \mathcal{F}_1$ for all $A \in B$.

Proof. Part a) follows from the fact that for every $A \in \mathcal{F}_3$ we have by the measurability of f and g that

$$(g \circ f)^{-1}(A) = f^{-1}(\underbrace{g^{-1}(A)}_{\in \mathcal{F}_2}) \in \mathcal{F}_1.$$

In order to proof b), note that the set

$$\mathcal{A} := \{A \subset \Omega_2 : f^{-1}(A) \in \mathcal{F}_1\}$$

is a σ-algebra on Ω_2 (this is left as an exercise for the interested reader, but as such it is not relevant for the exam). Since $B \subset \mathcal{A}$ by assumption it follows that $\mathcal{F}_2 = \sigma(B) \subset \sigma(\mathcal{A}) = \mathcal{A}$, where in the last equality we used that $\mathcal{A}$ is already a σ-algebra, hence the smallest one containing $\mathcal{A}$. Thus by definition of $\mathcal{A}$ it follows that f is measurable. $\square$

Our goal is to define the integral of as many measurable functions $f : \Omega \to \mathbb{R}$ as possible, where we equip *the image space* $\mathbb{R}$ with the Borel σ-algebra[7]. We start with the case of measurable functions that attain only finitely many values.

Definition 2.6. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Denote by $S_+(\Omega)$ the class of all non-negative $\mathcal{F}$-measurable functions that attain only finitely many values, i.e., we can write

$$f = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}$$

with $a_i \in [0, +\infty)$ for all $i = 1, \dots, n$, $(A_i)_{i=1}^n$ pairwise disjoint and measurable, and the indicator function

$$\mathbb{1}_{A_i}(\omega) = \begin{cases} 1 & \text{if } \omega \in A_i, \\ 0 & \text{otherwise.} \end{cases}$$

The functions in $S_+(\Omega)$ are called non-negative simple functions. We define their integral over Ω with respect to μ by

$$\int_\Omega f \, \mathrm{d}\mu = \sum_{i=1}^{n} a_i \, \mu(A_i).$$

Remark 2.7. The representation of simple functions is not unique. However, the definition of the integral does not depend on the representation of simple functions. To prove this, note that if $\sum_{i=1}^n a_i \mathbb{1}_{A_i} = \sum_{j=1}^m b_i \mathbb{1}_{B_i}$ and $A_i \cap B_j \neq \emptyset$ then, for $\omega \in A_i \cap B_j$ we have $a_i = f(\omega) = b_j$. Moreover, by adding $a_{n+1} = 0$, $b_{m+1} = 0$ and $A_{n+1} = \{f = 0\} = B_{m+1}$ (if necessary) we can assume that $\bigcup_i A_i = \Omega = \bigcup_j B_j$. Hence by finite additivity of μ

$$\sum_{i=1}^{n} a_i \mu(A_i) = \sum_{i=1}^{n} a_i \mu\Big(\underbrace{\bigcup_{j=1}^{m} B_j}_{=\Omega} \cap A_i\Big) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i \mu(A_i \cap B_j) = \sum_{j=1}^{m} \sum_{i=1}^{n} b_j \mu(A_i \cap B_j) = \sum_{j=1}^{m} b_j \mu(B_j).$$

As a next step we define the integral of general measurable non-negative functions $f : \Omega \to [0, +\infty)$. As announced, this is achieved via a discretization of the image as performed in the proof of the next proposition. We remark that the statement is independent of the measure μ.

Proposition 2.8. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f : \Omega \to [0, +\infty)$ be $\mathcal{F}$-measurable. Then there exists a sequence of simple functions $f_n \in S_+(\Omega)$ such that $f_n \leq f_{n+1}$ and $f = \sup_n f_n$.*

Proof. For fixed $n \in \mathbb{N}$ we discretize the image $[0, +\infty)$ with $n2^n$ consecutive intervals of size 2^{-n} and the unbounded rest $[n, +\infty)$. Here come the formulas: for any natural number $k \in \{0, \dots, n2^n\}$ define the sets $A_{k,n}$ by

$$A_{k,n} = \begin{cases} \{w \in \Omega : \; k2^{-n} \leq f(\omega) < (k+1)2^{-n}\} & \text{if } 0 \leq k < n2^n, \\ \{w \in \Omega : \; f(\omega) \geq n\} & \text{if } k = n2^n. \end{cases}$$

Note that $[a, b) \in \mathcal{B}(\mathbb{R})$ for all $a \in \mathbb{R}$ and $b \in \mathbb{R} \cup \{+\infty\}$ since it is the intersection of an open and a closed set. Hence it follows that every set $A_{k,n}$ is $\mathcal{F}$-measurable as the preimage of a Borel-measurable set. Moreover, for fixed $n \in \mathbb{N}$ the family $(A_{k,n})_{k=0}^{n2^n}$ is pairwise disjoint since the sets are preimages of disjoint intervals under f. Thus we can define a function $f_n \in S_+(\Omega)$ by

$$f_n = \sum_{k=0}^{n2^n} k2^{-n} \mathbb{1}_{A_{k,n}}.$$

[7] Sometimes also the values $\pm\infty$ are allowed which requires a σ-algebra on $[-\infty, \infty]$. One takes the σ-algebra generated by the half-open intervals $[a, b)$ with $a \in [-\infty, \infty)$ and $b \in \mathbb{R}$.

By construction $f_n \leq f_{n+1}$ since each interval in the discretization gets bisected when n increases by 1 (and new intervals in the range $[n, n+1)$ are added). Moreover, for any $\omega \in \Omega$ we have that

$$0 \leq f(\omega) - f_n(\omega) \leq 2^{-n} \qquad \forall n > f(\omega).$$

Thus by monotonicity in n we have $f(\omega) = \lim_n f_n(\omega) = \sup_n f_n(\omega)$. □

The previous proposition allows us to define the Lebesgue integral for general non-negative measurable functions.

Definition 2.9. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f : \Omega \to [0, +\infty)$ be $\mathcal{F}$-measurable. Then we define

$$\int_\Omega f \, \mathrm{d}\mu = \sup_{n\in\mathbb{N}} \int_\Omega f_n \, \mathrm{d}\mu,$$

where f_n is a sequence of simple functions as in Proposition 2.8. One can show that this definition does not depend on the particular choice of the sequence f_n of simple functions.

Note that the integral of a non-negative function might equal $+\infty$. Finally, we define the integral of measurable functions without any sign condition. To this end, we decompose a measurable function $f : \Omega \to \mathbb{R}$ into its positive and negative part f^+ and f^- defined as

$$f^+(\omega) = \max\{f(\omega), 0\}, \qquad f^-(\omega) = -\min\{f(\omega), 0\}.$$

Then $f = f^+ - f^-$ and both $f^+, f^- : \Omega \to [0, +\infty)$. Moreover, one can show that both f^+, f^- are measurable.[8] The idea is to define the integral of f as the difference of the integrals of f^+ and f^-. This is contained in the next definition.

Definition 2.10. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f : \Omega \to \mathbb{R}$ be measurable. Denote by f^+ and f^- the positive and negative part of f defined above. We say that f is Lebesgue integrable if and only if

$$\int_\Omega f^+ \, \mathrm{d}\mu < +\infty \quad \text{and} \quad \int_\Omega f^- \, \mathrm{d}\mu < +\infty,$$

where the integrals are defined as in Definition 2.9. In this case we define the integral of f over Ω by

$$\int_\Omega f \, \mathrm{d}\mu = \int_\Omega f^+ \, \mathrm{d}\mu - \int_\Omega f^- \, \mathrm{d}\mu.$$

Before we state some properties of the Lebesgue integral we need to study which real-valued functions are measurable. This question of course depends on the σ-algebra, but for now we treat only operations on measurable functions. We omit the proof of the following lemma which is rather technical using different sets that generate the Borel σ-algebra.

Lemma 2.11. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f, g : \Omega \to \mathbb{R}$ be measurable. Moreover, let $(f_n)_{n\in\mathbb{N}} : \Omega \to \mathbb{R}$ be a sequence of measurable functions. Then the following functions are measurable[9]:*

(i) $f + g$ *and* $f - g$;
(ii) $f \cdot g$ *and* f/g *(the latter provided* $g(\omega) \neq 0 \, \forall \, \omega \in \Omega$*)*;
(iii) $\sup_{n\in\mathbb{N}} f_n$ *and* $\inf_{n\in\mathbb{N}} f_n$;
(iv) $\liminf_{n\to+\infty} f_n$ *and* $\limsup_{n\to+\infty} f_n$.

In the next lemma we state some elementary properties of the Lebesgue integral. We leave the proof (which follows essentially by going through the approximation steps) to the reader.

Lemma 2.12. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f, g : \Omega \to \mathbb{R}$ be integrable (this always implies that they are measurable). Then*

[8] Note that the function $x \mapsto \max\{x, 0\}$ is continuous. By Lemma 2.5 any continuous function is $(\mathcal{B}(\mathbb{R})$-$\mathcal{B}(\mathbb{R}))$-measurable and the composition of measurable functions remains measurable. This clarifies the measurability of f^+. To show that f^- is measurable we note that the multiplication by -1 is also continuous, so that $-f$ is measurable and $-\min\{f(\omega), 0\} = \max\{-f(\omega), 0\}$.

[9] The quantities in (iii) and (iv) might take the values $\pm\infty$. Measurability is understood in the sense of Footnote 7.

(i) $\int_\Omega \alpha f + \beta g \,\mathrm{d}\mu = \alpha \int_\Omega f \,\mathrm{d}\mu + \beta \int_\Omega g \,\mathrm{d}\mu$ *for all* $\alpha, \beta \in \mathbb{R}$*;*

(ii) *if* $f \leq g$*, then* $\int_\Omega f \,\mathrm{d}\mu \leq \int_\Omega g \,\mathrm{d}\mu$*;*

(iii) $|f|$ *is integrable*[10] *and*

$$\left|\int_\Omega f \,\mathrm{d}\mu\right| \leq \int_\Omega |f| \,\mathrm{d}\mu.$$

Next we introduce the integral over a subset.

Definition 2.13. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f : \Omega \to \mathbb{R}$ be integrable. Given a measurable set $A \in \mathcal{F}$ we define

$$\int_A f \,\mathrm{d}\mu = \int_\Omega \mathbb{1}_A f \,\mathrm{d}\mu.$$

Note that the integrand in the previous definition is well-defined due to Lemma 2.11 since $\mathbb{1}_A$ is measurable if and only if A is a measurable set.

2.1. **Convergence theorems.** In this subsection we state some important convergence theorems for sequences of integrable functions. Before we start, we introduce the notion of null sets which are in a sense negligible for integration theory.

Definition 2.14. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. A set $A \in \mathcal{F}$ is called a null set if $\mu(A) = 0$.

Remark 2.15. Some authors don't require that a null set is measurable, but that it is contained in a measurable set with measure zero. This difference is somehow irrelevant since one can always complete a measure space so that every subset of a null set is measurable (cf. exercises). Measure spaces with that property are called complete.

In the next lemma we prove that if two functions differ only on a null set then their integrals agree.

Lemma 2.16. *Let* $(\Omega, \mathcal{F}, \mu)$ *be a measure space and* $f, g : \Omega \to \mathbb{R}$ *be integrable. If the measurable set* $\{f \neq g\}$ *is a null set, then*

$$\int_\Omega f \,\mathrm{d}\mu = \int_\Omega g \,\mathrm{d}\mu.$$

Proof. By linearity it suffices to consider the case $g \equiv 0$. We prove that the integral of the positive part f^+ vanishes (the proof for the negative part f^- is analogous). Set $A = \{f^+ \neq 0\}$ which is measurable since $\mathbb{R} \setminus \{0\}$ is open, hence a Borel set. Note that $\mathbb{1}_A$ is a simple function in $S_+(\Omega)$. Take a sequence of simple functions $f_n^+ \in S_+(\Omega)$ as in Proposition 2.8 for f^+. Write $f_n^+ = \sum_{i=1}^{m_n} a_i^n \mathbb{1}_{A_i^n}$. Then $g_n := f_n^+ \mathbb{1}_A$ is a sequence of simple functions such that $g_n \leq g_{n+1}$ and $\sup_n g_n = f^\pm \mathbb{1}_A = f^+$. Hence it is admissible for defining the Lebesgue integral of f^+ and thus

$$\int_\Omega f^+ \,\mathrm{d}\mu = \sup_{n\in\mathbb{N}} \int_\Omega \mathbb{1}_A f_n^+ \,\mathrm{d}\mu = \int_\Omega \sum_{i=1}^{m_n} a_i^n \mathbb{1}_{A_i^n \cap A} \,\mathrm{d}\mu = \sup_{n\in\mathbb{N}} \sum_{i=1}^{m_n} a_i^n \underbrace{\mu(A_i^n \cap A)}_{0 \leq \mu(A_i^n \cap A) \leq \mu(A) \leq \mu(\{f \neq 0\}) = 0} = 0.$$

Here we used that by definition $\mathbb{1}_A \mathbb{1}_{A_i^n} = \mathbb{1}_{A \cap A_i^n}$ and that measures are monotone with respect to set inclusion since for any measurable sets $B \subset C$ we have by additivity that

$$\mu(C) = \mu(C \setminus B \cup B) = \underbrace{\mu(C \setminus C)}_{\geq 0} + \mu(B) \geq \mu(B).$$

□

In order to understand null sets a bit better we next establish some general properties of measures.

Lemma 2.17. *Let* $(\Omega, \mathcal{F}, \mu)$ *be a measure space. Then* μ *satisfies the following properties for every* $A, B \in \mathcal{F}$ *and* $(A_n)_{n\in\mathbb{N}} \subset \mathcal{F}$*:*

[10] taking the modulus is a continuous operation on $\mathbb{R}$, so that $|f|$ is measurable by Lemma 2.5

(i) $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$;
(ii) [monotonicity] *If* $A \subset B$ *then* $\mu(A) \leq \mu(B)$;
(iii) *If* $A \subset B$ *and* $\mu(A) < +\infty$, *then* $\mu(B \setminus A) = \mu(B) - \mu(A)$.
(iv) [σ-subadditivity] $\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n)$.

Proof. Note that for $A \subset B$ we can write $B = (B \setminus A) \cup A$ and this union is disjoint. Thus by finite additivity

$$\mu(B) = \mu(A) + \mu(B \setminus A).$$

Since μ is non-negative this implies (ii) and moreover (iii) if $\mu(A) < +\infty$. Moreover, for general $A, B \in \mathcal{F}$ such that $\mu(A \cap B) = +\infty$ the property (i) follows from (ii) as both sides are $+\infty$. If $\mu(A \cap B) < +\infty$ property (i) follows from

$$\mu(A \cup B) = \mu(A \cup (B \setminus A)) = \mu(A) + \mu(B \setminus A),$$
$$\mu(B) = \mu((B \setminus A) \cup (B \cap A) = \mu(B \setminus A) + \mu(B \cap A)$$

by solving the second equation for $\mu(B \setminus A)$ and inserting the result in the first equation. In order to show the last property, we make the sets pairwise disjoint by defining $B_1 = A_1$ and

$$B_n = A_n \setminus \left(\bigcup_{i=1}^{n-1} A_i\right).$$

Then by construction the family $(B_n)_{n\in\mathbb{N}}$ is pairwise disjoint and

$$\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} B_n.$$

Since $B_n \subset A_n$ it follows from σ-additivity and (ii) that

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \mu\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \sum_{n=1}^{\infty} \mu(B_n) \leq \sum_{n=1}^{\infty} \mu(A_n).$$

□

The previous lemma implies that the countable union of null sets is still a null set and that adding or subtracting a null set to a measurable set does not change its measure. In what follows we say that a property holds for μ-almost every $\omega \in \Omega$ if it holds for all $\omega \in \Omega \setminus N$, where N is a null set. The following convergence results (especially (3)) are the reason why the Lebesgue integral should be preferred over the Riemann integral.

Theorem 2.18 (Convergence theorems)**.** *Let* $(\Omega, \mathcal{F}, \mu)$ *be a measure space and* $(f_n)_{n\in\mathbb{N}} : \Omega \to \mathbb{R}$ *be a sequence of measurable functions and* $f : \Omega \to \mathbb{R}$.

(1) [Monotone convergence theorem] *If* $0 \leq f_1(\omega) \leq f_2(\omega) \leq \ldots \leq f_n(\omega) \to f(\omega)$ *for* μ*-almost every* $w \in \Omega$ *then*
$$\lim_{n\to+\infty} \int_\Omega f_n \,\mathrm{d}\mu = \int_\Omega f \,\mathrm{d}\mu.$$
(2) [Fatou's lemma] *If* $f_n(\omega) \geq 0$ *for* μ*-almost every* $w \in \Omega$ *and* $f(\omega) := \liminf_{n\to+\infty} f_n(\omega)$ *we have*
$$\int_\Omega f \,\mathrm{d}\mu \leq \liminf_{n\to+\infty} \int_\Omega f_n \,\mathrm{d}\mu.$$
(3) [Dominated convergence theorem] *Suppose that* $f(\omega) := \lim_{n\to+\infty} f_n(\omega)$ *exists for* μ*-almost every* $\omega \in \Omega$ *and that there exists an integrable function* $g : \Omega \to [0, +\infty)$ *such that* $|f_n(\omega)| \leq g(\omega)$ *for* μ*-almost every* $\omega \in \Omega$. *Then*
$$\lim_{n\to+\infty} \int_\Omega f_n \,\mathrm{d}\mu = \int_\Omega f \,\mathrm{d}\mu.$$

Remark 2.19. The above theorems (1) and (2) also hold for functions taking the value $+\infty$ (the same construction as in the proof of Proposition 2.8 allows to approximate non-negative measurable functions taking the value $+\infty$ and we can define their integral again by approximation with simple functions in $S_+(\Omega)$). Also in the above formulation it is not guaranteed that the integrals in (1) and (2) are finite. They are in (3) since g is integrable and dominates $|f_n|$ and thus also $|f|$.

Proof. (1) Up to redefining $f_n(\omega) := f(\omega) := 0$ on the null set where $f_j(\omega) > f_{j+1}(\omega)$ for some $j \in \mathbb{N}$ we can assume that the sequence $(f_n)_{n\in\mathbb{N}}$ is pointwise increasing, so that it follows from Lemma 2.12 that

$$\int_\Omega f_n \,\mathrm{d}\mu \leq \int_\Omega f \,\mathrm{d}\mu.$$

Taking the supremum over $n \in \mathbb{N}$ we deduce again by monotonicity of f_n that

$$\lim_{n\to+\infty}\int_\Omega f_n \,\mathrm{d}\mu = \sup_{n\in\mathbb{N}}\int_\Omega f_n \,\mathrm{d}\mu \leq \int_\Omega f \,\mathrm{d}\mu.$$

In order to prove the reverse inequality, take for every $n \in \mathbb{N}$ a sequence $(f_{k,n})_{k\in\mathbb{N}} \subset S_+(\Omega)$ as in Proposition 2.8 for f_n. Thus $f_n = \sup_{k\in\mathbb{N}} f_{k,n}$ and the sequence $f_{k,n}$ is increasing in k. Define

$$g_n := \sup_{j\leq n} f_{n,j} \leq \sup_{j\leq n} f_j = f_n.$$

Then g_n is measurable by Lemma 2.11 and attains only finitely many non-negative values. Thus $g_n \in S_+(\Omega)$. Moreover, since $f_{n,j}$ is increasing in n it follows that $g_n \leq g_{n+1}$ since we also take more competitors with respect to j in the supremum. For $m, k \in \mathbb{N}$ and $n \geq \max\{m,k\}$ we have by the monotonicity of $(f_{k,n})$ with respect to k that

$$f_{k,m} \leq f_{n,m} \leq g_n \leq f_n.$$

Set $g := \sup_{n\in\mathbb{N}} g_n$. Taking first the supremum in n and then in m, k in the above inequality we conclude that

$$f = \sup_{m\in\mathbb{N}} f_m = \sup_{m\in\mathbb{N}}\sup_{k\in\mathbb{N}} f_{k,m} \leq \sup_{n\in\mathbb{N}} g_n.$$

Thus it follows from monotonicity and the definition of the Lebesgue integral of g as the supremum of integrals of g_n that

$$\int_\Omega f \,\mathrm{d}\mu \leq \int_\Omega g \,\mathrm{d}\mu = \sup_{n\in\mathbb{N}}\int_\Omega g_n \,\mathrm{d}\mu \leq \sup_{n\in\mathbb{N}}\int_\Omega f_n \,\mathrm{d}\mu = \lim_{n\to+\infty}\int_\Omega f_n \,\mathrm{d}\mu.$$

This concludes the proof of (1).

In order to show (2) note that by definition of the liminf we have

$$\liminf_{n\to+\infty} f_n = \sup_{n\in\mathbb{N}} g_n, \qquad g_n := \inf_{m\geq n} f_m.$$

The sequence g_n is monotone increasing in n and non-negative. We thus can apply (1). First note that $g_n \leq f_m$ for all $m \geq n$, so by monotonicity of the Lebesgue integral we have that

$$\int_\Omega g_n \,\mathrm{d}\mu \leq \inf_{m\geq n}\int_\Omega f_m \,\mathrm{d}\mu.$$

Thus the monotone convergence theorem yields that

$$\int_\Omega \liminf_{n\to+\infty} f_n \,\mathrm{d}\mu = \int_\Omega \sup_{n\in\mathbb{N}} g_n \,\mathrm{d}\mu = \sup_{n\in\mathbb{N}}\int_\Omega g_n \,\mathrm{d}\mu \leq \sup_{n\in\mathbb{N}}\inf_{m\geq n}\int_\Omega f_m \,\mathrm{d}\mu = \liminf_{n\to+\infty}\int_\Omega f_n \,\mathrm{d}\mu.$$

Finally, we prove (3). Note that $|f| \leq g$ for μ-almost every $\omega \in \Omega$. Up to changing f_n, f on a null set (not affecting their integrals) we may assume that $|f_n(\omega)|, |f(\omega)| \leq g(\omega)$ for all $\omega \in \Omega$. Thus $|f|$ is also integrable and by linearity also $h := g + |f|$. Moreover, by the triangle inequality we have that $|f_n(\omega) - f(\omega)| \leq |f_n(\omega)| + |f(\omega)| \leq g(\omega) + |f(\omega)| = h(\omega)$, so that

$$h(\omega) = \lim_{n\to+\infty} \underbrace{h(\omega) - |f_n(\omega) - f(\omega)|}_{\geq 0}.$$

Fatou's lemma yields

$$\int_\Omega h \,\mathrm{d}\mu \le \liminf_{n\to+\infty} \int_\Omega h - |f_n - f| \,\mathrm{d}\mu = \int_\Omega h \,\mathrm{d}\mu - \limsup_{n\to+\infty} \int_\Omega |f_n - f| \,\mathrm{d}\mu.$$

Since h is integrable we can cancel its integral to deduce that

$$0 \le \limsup_{n\to+\infty} \int_\Omega |f_n - f| \,\mathrm{d}\mu \le 0.$$

Hence the integral above converges to 0. By the triangle inequality for the Lebesgue integral this implies that

$$\left| \int_\Omega f_n - f \,\mathrm{d}\mu \right| \le \int_\Omega |f_n - f| \,\mathrm{d}\mu \to 0.$$

This proves the claim. □

In the exercises you will see that the assumption on non-negativity is essential in Fatou's lemma. The assumption in the dominated convergence theorem however is not sharp, but can be slightly weakened using the concept of equiintegrability which we will not introduce in this course.

The dominated convergence theorem can be used to prove the differentiability of parameter-dependent integrals as stated in the corollary below.

Corollary 2.20. *Let $U \subset \mathbb{R}^d$ be open and $(\Omega, \mathcal{F}, \mu)$ be a measure space. Let $f : U \times \Omega \to \mathbb{R}$ be such that for every $x \in U$ the function $\omega \mapsto f(x, \omega)$ is integrable. Assume further that the partial derivative $\partial_i f(x, \omega)$ exists for all $x \in U$ and $\omega \in \Omega$ and that there exists an integrable function $h : \Omega \to [0, +\infty)$ such that $|\partial_i f(x, \omega)| \le h(\omega)$ for all $x \in U$ and $\omega \in \Omega$. Then the function*

$$F(x) = \int_\Omega f(x, \omega) \,\mathrm{d}\mu(\omega)$$

is partially differentiable with respect to x_i on U and

$$\partial_i F(x) = \int_\Omega \partial_i f(x, \omega) \,\mathrm{d}\mu(\omega).$$

Here the notation $\mathrm{d}\mu(\omega)$ indicates that we integrate with respect to ω.

Proof. See exercise H 7.5. □

2.2. **The space $L^p(\Omega; \mu)$.** Next we introduce a vector space structure on integrable functions. In particular, we introduce the Hilbert space $L^2(\Omega; \mu)$ which appears in many applications. By Lemma 2.12 (i) the set of all Lebesgue integrable functions on a measure space $(\Omega, \mathcal{F}, \mu)$ forms a vector space by pointwise addition and scalar multiplication. For $p \in [1, +\infty]$ we define

$$\mathscr{L}^p(\Omega; \mu) := \begin{cases} \{f : \Omega \to \mathbb{R} \text{ measurable}, \displaystyle\int_\Omega |f|^p \,\mathrm{d}\mu < +\infty\} & \text{if } p < +\infty, \\ \\ \{f : \Omega \to \mathbb{R} \text{ measurable}, \exists C > 0 \text{ s.t. } |f(\omega)| \le C \ \mu\text{-almost everywhere}\} & \text{if } p = +\infty. \end{cases}$$

and for $f \in \mathscr{L}^p(\Omega; \mu)$ the semi-norms

$$\|f\|_p = \begin{cases} \left(\displaystyle\int_\Omega |f|^p \,\mathrm{d}\mu \right)^{\frac{1}{p}} & \text{if } p < +\infty, \\ \\ \inf\{c \in \mathbb{R} : \ |f(\omega)| \le c \ \mu\text{-almost everywhere}\} & \text{if } p = +\infty. \end{cases}$$

The quantity $\|f\|_p$ defines only a semi-norm as the following lemma shows:

Lemma 2.21. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f, g \in \mathscr{L}^p(\Omega; \mu)$. Then*

(i) $\|\lambda f\|_p = |\lambda| \, \|f\|_p$ *for all* $\lambda \in \mathbb{R}$*;*
(ii) $\|f + g\|_p \le \|f\|_p + \|g\|_p$ *(also known as* Minkowski inequality*);*
(iii) $\|f\|_p = 0 \iff f = 0 \quad \mu$*-almost everywhere.*

Proof. (i) follows by definition. (ii) will be proven in the exercises. In order to prove the equivalence in (iii), note that $f = 0$ μ-almost everywhere implies that $|f|^p = 0$ μ-almost everywhere, so that $\|f\|_p = 0$ by Lemma 2.16 in the case $p < +\infty$ while it follows from the definition if $p = +\infty$. To prove the converse statement, note that in the case $p = +\infty$ for every $n \in \mathbb{N}$ we have $|f(\omega)| \leq \frac{1}{n}$ for every $\omega \in \Omega \setminus N_n$, where N_n is a null set. Then $N := \bigcup_{n\in\mathbb{N}} N_n$ is still a null set as a countable union of null sets. On $\Omega \setminus N$ we have that $|f(\omega)| = 0$. Hence $f = 0$ μ-almost everywhere. If $1 \leq p < +\infty$ we have for every $n \in \mathbb{N}$ that

$$\frac{1}{n}\mu(\{|f|^p \geq \tfrac{1}{n}\}) \leq \int_{\{|f|^p\geq\frac{1}{n}\}} |f|^p \,\mathrm{d}\mu \leq \|f\|_p^p = 0.$$

Thus

$$\mu(\{|f|^p > 0\}) = \mu(\bigcup_{n\in\mathbb{N}}\{|f|^p \geq \tfrac{1}{n}\}) \leq \sum_{n=1}^{\infty}\mu(\{|f|^p \geq \tfrac{1}{n}\}) = 0.$$

Hence $f = 0$ μ-almost everywhere. □

Before we deal with the problem that $\|f\|_p = 0$ does not imply that $f = 0$ (which rules out that $\|\cdot\|_p$ defines a norm) we state another inequality that explains the relationship between the spaces $\mathscr{L}^p(\Omega;\mu)$ for different exponents p.

Proposition 2.22 (Hölder inequality)**.** *Let $f \in \mathscr{L}^p(\Omega;\mu)$ and $g \in \mathscr{L}^q(\Omega;\mu)$ with $\frac{1}{p} + \frac{1}{q} = 1$ (with the convention that $1/+\infty = 0$). Then $fg \in \mathscr{L}^1(\Omega;\mu)$ and*

$$\|fg\|_1 \leq \|f\|_p\|g\|_q.$$

Proof. See exercises. □

In order to resolve the problem of defining a norm on integrable functions one introduces equivalence classes by identifying two functions that agree μ-almost everywhere. We will not recall the mathematical properties of quotient spaces, but just that the equivalence class of $f \in \mathscr{L}^p(\Omega;\mu)$ is given by

$$[f] = \{g \in \mathscr{L}^p(\Omega;\mu) :\, g = f \quad \mu\text{-almost everywhere}\}.$$

The vector space $L^p(\Omega;\mu)$ is then defined as the quotient space

$$L^p(\Omega;\mu) := \{[f] :\, f \in \mathscr{L}^p(\Omega;\mu)\}.$$

In that sense elements in $L^p(\Omega;\mu)$ are not functions but equivalence classes of functions. On this space $\|f\|_p$ defines a norm. We recall properties of this space in the following theorem which we will not prove here.

Theorem 2.23. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. Then for $p \in [1,+\infty]$ the space $L^p(\Omega;\mu)$ is a Banach space with the norm $\|\cdot\|_p$. If $p = 2$ it is a Hilbert space with the scalar product*

$$\langle [f],[g]\rangle = \int_\Omega fg\,\mathrm{d}\mu,$$

which does not depend on the representatives of the equivalence classes $[f]$ and $[g]$.

Remark 2.24. Since the elements in $L^p(\Omega;\mu)$ are equivalence classes, one has to be careful with pointwise evaluations $[f](\omega)$ since they might depend on the representative of the equivalence class. Nevertheless, it is customary to write f instead of $[f]$ whenever it is clear that the quantity containing f does not depend on the representative (e.g. integrals or $\mu(\{f > 0\})$) or one emphasizes that one takes a representative if this is not the case. In general, if Ω is a subset of $\mathbb{R}^d$ then statements like f is continuous, differentiable etc. are not well-defined. However, for the Lebesgue measure (see next subsection) this usually means that there is a representative in the equivalence class that satisfies these properties.

2.3. **The Lebesgue measure on $\mathbb{R}^d$.** In this subsection we introduce a measure on $\mathbb{R}^d$ with the Borel σ-algebra. Integration with respect to this measure replaces the Riemann integral. The following theorem contains the defining property.

Theorem 2.25. *Let $\mathcal{B}(\mathbb{R}^d)$ be the Borel σ-algebra on $\mathbb{R}^d$. Then there exists a unique measure $\mu : \mathcal{B}(\mathbb{R}^d) \to [0, +\infty]$ such that*

$$\mu([a_1, b_1) \times \ldots \times [a_d, b_d)) = \prod_{i=1}^{d} (b_i - a_i) \qquad \forall a_i, b_i \in \mathbb{R},\ a_i < b_i.$$

The completion (cf. exercise H 7.2) is denoted by $(\Omega, \mathcal{L}^d, \lambda_d)$ and λ_d is the d-dimensional Lebesgue measure while sets in the completed σ-algebra $\mathcal{L}^d$ are called Lebesgue-measurable.

We will not prove this theorem. Skipping the proof is the main reason for emphasizing that the introduction is 'mostly without proofs'. Indeed, the argument is quite involved studying in detail which classes of sets already determine a measure and how to construct measures with a given property.

Let us state some elementary properties of the Lebesgue measure.

- $\lambda_d(\mathbb{R}^d) = +\infty$;
- $\lambda_d(\{x\}) = 0$ for all $x \in \mathbb{R}^d$;
- $\lambda_d(x + A) = \lambda_d(A)$ for all $x \in \mathbb{R}^d$, $A \in \mathcal{L}^d$;
- $\lambda_d(\alpha A) = \alpha^d \lambda_d(A)$ for all $\alpha > 0$, $A \in \mathcal{L}^d$.

Here we used the notation $x + A := \{x + a : a \in A\}$ and $\alpha A := \{\alpha a : a \in A\}$. We collect some further properties of the Lebesgue measure space in the following lemma.

Lemma 2.26. *Let $(\mathbb{R}^d, \mathcal{L}^d, \lambda_d)$ be the Lebesgue measure space. Then*

(i) $\mathcal{L}^d \neq \mathcal{P}(\mathbb{R}^d)$[11]*;*
(ii) $\lambda_d(\mathbb{R}^{d-1} \times \{0\}) = 0$*;*
(iii) $\lambda_d(K) < +\infty$ *for every bounded set* $K \in \mathcal{L}^d$*;*
(iv) $\lambda_d(U) > 0$ *for every non-empty open set* $U \subset \mathbb{R}^d$*.*

Proof. (i) The interested reader can find the construction in the optional exercise H 8.6.
(ii) is part of exercise H 8.2.
(iii) Note than any bounded set is contained in some large cube $[-n, n)^d$. Hence by monotonicity its Lebesgue measure is finite.
(iv) holds true since any non-empty open set contains a small half-open cube which has positive measure. □

Often we consider functions f defined on measurable subsets $\Omega \subset \mathbb{R}^d$. In this case we can consider the trace σ-algebra of $\mathcal{L}^d$ on Ω (cf. Definition 2.2) which in the case of a measurable set Ω coincides with the Lebesgue measurable subsets of Ω. The Lebesgue measure can then naturally be restricted to the trace σ-algebra and we can speak about Lebesgue-measurable functions on Ω. This allows to define the integral

$$\int_\Omega f \, \mathrm{d}\lambda_d$$

for those functions.

Next we want to state without proof some rather involved theorems on the Lebesgue integral. But before doing so we compare the Lebesgue integral in 1D with the Riemann integral.

---

[11]This property is based on the axiom of choice. An axiom is a property that you assume to be true but cannot be proven. Very few mathematicians don't accept the axiom of choice. Without this axiom even $\mathcal{B}(\mathbb{R}^d) = \mathcal{P}(\mathbb{R}^d)$. The axiom of choice is a rather abstract concept. An informal saying goes that every set/function that you can write down is Borel-measurable. Concerning the difference between Lebesgue and Borel-measurable, one can show that there exists a bijection between $\mathcal{B}(\mathbb{R}^d)$ and $\mathbb{R}$ while there exists a bijection between $\mathcal{L}^d$ and $\mathcal{P}(\mathbb{R})$. A general theorem from set theory says that there cannot exist a bijection between a set and its power set. Hence there are many more Lebesgue measurable sets than Borel measurable sets on $\mathbb{R}^d$. Note that the existence of a bijection is not necessarily a good measure to compare infinite sets but the non-existence of a bijection really says that the sets are different. The class of non-measurable subsets also has the same cardinality as $\mathcal{P}(\mathbb{R})$, no matter if you speak about Borel or Lebesgue measurable sets.

Theorem 2.27. *Let $f : [a, b] \to \mathbb{R}$ be bounded and Riemann integrable. Then f is also Lebesgue-integrable and the two integrals agree. Hence we adopt the notation*

$$\int_a^b f \, \mathrm{d}\lambda_1 = \int_a^b f(x) \, \mathrm{d}x.$$

Remark 2.28. The assumption on boundedness is natural since the (proper) Riemann integral is only defined for bounded functions on a closed interval.

Proof of Theorem 2.27. By definition of Riemann integrability there exists a sequence P_k of partitions of the interval $[a, b]$ and piecewise constants functions s_k, S_k with respect to those partitions such that P_{k+1} is a refinement of P_k, $s_k \leq f \leq S_k$ and

$$\lim_{k\to+\infty} \int_a^b s_k(x) \, \mathrm{d}x = \lim_{k\to+\infty} \int_a^b S_k(x) \, \mathrm{d}x = \int_a^b f(x) \, \mathrm{d}x.$$

If $P_k = \{a = x_0^k < x_1^k < \ldots < x_{n(k)}^k = b\}$, then s_k and S_k can be defined by

$$s_k = \sum_{i=1}^{n(k)} \left(\inf_{x \in (x_{i-1}^k, x_i^k]} f(x) \right) \mathbb{1}_{(x_{i-1}^k, x_i^k]}, \qquad S_k = \sum_{i=1}^{n(k)} \left(\sup_{x \in (x_{i-1}^k, x_i^k]} f(x) \right) \mathbb{1}_{(x_{i-1}^k, x_i^k]}$$

Since P_{k+1} is a refinement of P_k it follows that $s_k \leq s_{k+1}$ and $S_k \geq S_{k+1}$. Due to the boundedness of f the sequences $(s_k)_{k\in\mathbb{N}}$ and $(S_k)_{k\in\mathbb{N}}$ are also bounded and therefore they converge pointwise for every $x \in [a, b]$ by monotonicity. Denote their limits by $s, S : [a, b] \to \mathbb{R}$. Since each s_k, S_k is measurable (half-open intervals are Borel-measurable) it follows that s and S are Borel-measurable functions as the limits of measurable functions. Moreover, by construction $s(x) \leq f(x) \leq S(x)$ for all $x \in [a, b]$. By definition the Riemann integral agrees with the Lebesgue integral for s_k and S_k since those are also simple functions attaining only finitely many values. By the dominated convergence theorem for the Lebesgue integral (one can take the constant $\|f\|_\infty$ as dominating integrable function) we obtain

$$\begin{aligned}\int_a^b S \, \mathrm{d}\lambda_1 &= \lim_{k\to+\infty} \int_a^b S_k \, \mathrm{d}\lambda_1 = \lim_{k\to+\infty} \int_a^b S_k(x) \, \mathrm{d}x = \int_a^b f(x) \, \mathrm{d}x = \lim_{k\to+\infty} \int_a^b s_k(x) \, \mathrm{d}x \\ &= \lim_{k\to+\infty} \int_a^b s_k \, \mathrm{d}\lambda_1 = \int_a^b s \, \mathrm{d}\lambda_1.\end{aligned}$$

Since $S(x) \geq s(x)$ for all $x \in [a, b]$ it follows from the monotonicity of the Lebesgue integral that

$$0 \leq \int_a^b |S - s| \, \mathrm{d}\lambda_1 = \int_a^b S - s \, \mathrm{d}\lambda_1 = 0.$$

Thus $S(x) = s(x) = f(x)$ for λ_1-a.e. $x \in [a, b]$. This implies that f is Lebesgue-measurable as it agrees with a Borel-measurable function a.e. Moreover,

$$\int_a^b f \, \mathrm{d}\lambda_1 = \int_a^b S \, \mathrm{d}\lambda_1 = \int_a^b f(x) \, \mathrm{d}x.$$

□

Remark 2.29. While the above theorem clarifies that the Lebesgue integral extends the Riemann integral (for instance the function $\mathbb{1}_{\mathbb{Q}}$ is Lebesgue integrable on $[0, 1]$ but not Riemann integrable) the situation is different for improper Riemann integrals. For instance, on the one hand the function $\mathbb{1}_{\mathbb{Q}}$ is also Lebesgue integrable on $(0, +\infty)$ but the improper Riemann integral is not defined. On the other hand, the function $f(x) = \frac{\sin(x)}{x}$ is Riemann integrable on $(0, +\infty)$, while it is not Lebesgue integrable because $|f|$ has no finite integral on $(0, +\infty)$. In general the only reason why a function can have an improper Riemann integral but no Lebesgue integral is that it changes sign which causes cancellations in the Riemann integral, while the Lebesgue integral needs the integrability of $|f|$ which is less sensitive to sign changes.

Theorem 2.27 implies in particular that all (proper) Riemann integrals that have been calculated in previous courses remain the same for the Lebesgue integral. However, the Lebesgue integral provides many more tools than the Riemann integral (e.g., the dominated convergence theorem). Moreover, with the Riemann integral it is impossible to define a nice Banach space of integrable functions so that a functional analytic point of view is not easily available.

The definition of the Lebesgue integral in higher dimensions does not allow to compute integrals easily since there is no analogue of the fundamental theorem of calculus (you might consider the divergence theorem as a generalization but this requires to calculate a complicated integral over the boundary of sets and can only be applied if you integrate a divergence). In what follows we state two powerful tools to calculate some integrals. As mentioned before we omit the technical proofs.

Theorem 2.30 (Fubini's theorem)**.** *Let $\Omega_1 \subset \mathbb{R}^n, \Omega_2 \subset \mathbb{R}^m$ be Lebesgue-measurable sets and $f : \Omega_1 \times \Omega_2 \subset \mathbb{R}^{n+m} \to \mathbb{R}$ be Lebesgue-measurable. If $f \in L^1(\Omega_1 \times \Omega_2)$, then*

$$\int_{\Omega_1 \times \Omega_2} f \, \mathrm{d}\lambda_{n+m} = \int_{\Omega_1} \int_{\Omega_2} f(x,y) \, \mathrm{d}\lambda_m(y) \, \mathrm{d}\lambda_n(x) = \int_{\Omega_2} \int_{\Omega_1} f(x,y) \, \mathrm{d}\lambda_n(x) \, \mathrm{d}\lambda_m(y).$$

Remark 2.31. This version of Fubini's theorem is for complete measure spaces. In particular, the quantities appearing in the above formula are only well-defined up to null sets with respect to the corresponding measures. The above version is build on many properties of the Lebesgue measure space and might differ from what you find in textbooks for more abstract measure spaces. Many authors formulate the theorem using the Borel σ-algebras.

In what follows we also use the notation $\mathrm{d}x = \mathrm{d}\lambda_d$ for the d-dimensional integration with respect to the Lebesgue measure whenever this is clear from the context.

The next theorem is similar to Fubini's theorem but does not require the integrability of f on $\Omega_1 \times \Omega_2$ but only the non-negativity of f.

Theorem 2.32 (Tonelli's's theorem)**.** *Let $\Omega_1 \subset \mathbb{R}^n, \Omega_2 \subset \mathbb{R}^m$ be Lebesgue-measurable sets and $f : \Omega_1 \times \Omega_2 \subset \mathbb{R}^{n+m} \to \mathbb{R}$ be Lebesgue-measurable. If $f \geq 0$, then*

$$\int_{\Omega_1 \times \Omega_2} f \, \mathrm{d}\lambda_{n+m} = \int_{\Omega_1} \int_{\Omega_2} f(x,y) \, \mathrm{d}\lambda_m(y) \, \mathrm{d}\lambda_n(x) = \int_{\Omega_2} \int_{\Omega_1} f(x,y) \, \mathrm{d}\lambda_n(x) \, \mathrm{d}\lambda_m(y).$$

The integrals can be $+\infty$.

Remark 2.33. Tonelli's theorem on abstract measure spaces allows to derive an alternative formula for the Lebesgue integral using just the 1-dimensional Riemann integral. It states that

$$\int_\Omega f \, \mathrm{d}\mu = \int_0^\infty \mu\left(\{\omega \in \Omega : \, f(\omega) > t\}\right) \, \mathrm{d}t$$

for every measurable function $f : \Omega \to [0, +\infty]$. Some authors use the right hand side to define the Lebesgue integral.

The next result is the so-called area formula (or change of variables). Recall that a C^1-diffeomorphism between open sets $U, V \subset \mathbb{R}^d$ is a C^1-function $f : U \to V$ that is bijective and such that its inverse f^{-1} is also a C^1-function.

Theorem 2.34 (Change of variables formula)**.** *Let $U, V \subset \mathbb{R}^d$ be open sets, $T : U \to V$ be a C^1-diffeomorphism and $f : V \to \mathbb{R}$ be measurable. Then*

$$f \in L^1(V) \iff (f \circ T) \cdot |\det(DT)| \in L^1(U)$$

and in this case

$$\int_{T(U)} f(x) \, \mathrm{d}x = \int_U f(T(x)) |\det(DT(x))| \, \mathrm{d}x.$$

Remark 2.35. There is a far-reaching generalization of the change of variables for Lipschitz functions T that are not necessarily one-to-one. It states then when $T : U \to \mathbb{R}^d$ is a Lipschitz mapping and $f : U \to \mathbb{R}$ is such that $f|\det(DT)| \in L^1(U)$[12] then

$$\int_U f(x)|\det(DT(x))|\,\mathrm{d}x = \int_{\mathbb{R}^d}\int_{T^{-1}(y)} f(x)\,\mathrm{d}\mu_{\#}(x)\,\mathrm{d}y,$$

where $\mu_{\#}$ is the counting measure. Such formulas belong to the field of geometric measure theory.

Coming back to the classical change of variables we discuss the well-known polar coordinates in $\mathbb{R}^2$.

Corollary 2.36. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be measurable. Define $T : (0,+\infty)\times(0,2\pi) \to \mathbb{R}^2$ by $T(r,\varphi) = (r\cos(\varphi), r\sin(\varphi))$ and $\widetilde{f} : (0,+\infty)\times(0,2\pi)\to\mathbb{R}$ by*

$$\widetilde{f}(r,\varphi) = f(r\cos(\varphi), r\sin(\varphi))\cdot r.$$

Then $f \in L^1(\mathbb{R}^2)$ if and only if $\widetilde{f} \in L^1((0,+\infty)\times(0,2\pi))$. Moreover, in this case

$$\int_{\mathbb{R}^2} f(x,y)\,\mathrm{d}(x,y) = \int_0^{2\pi}\int_0^{\infty} f(r\cos(\varphi), r\sin(\varphi))r\,\mathrm{d}r\,\mathrm{d}\varphi.$$

Proof. We apply Theorem 2.34 with $U = (0,+\infty)\times(0,2\pi)$ and $V := T(U) = \mathbb{R}^2\setminus\{(x,0) : x \geq 0\}$. By Lemma 2.26 (ii) it follows that $\mathbb{R}^2\setminus V$ is a null set, so that

$$f \in L^1(\mathbb{R}^2) \iff f \in L^1(V) \quad \text{and} \quad \int_{\mathbb{R}^2} f\,\mathrm{d}(x,y) = \int_V f\,\mathrm{d}(x,y).$$

Since T is a C^1-diffeomorphism from U to V (see Analysis 2) and $|\det(DT(r,\varphi))| = r$ the claim follows from Theorem 2.34 and Fubini's theorem. ☐

As a consequence of the polar coordinate transformation we can compute the limit of the error function. Note that setting $f(x,y) = e^{-(x^2+y^2)}$ the previous corollary implies that

$$\int_{\mathbb{R}^2} e^{-(x^2+y^2)}\,\mathrm{d}(x,y) = \int_0^{2\pi}\int_0^{\infty} e^{-r^2} r\,\mathrm{d}r\,\mathrm{d}\varphi = \int_0^{2\pi}[-\tfrac{1}{2}e^{-r^2}]_{r=0}^{\infty}\,\mathrm{d}\varphi = \pi.$$

On the other hand, since $e^{-(x^2+y^2)} = e^{-x^2}e^{-y^2}$ the left hand side can be simplified using Tonelli's theorem which yields

$$\int_{\mathbb{R}^2} e^{-(x^2+y^2)}\,\mathrm{d}(x,y) = \int_{\mathbb{R}} e^{-x^2}\int_{\mathbb{R}} e^{-y^2}\,\mathrm{d}y\,\mathrm{d}x = \left(\int_{\mathbb{R}} e^{-x^2}\,\mathrm{d}x\right)^2.$$

Thus $\int_{\mathbb{R}} e^{-x^2}\,\mathrm{d}x = \sqrt{\pi}$.

A further application concerns the volume of a disc. Setting $f(x,y) = \mathbb{1}_{\{x^2+y^2<1\}}$ we deduce that

$$\lambda_2(B_1(0)) = \int_{\mathbb{R}^2} f(x,y)\,\mathrm{d}(x,y) = \int_0^{2\pi}\int_0^{\infty}\mathbb{1}_{\{r<1\}}(r)r\,\mathrm{d}r\,\mathrm{d}\varphi = \int_0^{2\pi}\int_0^1 r\,\mathrm{d}r\,\mathrm{d}\varphi = \pi.$$

Another important change of variables concerns spherical coordinates in $\mathbb{R}^3$ as defined in the following corollary.

Corollary 2.37. *Let $f : \mathbb{R}^3 \to \mathbb{R}$ be measurable. Define $T : U := (0,+\infty)\times(0,\pi)\times(0,2\pi) \to \mathbb{R}^3$ by*

$$T(r,\theta,\varphi) = (r\sin(\theta)\cos(\varphi), r\sin(\theta)\sin(\varphi), r\cos(\theta)).$$

Then $f \in L^1(\mathbb{R}^3)$ if and only if $f(T(r,\theta,\varphi))\cdot r^2\sin(\theta) \in L^1(U)$ and in this case

$$\int_{\mathbb{R}^3} f\,\mathrm{d}(x,y,z) = \int_0^{2\pi}\int_0^{\pi}\int_0^{\infty} f(T(r,\theta,\varphi))r^2\sin(\theta)\,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}\varphi.$$

[12] A function $T : U \to \mathbb{R}^d$ is called Lipschitz continuous if there exists $L > 0$ such that $|T(x) - T(y)| \leq L|x-y|$ for all $x, y \in U$. One can prove that Lipschitz maps are differentiable a.e. so that taking the determinant of DT makes sense in a measure theoretic sense.

Proof. The argument is similar to the proof of Corollary 2.36, so we only sketch it. First note that $V := T(U)$ satisfies $\mathbb{R}^3 \setminus V = \{(x,0,z) \in \mathbb{R}^3 : x \geq 0, z \in \mathbb{R}\}$, so that it is a null set due to Lemma 2.26 (ii). Moreover, a direct calculation yields $|\det(DT(r,\theta,\varphi))| = r^2 \sin(\theta)$. From here on the proof is the same as for the polar coordinates. We skip the details. □

As a last application of the change of variables formula we show that a C^1-diffeomorphism $T : U \to V$ maps null sets to null sets[13]. Let $N \subset U$ be a null set. By the bijectivity of T we have that $T(x) \in T(N)$ if and only if $x \in N$, so that by the change of variables applied to $\mathbb{1}_N$ we obtain

$$\lambda_d(T(N)) = \int_{T(U)} \mathbb{1}_{T(N)}(x)\,\mathrm{d}x = \int_U \underbrace{\mathbb{1}_{T(N)}(T(x))|\det(DT(x))|\,\mathrm{d}x}_{=\mathbb{1}_N(x)} = \int_N |\det(DT(x))|\,\mathrm{d}x = 0.$$

2.4. **Convolution of integrable functions.** Convolution is an important operation on integrable functions, both from a 'pure mathematics' and an applied point of view. We first give the definition.

Definition 2.38 (Convolution)**.** Let $f, g \in L^1(\mathbb{R}^d)$. Then the convolution is defined as the Lebesgue-measurable function $f * g : \mathbb{R}^d \to \mathbb{R}$ given for a.e. $x \in \mathbb{R}^d$ by

$$(f * g)(x) = \int_{\mathbb{R}^d} f(y)g(x-y)\,\mathrm{d}y.$$

Remark 2.39. The fact that the convolution is well-defined for a.e. $x \in \mathbb{R}^d$ is part of the statement of Fubini's theorem. One can further show that the convolution does not depend on the representatives of f and g except on a negligible set (we omit these proofs). We will discuss more properties of the convolution in the lemma below.

Lemma 2.40. *Let $f, g, h \in L^1(\mathbb{R}^d)$. Then the convolution satisfies the following properties:*

(i) $f * g \in L^1(\mathbb{R}^d)$ *and* $\|f * g\|_1 \leq \|f\|_1\|g\|_1$*;*
(ii) $f * g = g * f$*;*
(iii) $f * (g * h) = (f * g) * h$*;*
(iv) $f * (g + h) = (f * g) + (f * h)$*;*
(v) $\alpha(f * g) = (\alpha f) * g$ *for all* $\alpha \in \mathbb{R}$*.*

If in addition $g \in C^1(\mathbb{R}^d)$ has compact support[14]*, then $f * g$ is differentiable and $\partial_i(f * g) = f * \partial_i g$ for all partial derivatives ∂_i.*

Proof. (i) As mentioned in the previous remark we do not prove the measurability of the convolution and the fact that it is independent of the choice of representatives. Hence we focus on the integrability. By definition we have

$$\|f * g\|_1 = \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} f(y)g(x-y)\,\mathrm{d}y \right| \mathrm{d}x \leq \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |f(y)||g(x-y)|\,\mathrm{d}y\,\mathrm{d}x = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |f(y)||g(x-y)|\,\mathrm{d}x\,\mathrm{d}y,$$

where we used Tonelli's theorem for non-negative functions to switch the order of integration in the last equality. Using then the change of variables $z = T(x) := x - y$ which satisfies $T(\mathbb{R}^d) = \mathbb{R}^d$ and $\det(DT(x)) = 1$, Theorem 2.34 implies that

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |f(y)||g(x-y)|\,\mathrm{d}x\,\mathrm{d}y = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |f(y)||g(z)|\,\mathrm{d}z\,\mathrm{d}y = \int_{\mathbb{R}^d} |f(y)| \int_{\mathbb{R}^d} |g(z)|\,\mathrm{d}z\,\mathrm{d}y = \|f\|_1\|g\|_1.$$

This concludes the proof of (i).
(ii) We apply the change of variables $z = T(y) = x - y$ which satisfies $T(\mathbb{R}^d) = \mathbb{R}^d$ and $|\det(DT(x))| = 1$, so that Theorem 2.34 yields

$$(f * g)(x) = \int_{\mathbb{R}^d} f(y)g(x-y)\,\mathrm{d}y = \int_{\mathbb{R}^d} f(x-z)g(z)\,\mathrm{d}z = (g * f)(x).$$

---

[13]This conclusion is a little cheating since this property is used to prove the change of variables formula.
[14]The support of a continuous function is defined as the closure of the set $\{g \neq 0\}$.

(iii) The calculations look complicated, but essentially it suffices to apply Fubini's theorem. Indeed, we have

$$\begin{aligned} f*(g*h)(x) &= \int_{\mathbb{R}^d} f(y) \underbrace{\int_{\mathbb{R}^d} g(z)h(x-y-z)\,\mathrm{d}z}_{=(g*h)(x-y)}\,\mathrm{d}y = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(y)g(z)h(x-y-z)\,\mathrm{d}y\,\mathrm{d}z \\ &\overset{z+y=:\omega}{=} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(\omega-z)g(z)h(x-\omega)\,\mathrm{d}\omega\,\mathrm{d}z = \int_{\mathbb{R}^d} \underbrace{\int_{\mathbb{R}^d} f(\omega-z)g(z)\,\mathrm{d}z}_{=(f*g)(\omega)}\, h(x-\omega)\,\mathrm{d}\omega \\ &= (f*g)*h(x) \end{aligned}$$

(iv) and (v) are a consequence of the linearity of the Lebesgue integral.

In order to prove the last statement, fix $x_0 \in \mathbb{R}^d$. Note that a C^1-function with compact support has a uniformly bounded derivative. We now apply Corollary 2.20 about the differentiability of parameter-dependent integrals. Define $\ell : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ by $\ell(x,y) = f(y)g(x-y)$. Then by assumption ℓ is integrable in y for every $x \in \mathbb{R}^d$[15] and differentiable with respect to x for all $y \in \mathbb{R}^d$. Moreover, the partial derivative is given by

$$\partial_{x_i}\ell(x,y) = f(y)\partial_i g(x-y).$$

We need to bound this derivative uniformly in $x \in \mathbb{R}^d$. Clearly $|f(y)\partial_i g(x-y)| \leq |f(y)|\|Dg\|_\infty$ and the right hand side is integrable since f is integrable by assumption. Thus the claim follows from Corollary 2.20. □

The convolution can also be defined on other spaces than $L^1(\mathbb{R}^d)$. For instance, using tools from functional analysis one can define the convolution of certain measures. In particular, it turns out that the Dirac measure δ_0 satisfies $f * \delta_0 = f$ for any reasonably nice function such that the convolution with measures is well-defined (for instance, $f \in C_c(\mathbb{R}^d)$). One powerful aspect of convolutions becomes evident when one weakens the notion of derivative (so-called distributional derivatives). Then one can try to solve PDEs by finding the so-called fundamental solution. Consider for instance the Poisson equation $-\Delta u = f$ on $\mathbb{R}^d$. Solving the PDE $-\Delta u_0 = \delta_0$ (the solution is called the fundamental solution) then suffices to treat the inhomogeneous case by setting $u_f = f * u_0$. Indeed, a formal calculation (which can be made rigorous in the framework of distributions) yields that

$$-\Delta u_f = -\Delta(f*u_0) = f*(-\Delta u_0) = f*\delta_0 = f.$$

In this subsection we only treat the case of L^p spaces with exponents different from 1.

Proposition 2.41 (Young's convolution inequality)**.** *Let $f \in L^p(\mathbb{R}^d)$ and $g \in L^q(\mathbb{R}^d)$. Let $1 \leq p,q \leq r \leq +\infty$ satisfy $1/p + 1/q = 1 + 1/r$. Then $f * g \in L^r(\mathbb{R}^d)$ and*

$$\|f*g\|_r \leq \|f\|_p\|g\|_q.$$

In particular, when $q = 1$ then $r = p$ and $\|f*g\|_p \leq \|f\|_p\|g\|_1$.

Proof. See exercise H 9.4. □

Next we take a look at an example to see the effect of the convolution. Let $f \in L^1(\mathbb{R}) \cap C(\mathbb{R})$ with primitive $F \in C^1(\mathbb{R})$ and set $g = \mathbb{1}_{[-1,0]}$. Then by definition

$$(f*g)(x) = \int_{\mathbb{R}} f(y) \underbrace{\mathbb{1}_{[-1,0]}(x-y)}_{=\mathbb{1}_{[x,x+1]}(y)}\,\mathrm{d}y = \int_x^{x+1} f(y)\,\mathrm{d}y = F(x+1) - F(x).$$

[15] Fubini's theorem only yields the integrability for a.e. $x \in \mathbb{R}^d$. However, since g is uniformly bounded it follows that the integral is finite for all $x \in \mathbb{R}^d$.

Hence in this case the convolution filters the behavior of the function f locally around x. Considering instead the function $g_\varepsilon = \frac{1}{\varepsilon}\mathbb{1}_{[-\varepsilon,0]}$ the same calculation shows that

$$(f * g_\varepsilon)(x) = \frac{F(x+\varepsilon) - F(x)}{\varepsilon} \overset{\varepsilon\to 0}{\to} F'(x) = f(x).$$

Thus g_ε is close to an identity for the convolution (and indeed, one can show that g_ε converges in a suitable sense to a Dirac measure). Now assume for the moment that the function g_ε would be regular. Then the convolution $(f * g_\varepsilon)$ has at least the same regularity as g_ε by Lemma 2.40, so that we could approximate f by more regular functions.

In order to find a smoother version of g_ε one can use the following method (also valid in higher dimensions): define $g, g_\varepsilon : \mathbb{R}^d \to \mathbb{R}$ by

$$g(x) = \begin{cases} c\, e^{-\frac{1}{1-|x|^2}} & \text{if } x \in B_1(0), \\ 0 & \text{otherwise}, \end{cases}, \qquad g_\varepsilon(x) = \varepsilon^{-d} g(\tfrac{x}{\varepsilon}), \tag{6}$$

where $c > 0$ is a constant chosen such that $\int_{\mathbb{R}^d} g \,\mathrm{d}x = 1$. Then one can show that $g, g_\varepsilon \in C_c^\infty(\mathbb{R}^d)$.

The next theorem shows that the convolution with g_ε with L^p-functions (with $p < +\infty$) provides indeed a smooth approximation. We just sketch the proof which relies on the non-trivial fact that every L^p-function can be approximated in the L^p-norm by $C_c(\mathbb{R}^d)$-functions.[16]

Theorem 2.42 (Smooth functions are dense in L^p for $p < +\infty$)**.** *Let $\Omega \subset \mathbb{R}^d$ be open and $1 \le p < +\infty$. Let $g_\varepsilon \in C_c^\infty(\mathbb{R}^d)$ be defined as in (6). For $f \in L^p(\Omega)$ define its extension to $\mathbb{R}^d$ by $\widetilde{f}(x) = 0$ for $x \in \mathbb{R}^d \setminus \Omega$. Then the sequence $(\widetilde{f} * g_\varepsilon)|_\Omega$ belongs to $L^p(\Omega) \cap C^\infty(\Omega)$ and*

$$\lim_{\varepsilon\to 0} \|(\widetilde{f} * g_\varepsilon)|_\Omega - f\|_{L^p(\Omega)} = 0.$$

Proof. Note that $\widetilde{f} \in L^p(\mathbb{R}^d)$ and by Lemma 2.40 and Proposition 2.41 we know that $\widetilde{f} * g_\varepsilon \in L^p(\mathbb{R}^d) \cap C^\infty(\mathbb{R}^d)$, so that $(\widetilde{f} * g_\varepsilon)|_\Omega \in L^p(\Omega) \cap C^\infty(\Omega)$. As announced we use without proof that for any $\delta > 0$ there exists a function $f_\delta \in C_c(\mathbb{R}^d)$ such that $\|\widetilde{f} - f_\delta\|_p < \delta$ (this is false for $p = +\infty$). Fix such a function f_δ. Next note that by the change of variables $y = x/\varepsilon$ we have $\mathrm{d}x = \varepsilon^d \mathrm{d}y$. Hence

$$\|g_\varepsilon\|_1 = \int_{\mathbb{R}^d} |g_\varepsilon(x)| \,\mathrm{d}x = \int_{\mathbb{R}^d} \varepsilon^{-d} g(\tfrac{x}{\varepsilon}) \,\mathrm{d}x = \int_{\mathbb{R}^d} g(y) \,\mathrm{d}y = 1.$$

Hence by the triangle inequality and Proposition 2.41 we have

$$\begin{aligned} \|(\widetilde{f} * g_\varepsilon)|_\Omega - f\|_{L^p(\Omega)} &\le \|(\widetilde{f} * g_\varepsilon) - \widetilde{f}\|_p \le \|(\widetilde{f} - f_\delta) * g_\varepsilon\|_p + \|f_\delta * g_\varepsilon - f_\delta\|_p + \|f_\delta - \widetilde{f}\|_p \\ &\le \underbrace{\|\widetilde{f} - f_\delta\|_p}_{\le\delta} \underbrace{\|g_\varepsilon\|_1}_{=1} + \|f_\delta * g_\varepsilon - f_\delta\|_p + \delta \le 2\delta + \|f_\delta * g_\varepsilon - f_\delta\|_p. \end{aligned}$$

We show that $\lim_{\varepsilon\to 0} \|f_\delta * g_\varepsilon - f_\delta\|_p$. Then the claim follows from the arbitrariness of $\delta > 0$.

Since $\int_{\mathbb{R}^d} g_\varepsilon \,\mathrm{d}x = 1$ we have

$$\|f_\delta * g_\varepsilon - f_\delta\|_p^p = \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} (f_\delta(y) - f_\delta(x)) g_\varepsilon(x-y) \,\mathrm{d}y \right|^p \mathrm{d}x \le \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |f_\delta(y) - f_\delta(x)| g_\varepsilon(x-y) \,\mathrm{d}y \right)^p \mathrm{d}x$$

[16] This fact relies in particular on a regularity property of the Lebesgue-measure that requires to treat its construction in detail.

Next note that $g_\varepsilon(x-y)=0$ for $y\notin B_\varepsilon(x)$. Hence the inner integral can be restricted to $B_\varepsilon(x)$, so that

$$\begin{aligned}\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}|f_\delta(y)-f_\delta(x)|g_\varepsilon(x-y)\,\mathrm{d}y\right)^p\mathrm{d}x &= \int_{\mathbb{R}^d}\left(\int_{B_\varepsilon(x)}|f_\delta(y)-f_\delta(x)|g_\varepsilon(x-y)\,\mathrm{d}y\right)^p\mathrm{d}x\\ &\le \int_{\mathbb{R}^d}\sup_{y\in B_\varepsilon(x)}|f_\delta(y)-f_\delta(x)|^p\Bigg(\underbrace{\int_{B_\varepsilon(x)}g_\varepsilon(x-y)\,\mathrm{d}y}_{=1}\Bigg)^p\mathrm{d}x\\ &\le \int_{\mathbb{R}^d}\sup_{y\in B_\varepsilon(x)}|f_\delta(y)-f_\delta(x)|^p\,\mathrm{d}x.\end{aligned}$$

Since f_δ has compact support (denoted by K), the last integral can be restricted to the bounded set $K+B_\varepsilon(0)\subset K+B_1(0)$ for $\varepsilon<1$. Moreover, being f uniformly continuous (since its support is compact), it follows that

$$\lim_{\varepsilon\to0}\sup_{x\in\mathbb{R}^d}\sup_{y\in B_\varepsilon(x)}|f_\delta(y)-f_\delta(x)|^p=0.$$

Hence

$$\int_{\mathbb{R}^d}\sup_{y\in B_\varepsilon(x)}|f_\delta(y)-f_\delta(x)|^p\,\mathrm{d}x\le\sup_{x\in\mathbb{R}^d}\sup_{y\in B_\varepsilon(x)}|f_\delta(y)-f_\delta(x)|^p\,\lambda_d(K+B_1(0))\overset{\varepsilon\to0}{\to}0.$$

This concludes the proof. □

3. The Fourier transform

In this third chapter we discuss the so-called Fourier transformation. While Fourier series represent periodic functions by a superposition of sin and cos with variable frequencies the Fourier transformation treats non-periodic functions. In general, a countable superposition of elementary waves is no longer sufficient, but we need a continuum of superpositions. The mathematically rigorous definition is given below. In what follows the Lebesgue integral of complex-valued functions is understood by integrating separately the real-and imaginary part.

Definition 3.1. Let $f\in L^1(\mathbb{R}^d)$. The Fourier-transform of f is the function $\mathcal{F}[f]:\mathbb{R}^d\to\mathbb{C}$ given by

$$\mathcal{F}[f](k)=\int_{\mathbb{R}^d}f(x)e^{-ik\cdot x}\,\mathrm{d}x,$$

where $k\cdot x$ denotes the scalar-product between the vectors $x,k\in\mathbb{R}^d$.

Remark 3.2. The definition of the Fourier transform varies in the literature. Sometimes you find the factor 2π in the exponential or also a normalizing factor in front of the integral. In any case the Fourier transform is well-defined for $f\in L^1(\mathbb{R}^d)$ since the exponential has modulus 1.

The Fourier transform is often interpreted as a function defined on frequencies and indeed this viewpoint is exploited in the field of signal processing. We will also see that some PDEs can be solved quite easily after Fourier transformation since differential operators are transformed to polynomial products. However, in order to get back to the relevant physical space one needs to define the inverse Fourier transformation. Before we come to that topic we collect some elementary properties of the Fourier transform.

Lemma 3.3. *Let $f\in L^1(\mathbb{R}^d)$, $h\in\mathbb{R}^d$ and $\lambda>0$. Then*

- *(i)* $f\mapsto\mathcal{F}[f]$ *is linear in* f*;*
- *(ii)* $\mathcal{F}[f]\in C(\mathbb{R}^d)$ *and* $\|\mathcal{F}[f]\|_\infty\le\|f\|_1$*;*
- *(iii)* $g(x):=f(x-h)\implies\mathcal{F}[g](k)=e^{-ik\cdot h}\mathcal{F}[f](k)$*;*
- *(iv)* $g(x):=e^{ih\cdot x}f(x)\implies\mathcal{F}[g](k)=\mathcal{F}[f](k-h)$*;*
- *(v)* $g(x):=f(\frac{x}{\lambda})\implies\mathcal{F}[g](k)=\lambda^d\mathcal{F}[f](\lambda k)$*.*

Proof. (i) follows from the linearity of the Lebesgue integral.
(ii) Continuity of the Fourier transformation was proven in H 8.2 b). The bound on the L^∞-norm follows by taking the supremum over $k \in \mathbb{R}^d$ in the estimate

$$|\mathcal{F}[f](k)| \leq \int_{\mathbb{R}^d} |f(x)| \underbrace{|e^{-ik\cdot x}|}_{=1} \,\mathrm{d}x = \|f\|_1.$$

(iii) By definition and the change of variables $T(x) = x - h$ we have

$$\mathcal{F}[g](k) = \int_{\mathbb{R}^d} f(x-h)e^{-ik\cdot x} \,\mathrm{d}x \overset{y:=x-h}{=} \int_{\mathbb{R}^d} f(y) \underbrace{e^{-ik\cdot(y+h)}}_{=e^{-ik\cdot y}e^{-ik\cdot h}} \,\mathrm{d}y = e^{-ik\cdot h} \int_{\mathbb{R}^d} f(y)e^{-ik\cdot y} \,\mathrm{d}y = e^{-ik\cdot h}\mathcal{F}[f](k).$$

(iv) By definition we find that

$$\mathcal{F}[g](k) = \int_{\mathbb{R}^d} f(x)e^{ih\cdot x}e^{-ik\cdot x} \,\mathrm{d}x = \int_{\mathbb{R}^d} f(x)e^{-ik\cdot(k-h)} \,\mathrm{d}x = \mathcal{F}[f](k-h).$$

(v) We apply the change of variables $y = T(x) = \frac{x}{\lambda}$ which yields $\mathrm{d}x = \lambda^d \mathrm{d}y$, so that

$$\mathcal{F}[g](k) = \int_{\mathbb{R}^d} f(\tfrac{x}{\lambda})e^{-ik\cdot x} \,\mathrm{d}x = \lambda^d \int_{\mathbb{R}^d} f(y)e^{-ik\cdot\lambda y} \,\mathrm{d}y = \lambda^d \mathcal{F}[f](\lambda k).$$

□

Next we treat some illustrative examples of the Fourier transform. You will prove the corresponding statements in the exercises.

Example 3.4. a) Let $A \in \mathbb{R}^{d\times d}$ be symmetric and positive definite. Then $f_A(x) = e^{-\frac{1}{2}x^T Ax}$ belongs to $L^1(\mathbb{R}^d)$ and

$$\mathcal{F}[f_A](k) = \frac{\sqrt{(2\pi)^d}}{\sqrt{\det(A)}} e^{-\frac{1}{2}k^T A^{-1} k}.$$

b) For $R > 0$ let $f_R : \mathbb{R} \to \mathbb{R}$ be defined by $f_R = \mathbb{1}_{[-R,R]} \in L^1(\mathbb{R})$. Then

$$\mathcal{F}[f_R](k) = \frac{2\sin(Rk)}{k}.$$

In particular, $\mathcal{F}[f_R] \notin L^1(\mathbb{R})$ (cf. H 9.1), so that in general the Fourier transform of an L^1-function is not necessarily integrable.

The fact that the Fourier transform does not belong to $L^1(\mathbb{R}^d)$ is the main difficulty to define the reverse Fourier transform. Before we discuss the inverse transformation we investigate the connection between differentiability and the decay at infinity of the Fourier transform, which also shows later why we are interested in an inverse operation.

Proposition 3.5. *Let $f \in L^1(\mathbb{R}^d) \cap C^m(\mathbb{R}^d)$ for some $m \in \mathbb{N}$. Assume that $\partial^\alpha f \in L^1(\mathbb{R}^d)$ for all multi-indices $\alpha \in \mathbb{N}_0^d$ with $|\alpha| \leq m$. Then*

$$\mathcal{F}[\partial^\alpha f](k) = i^{|\alpha|} k^\alpha \mathcal{F}[f](k),$$

where $|\alpha| := \sum_{i=1}^d \alpha_i$, $k^\alpha := \prod_{i=1}^d k_i^{\alpha_i}$ and $\partial^\alpha := \frac{\partial^{|\alpha|}}{(\partial x_1)^{\alpha_1} \ldots (\partial x_d)^{\alpha_d}}$.

Proof. We first show the statement when $d = m = 1$. By dominated convergence and integration by parts we have

$$\begin{aligned}\mathcal{F}[f'](k) &= \int_{\mathbb{R}} f'(x)e^{-ikx} \,\mathrm{d}x = \lim_{t\to+\infty} \int_{-t}^{t} f'(x)e^{-ikx} \,\mathrm{d}x \\ &= \lim_{t\to+\infty} \left(f(t)e^{-ikt} - f(-t)e^{ikt} + ik\int_{-t}^{t} f(x)e^{-ikx} \,\mathrm{d}x \right).\end{aligned}$$

Again by dominated convergence it holds that

$$\lim_{t\to+\infty} ik\int_{-t}^{t} f(x)e^{-ikx}\,\mathrm{d}x = ik\mathcal{F}[f](k).$$

Hence we deduce that

$$\mathcal{F}[f'](k) = \lim_{t\to+\infty}\left(f(t)e^{-ikt} - f(-t)e^{ikt}\right) + ik\mathcal{F}[f](k).$$

It suffices to prove that the limit equals 0. Since $f' \in L^1(\mathbb{R})$ it follows one more time from dominated convergence that there exists the limits

$$f(0) + \lim_{t\to\pm\infty}\int_0^t f'(x)\,\mathrm{d}x = \lim_{t\to\pm\infty} f(t).$$

Since $f \in L^1(\mathbb{R})$ the only possibility is that these limits[17] are zero. Since $|e^{\pm ikt}| = 1$ we conclude that

$$\lim_{t\to+\infty}\left(f(t)e^{-ikt} - f(-t)e^{ikt}\right) = 0.$$

Hence $\mathcal{F}[f'](k) = ik\mathcal{F}[f](k)$ as claimed.

The case $d = 1$ and $m > 1$ follows by iterating the above proof which yields an additional prefactor ik in every step.

Next we treat the case $d > 1$ and $m = 1$. In this case the only admissible differential operators are the partial derivatives ∂_j. Then we have by Fubini's theorem and the same reasoning as in the one-dimensional case that

$$\begin{aligned}\mathcal{F}[\partial_j f](k) &= \int_{\mathbb{R}^{d-1}}\int_{\mathbb{R}} \partial_j f(x)e^{-ik\cdot x}\,\mathrm{d}x_j\,\mathrm{d}(x_1,\dots,x_{j-1},x_{j+1},\dots,x_d)\\ &= \int_{\mathbb{R}^{d-1}}\int_{\mathbb{R}} ik_j f(x)e^{-ik\cdot x}\,\mathrm{d}x_j\,\mathrm{d}(x_1\dots,x_{j-1},x_{j+1},\dots,x_d) = ik_j\int_{\mathbb{R}^d} f(x)e^{-ik\cdot x}\,\mathrm{d}x\\ &= i^{|\alpha|}k^\alpha\mathcal{F}[f](k),\end{aligned}$$

where we used that $\alpha = e_j$.

The general case $m > 1$ follows again by iterating the above argument. □

Remark 3.6. One can prove that for $f \in L^1(\mathbb{R}^d)$ the Fourier transform vanishes at infinity, i.e.,

$$\lim_{|k|\to+\infty}\mathcal{F}[f](k) = 0.$$

The previous proposition yields a more quantitative decay depending on the smoothness of f. In particular, combining Lemma 3.3 (ii) and Proposition 3.5 we deduce that any $f \in L^1(\mathbb{R}^d)\cap C^m(\mathbb{R}^d)$ with all derivatives also in $L^1(\mathbb{R}^d)$ satisfies

$$|\mathcal{F}[f](k)| \leq C\frac{\|D^m f\|_1}{\|k\|^m}.$$

The next result shows that when f decays with a certain rate at infinity then its Fourier transform possesses derivatives up to a certain order.

Proposition 3.7. *Let $f \in L^1(\mathbb{R}^d)$ and $m \in \mathbb{N}$. Assume that $x \mapsto g_\alpha(x) = x^\alpha f(x) \in L^1(\mathbb{R}^d)$ for all multi-indices $\alpha \in \mathbb{N}_0^d$ with $|\alpha| \leq m$. Then $\mathcal{F}[f] \in C^m(\mathbb{R}^d)$ and $\partial^\alpha\mathcal{F}[f] = (-i)^{|\alpha|}\mathcal{F}[g_\alpha]$.*

Proof. We apply Corollary 2.20 about the differentiability of parameter-dependent integrals. Note that the partial derivative with respect to k_j of the integrand is given by

$$\partial_{k_j}\left(f(x)e^{-ik\cdot x}\right) = -ix_j f(x)e^{-ik\cdot x}.$$

[17] In general integrability does not imply that the limits at $\pm\infty$ are zero since they might not even exist, but if the limits exist then they have to be zero.

Since $|e^{-ik\cdot x}| = 1$ we have that $|\partial_{k_j}(f(x)e^{-ik\cdot x})| \le |x_j||f(x)|$ which is independent of k and integrable by assumption. Hence Corollary 2.20 yields that

$$\partial_{k_j}\mathcal{F}[f](k) = -i\int_{\mathbb{R}^d} x_j f(x)e^{-ik\cdot x}\,\mathrm{d}x = -i\mathcal{F}[g_{e_j}](k).$$

Due to Lemma 3.3 (ii) the right hand side is a continuous function. Thus $\mathcal{F}[f] \in C^1(\mathbb{R}^d)$ and the derivative has the claimed structure. For higher order derivatives we repeat the above reasoning to conclude the proof. □

In order to apply the relation between Fourier transform and derivatives we consider a linear partial differential equation with constant coefficients, i.e., of the form

$$\sum_{\substack{\alpha\in\mathbb{N}_0^d\\ |\alpha|\le m}} c_\alpha \partial^\alpha u(x) = f(x) \qquad \text{on } \mathbb{R}^d.$$

Taking the Fourier transform of the equation then leads to the algebraic equation

$$\Big(\sum_{\substack{\alpha\in\mathbb{N}_0^d\\ |\alpha|\le m}} c_\alpha i^{|\alpha|}k^\alpha\Big)\mathcal{F}[u](k) = \mathcal{F}[f](k).$$

The polynomial $\sum_\alpha c_\alpha i^{|\alpha|}k^\alpha$ is called the symbol of the differential operator. If it is not degenerate (this depends on the structure of the differential operator) then we can solve for $\mathcal{F}[u]$ by dividing the above equality by the symbol. In order to obtain the solution u we have to invert the Fourier transformation. A similar procedure can be used for time-dependent PDEs for which the spatial[18] Fourier transformation yields an ODE instead of a purely algebraic equation. In both cases we need to undo the Fourier transform. Hence we next discuss how to invert the Fourier transformation.

Definition 3.8. Let $f \in L^1(\mathbb{R}^d)$. Then the inverse Fourier transform $\mathcal{F}^{-1}[f] : \mathbb{R}^d \to \mathbb{C}$ of f is defined by

$$\mathcal{F}^{-1}[f](x) = \frac{1}{(2\pi)^d}\mathcal{F}[f](-x) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} f(k)e^{ik\cdot x}\,\mathrm{d}k.$$

It is at the moment unclear that $\mathcal{F}^{-1}$ is the inverse operator of the Fourier transform. And indeed, there are some obstructions. For instance, to define $\mathcal{F}^{-1}[\mathcal{F}[f]]$ we need that $\mathcal{F}[f] \in L^1(\mathbb{R}^d)$, which we have seen is not true in general. Next, we know that $\mathcal{F}^{-1}[\mathcal{F}[f]]$ is a bounded, continuous function. This cannot be true for every function $f \in L^1(\mathbb{R}^d)$. The Fourier inversion theorem can be formulated in several ways. There are versions which make assumptions on $\mathcal{F}[f]$ (integrability) or which define the Fourier transformation on smaller spaces (which then ensure that the additional assumption is automatically satisfied). We present both versions.

Before we state the inversion theorems we prove an auxiliary lemma.

Lemma 3.9. *Let $f \in L^1(\mathbb{R}^d)$ be such that $\mathcal{F}[f] \in L^1(\mathbb{R}^d)$. For $L \in \mathbb{N}$ define the function $\delta_L : \mathbb{R}^d \to \mathbb{R}$ by $\delta_L(x) = \frac{L^d}{\sqrt{(2\pi)^d}}e^{-\frac{1}{2}L^2\|x\|^2}$. Then for all $x \in \mathbb{R}^d$ we have*

$$\lim_{L\to+\infty}(f*\delta_L)(x) = \mathcal{F}^{-1}[\mathcal{F}[f]](x).$$

Proof. By Example 3.4 it holds that $\mathcal{F}[\delta_1](x) = \sqrt{(2\pi)^d}\delta_1(x)$, so that

$$\delta_L(x) = \delta_L(-x) = L^d\delta_1(-Lx) = \frac{L^d}{\sqrt{(2\pi)^d}}\mathcal{F}[\delta_1](-Lx) = \frac{L^d}{(2\pi)^d}\int_{\mathbb{R}^d} e^{-\frac{1}{2}\|y\|^2}e^{iy\cdot Lx}\,\mathrm{d}y.$$

Using the change of variables $k = Ly$, we obtain

$$\delta_L(x) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{-\frac{\|k\|^2}{2L^2}}e^{ix\cdot k}\,\mathrm{d}k.$$

---

[18]One could also try a full Fourier transform but most time-dependent equations are such that the full symbol of the differential operator is degenerate.

Hence by Fubini's theorem we have

$$(f*\delta_L)(x) = \int_{\mathbb{R}^d} f(y)\delta_L(x-y)\,\mathrm{d}y = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} f(y)\int_{\mathbb{R}^d} e^{-\frac{\|k\|^2}{2L^2}} e^{i(x-y)\cdot k}\,\mathrm{d}k$$
$$= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \underbrace{\int_{\mathbb{R}^d} f(y)e^{-ik\cdot y}\,\mathrm{d}y}_{=\mathcal{F}[f](k)}\, e^{-\frac{\|k\|^2}{2L^2}} e^{ix\cdot k}\,\mathrm{d}k = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \underbrace{\mathcal{F}[f](k)\, e^{-\frac{\|k\|^2}{2L^2}} e^{ix\cdot k}}_{=:g_L(k)}\,\mathrm{d}k.$$

Note that $g_L(k) \to \mathcal{F}[f](k)e^{ix\cdot k}$ as $L \to +\infty$. Moreover, $|g_L(k)| \leq |\mathcal{F}[f](k)|$ and the last function is integrable by assumption. Hence by Lebesgue's dominated convergence theorem we deduce that

$$(f*\delta_L)(x) \to \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \mathcal{F}[f](k)e^{ix\cdot k}\,\mathrm{d}k = \mathcal{F}^{-1}[\mathcal{F}[f]](x).$$

□

In order to prove a version of the Fourier inversion theorem we need to show that $(f*\delta_L)(x) \to f(x)$ as $L \to +\infty$. To motivate why this can be true, note that δ_L decays exponentially for $x \gg 1/L$ and blows up like L^d if $\|x\| \leq 1/L$. Hence the behavior is similar to the function g_ε constructed to prove Theorem 2.42 (except that its support is not compact). However, in Theorem 2.42 we proved convergence in L^p but not pointwise. Indeed, we cannot expect pointwise convergence for every $x \in \mathbb{R}^d$ except when the function f is continuous. In order to state the result we use another notation from Lebesgue integration.

Definition 3.10. Let $f : \mathbb{R}^d \to \mathbb{R}$ be Lebesgue-integrable on every compact set $K \subset \mathbb{R}^d$. We say that $x_0 \in \mathbb{R}^d$ is a Lebesgue point of f if

$$\lim_{r\to 0}\frac{1}{r^d}\int_{B_r(x_0)} |f(y) - f(x_0)|\,\mathrm{d}y = 0.$$

Note that when f is continuous in x_0 then x_0 is a Lebesgue point of f. In the general case we have the following powerful result that we do not prove in this course.

Theorem 3.11 (Lebesgue's differentiation theorem)**.** *Let $f : \mathbb{R}^d \to \mathbb{R}$ be Lebesgue-integrable on every compact subset $K \subset \mathbb{R}^d$. Then a.e. $x_0 \in \mathbb{R}^d$ is a Lebesgue point of f.*

With the help of Lebesgue's differentiation theorem we can state and prove the first version of the Fourier inversion theorem.

Theorem 3.12 (Fourier inversion theorem; L^1-version)**.** *Let $f \in L^1(\mathbb{R}^d)$ be such that $\mathcal{F}[f] \in L^1(\mathbb{R}^d)$. Then $\mathcal{F}^{-1}[\mathcal{F}[f]](x) = f(x)$ for every Lebesgue point of f. In particular, $\mathcal{F}^{-1}[\mathcal{F}[f]] = f$ a.e. and thus as elements in $L^1(\mathbb{R}^d)$. If f is in addition continuous, then $\mathcal{F}^{-1}[\mathcal{F}[f]](x) = f(x)$ for all $x \in \mathbb{R}^d$.*

Proof. By Lemma 3.9 it suffices to prove that $\lim\limits_{L\to+\infty}(f*\delta_L)(x_0) = f(x_0)$ for every Lebesgue point x_0 of f, where $\delta_L \in L^1(\mathbb{R}^d)\cap C^\infty(\mathbb{R}^d)$ is defined in Lemma 3.9. Thus fix a Lebesgue point $x_0 \in \mathbb{R}^d$ of f. Due to Young's convolution inequality also $f(x_0)*\delta_L$ is well-defined and

$$(f(x_0)*\delta_L)(x_0) = f(x_0)\int_{\mathbb{R}^d}\delta_L(x_0-y)\,\mathrm{d}y = f(x_0)\int_{\mathbb{R}^d}\delta_L(y)\,\mathrm{d}y = f(x_0)\underbrace{\int_{\mathbb{R}^d}\delta_1(y)\,\mathrm{d}y}_{=1} = f(x_0).$$

Thus it suffices to prove that $\lim\limits_{L\to+\infty}((f-f(x_0))*\delta_L)(x_0) = 0$. First we show that asymptotically we can restrict the convolution integral to a very small set. Indeed, for any $r_0 > 0$ we have

$$((f-f(x_0))*\delta_L(x_0) = \int_{B_{r_0}(0)}(f(x_0-y)-f(x_0))\delta_L(y)\,\mathrm{d}y + \int_{\|y\|\geq r_0}(f(x_0-y)-f(x_0))\delta_L(y)\,\mathrm{d}y.$$

As we will see the second integral vanishes when $L \to +\infty$. Indeed, for any $y \in \mathbb{R}^d$ with $\|y\| \geq r_0$ we have

$$|f(x_0-y)-f(x_0)|\delta_L(y)| \leq |f(x_0-y)-f(x_0)|\underbrace{L^d e^{-\frac{1}{4}L^2 r_0^2}}_{=:a_L} e^{-\frac{1}{4}\|y\|^2} \leq \left(\sup_{L\in\mathbb{N}} a_L\right)\underbrace{|f(x_0-y)-f(x_0)|e^{-\frac{1}{4}\|y\|^2}}_{\in L^1(\mathbb{R}^d)}.$$

Since $a_L \to 0$ when $L \to +\infty$ we obtain that $(f(x_0-y)-f(x_0))\delta_L(y)$ converges to 0 pointwise on $\mathbb{R}^d \setminus B_{r_0}(0)$. Moreover, since a_L is in particular bounded, we can apply Lebesgue's dominated convergence theorem to deduce that for any $r_0 > 0$

$$\lim_{L\to+\infty} \int_{\|y\|\geq r_0} (f(x_0-y)-f(x_0))\delta_L(y)\,\mathrm{d}y = 0.$$

Hence it suffices to prove that

$$\lim_{r_0\to 0}\lim_{L\to+\infty} \int_{B_{r_0}(0)} (f(x_0-y)-f(x_0))\delta_L(y)\,\mathrm{d}y = 0.$$

To this end, let us assume that $r_0 = 2^{-k}$ for some $k \in \mathbb{N}$ and divide the ball $B_{2^{-k}}(0)$ in countably many annuli setting $A_n = \overline{B_{2^{n+1}}(0)} \setminus B_{2^n}(0)$ for $n \in \mathbb{Z}$, so that

$$\overline{B_{2^{-k}}(0)} = \{0\} \cup \bigcup_{n<-k} A_n.$$

Since $\partial B_{2^{-k}}(0)$ is a null set, we have

$$\begin{aligned}\left|\int_{B_{r_0}(0)} (f(x_0-y)-f(x_0))\delta_L(y)\,\mathrm{d}y\right| &\leq \int_{B_{r_0}(0)} |f(x_0-y)-f(x_0)|\delta_L(y)\,\mathrm{d}y \\ &\leq \sum_{n<-k}\int_{A_n} |f(x_0-y)-f(x_0)|\delta_L(y)\,\mathrm{d}y.\end{aligned}$$

Note that on each annulus A_n we have that $\delta_L(y) \leq L^d e^{-\frac{1}{2}L^2 4^n}$, so that

$$\begin{aligned}\int_{A_n} |f(x_0-y)-f(x_0)|\delta_L(y)\,\mathrm{d}y &\leq L^d e^{-\frac{1}{2}L^2 4^n} \int_{B_{2^{n+1}}(0)} |f(x_0-y)-f(x_0)|\,\mathrm{d}y \\ &\leq (2^{n+1}L)^d e^{-\frac{1}{8}(L2^{n+1})^2} \frac{1}{(2^{n+1})^d}\int_{B_{2^{n+1}}(0)} |f(x_0-y)-f(x_0)|\,\mathrm{d}y.\end{aligned}$$

Inserting this estimate in the previous one and using that $2^{n+1} \leq r_0$ for $n < -k$, we obtain by a change of variables that that

$$\left|\int_{B_{r_0}(0)} (f(x_0-y)-f(x_0))\delta_L(y)\,\mathrm{d}y\right| \leq \left(\sum_{n<-k} (2^{n+1}L)^d e^{-\frac{1}{8}(L2^{n+1})^2}\right) \sup_{r\leq r_0}\frac{1}{r^d}\int_{B_r(x_0)} |f(y)-f(x_0)|\,\mathrm{d}y.$$

The term outside of the brackets is independent of L and since x_0 is a Lebesgue point of f it follows that

$$\lim_{r_0\to 0}\sup_{r\leq r_0}\frac{1}{r^d}\int_{B_r(x_0)} |f(y)-f(x_0)|\,\mathrm{d}y = 0.$$

Therefore it suffices to prove that the term in brackets remains bounded when $L \to +\infty$ and $r_0 \to 0$ (the latter is equivalent to $k \to +\infty$). We achieve this by comparison to an integral. Note that

$$\begin{aligned}\lambda_d(LA_{n+1}) &= L^d\lambda_d(A_{n+1})) = L^d\left(\lambda_d(B_{2^{n+2}}(0)) - \lambda_d(B_{2^{n+1}}(0))\right) \\ &= L^d\lambda_d(B_1(0))\left((2^{n+2})^d - (2^{n+1})^d\right) = L^d\lambda_d(B_1(0))(2^d-1)(2^{n+1})^d \geq \lambda_d(B_1(0))(2^{n+1}L)^d\end{aligned}$$

and for $x \in LA_{n+1}$ it holds that $\|x\| \leq L2^{n+2}$, so that

$$e^{-\frac{1}{32}\|x\|^2} \geq e^{-\frac{(L2^{n+1})^2}{8}}.$$

Thus we can write

$$\sum_{k<-n}(2^{n+1}L)^d e^{-\frac{1}{8}(L2^{n+1})^2} \leq \frac{1}{\lambda_d(B_1(0))}\sum_{n<-k}\int_{LA_{n+1}} e^{-\frac{1}{32}\|x\|^2}\,\mathrm{d}x \leq \frac{1}{\lambda_d(B_1(0))}\int_{\mathbb{R}^d} e^{-\frac{1}{32}\|x\|^2}\,\mathrm{d}x < +\infty$$

independently of L and r_0. This proves the claim □

After proving the first version of the Fourier inversion theorem we give a sufficient condition for the assumptions to be satisfied.

Corollary 3.13. *Let $f \in L^1(\mathbb{R}^d) \cap C^m(\mathbb{R}^d)$ with $m \geq d+1$ and such that $\partial^\alpha f \in L^1(\mathbb{R}^d)$ for all multi-indices $\alpha \in \mathbb{N}_0^d$ with $|\alpha| \leq m$. Then $\mathcal{F}[f] \in L^1(\mathbb{R}^d)$ and $\mathcal{F}^{-1}[\mathcal{F}[f]](x) = f(x)$ for all $x \in \mathbb{R}^d$.*

Proof. Combine exercise H 9.3 with Theorem 3.12. □

Note that under those assumptions the Fourier transform is not necessarily m-times differentiable itself. Later we will introduce a space that is mapped by the Fourier transform onto itself.

But first we present an application (further examples will be treated in the exercises). Consider the heat equation on $\mathbb{R}^d$. More precisely, we look for solutions of the PDE

$$\begin{cases} \partial_t u(t,x) = D\Delta u(t,x) & \forall (t,x) \in (0,+\infty) \times \mathbb{R}^d, \\ u(0,x) = u_0(x) & \forall x \in \mathbb{R}^d. \end{cases}$$

Here $D > 0$ is the so-called diffusion coefficient and the Laplace operator is the second order differential operator given by $\Delta = \sum_{j=1}^d \partial_j^2$. We assume that the initial condition satisfies $u_0 \in L^1(\mathbb{R}^d)$. Taking the spatial Fourier transform of the heat equation we obtain the ODE[19]

$$\begin{cases} \partial_t \mathcal{F}[u(t,\cdot)](k) = -D\|k\|^2 \mathcal{F}[u(t,\cdot)](k) & \forall (t,k) \in (0,+\infty) \times \mathbb{R}^d, \\ \mathcal{F}[u(0,\cdot)](k) = \mathcal{F}[u_0](k) & \forall x \in \mathbb{R}^d. \end{cases}$$

Here we see k as a parameter. The solution is given by

$$\mathcal{F}[u(t,\cdot)](k) = \mathcal{F}[u_0](k) e^{-tD\|k\|^2} = \mathcal{F}[u_0](k)\mathcal{F}[G(t,\cdot)](k),$$

where, according to Example 3.4, we have

$$G(t,x) = \frac{1}{\sqrt{(4\pi t D)^d}} e^{-\frac{\|x\|^2}{4tD}}.$$

From exercise H 9.2 we know that $\mathcal{F}[f * g] = \mathcal{F}[f]\mathcal{F}[g]$, so we deduce that

$$\mathcal{F}[u](t,k) = \mathcal{F}[u_0 * G(t,\cdot)](k).$$

In order to apply the inverse Fourier transform, we need to check that $\mathcal{F}[u_0 * G(t,\cdot)] \in L^1(\mathbb{R}^d)$. The assumption that $u_0 \in L^1(\mathbb{R}^d)$ ensures that $\mathcal{F}[u_0] \in L^\infty(\mathbb{R}^d)$ by Lemma 3.3. Since $\mathcal{F}[f*g] = \mathcal{F}[f]\mathcal{F}[g]$ and $\mathcal{F}[G(t,\cdot)] \in L^1(\mathbb{R}^d)$, Hölder's inequality implies the claim. Hence we can apply Theorem 3.12 to deduce that

$$u(t,x) = (u_0 * G(t,\cdot))(x) = \frac{1}{\sqrt{(4\pi t D)^d}} \int_{\mathbb{R}^d} u_0(y) e^{-\frac{\|x-y\|^2}{4tD}} \,\mathrm{d}y,$$

which provides a semi-explicit formula for the solution.

3.1. **The Fourier transform in the Schwartz space.** In this subsection we introduce a space that is invariant under the Fourier transform. It is called the Schwartz space and the elements are sometimes called Schwartz functions or functions in the Schwartz class. Those are smooth functions that decay rapidly at infinity as classified in the definition below.

Definition 3.14. The Schwartz space $\mathscr{S}(\mathbb{R}^d)$ is defined as

$$\mathscr{S}(\mathbb{R}^d) = \{f \in C^\infty(\mathbb{R}^d;\mathbb{C}) : \forall \alpha,\beta \in \mathbb{N}_0^d : \sup_{x\in\mathbb{R}^d} |x^\alpha \partial^\beta f(x)| < +\infty\}.$$

Note that the set of compactly supported smooth functions $C_c^\infty(\mathbb{R}^d)$ is contained in the Schwartz space. Another example is given by the function $f(x) = e^{-x^T A x}$, where $A \in \mathbb{R}^{d\times d}$ is symmetric and positive definite. In the next lemma we collect some elementary properties of $\mathscr{S}(\mathbb{R}^d)$ whose proof is left to the interested reader.

[19] Here we assume that the time derivative commutes with the Fourier transform. This cannot be justified without any a priori assumptions on the solution. However, although being formal at this point, at the very end we obtain a solution for which this procedure can be justified a posteriori.

Lemma 3.15. *Let $f, g \in \mathscr{S}(\mathbb{R}^d)$, $c \in \mathbb{C}$ and $\alpha \in N_0^d$. Then the following functions belong also to $\mathscr{S}(\mathbb{R}^d)$:*

(i) $x \mapsto \overline{f(x)}$;
(ii) $x \mapsto f(x)g(x)$;
(iii) $x \mapsto cf(x) + g(x)$;
(iv) $x \mapsto x^\alpha f(x)$;
(v) $x \mapsto \partial^\alpha f(x)$.

As shown in the lemma below a Schwartz function belongs to $L^p(\mathbb{R}^d)$ for every $p \in [1, +\infty]$.

Lemma 3.16. *It holds that $\mathscr{S}(\mathbb{R}^d) \subset L^p(\mathbb{R}^d)$ for any $p \in [1, +\infty]$.*

Proof. First note that the case $p = +\infty$ is covered by the definition for the multi-indices $\alpha = \beta = 0$. Hence assume that $p \in [1, +\infty)$. As proven in the solution of H 9.3 the function $g_m(x) = \mathbb{1}_{\{\|x\| \geq 1\}}(x)\|x\|^{-m}$ is integrable on $\mathbb{R}^d$ as soon as $m > d$. Due to Hölder's inequality we have for any $f \in \mathscr{S}(\mathbb{R}^d)$ that

$$\begin{aligned}\|f\|_p &= \|f\mathbb{1}_{B_1(0)} + f\mathbb{1}_{\{\|x\|\geq 1\}}\|_p \leq \|f\mathbb{1}_{B_1(0)}\|_p + \|f\mathbb{1}_{\{\|x\|\geq 1\}}\|_p \\ &\leq \|f\|_\infty \|\mathbb{1}_{B_1(0)}\|_p + \underbrace{\left(\sup_{x\in\mathbb{R}^d} \|x\|^m |f(x)|\right)}_{=:C_{f,m}} \|g_m\|_p = \|f\|_\infty \lambda_d(B_1(0))^{\frac{1}{p}} + C_{f,m}\|g_m\|_p.\end{aligned}$$

In order to conclude the proof, it suffices to show that $g_m \in L^p(\mathbb{R}^d)$. To this end, note that $|g_m|^p = g_{pm}$ and $pm \geq m > d$. This proves the claim. □

The next result is about the density of Schwartz functions in $L^p(\mathbb{R}^d)$. As they are smooth, we have to exclude the case $p = +\infty$.

Proposition 3.17. *Let $1 \leq p < +\infty$. Then for every $f \in L^p(\mathbb{R}^d)$ there exists a sequence $(f_n)_{n\in\mathbb{N}} \subset \mathscr{S}(\mathbb{R}^d)$ such that $\|f - f_n\|_p \to 0$ as $n \to +\infty$.*

Proof. Arguing separately for the real- and imaginary part, we may assume that f is real-valued. In the exercises it is shown that we can even take $f_n \in C_c^\infty(\mathbb{R}^d)$. This yields the claim. □

The previous proposition will be useful to define the Fourier transform on the Hilbert space $L^2(\mathbb{R}^d)$ which is the setting most relevant for applied mathematics. Before we can switch to this setting we need to understand the Fourier transform on $\mathscr{S}(\mathbb{R}^d)$.

Theorem 3.18 (Fourier inversion theorem; Schwartz space version)**.** *For every $f \in \mathscr{S}(\mathbb{R}^d)$ it holds that $\mathcal{F}[f] \in \mathscr{S}(\mathbb{R}^d)$. Moreover, the Fourier transform $\mathcal{F} : \mathscr{S}(\mathbb{R}^d) \to \mathscr{S}(\mathbb{R}^d)$ is bijective.*

Proof. As $\mathscr{S}(\mathbb{R}^d) \subset L^1(\mathbb{R}^d)$ the Fourier transform is well-defined on $\mathscr{S}(\mathbb{R}^d)$. Moreover, by Lemma 3.15 it holds that $x \mapsto x^\alpha f(x) \in \mathscr{S}(\mathbb{R}^d) \subset L^1(\mathbb{R}^d)$. Thus Proposition 3.7 yields that $\mathcal{F}[f] \in C^\infty(\mathbb{R}^d; \mathbb{C})$. Moreover, for any two multi-indices $\alpha, \beta \in \mathbb{N}_0^d$ we have the formula

$$k^\alpha \partial^\beta \mathcal{F}[f](k) = k^\alpha (-i)^{|\beta|}\mathcal{F}[x \mapsto x^\beta f(x)](k) = (-i)^{|\alpha|+|\beta|}\mathcal{F}[x \mapsto \partial^\alpha \left(x^\beta f(x)\right)](k).$$

Since $x \mapsto \partial^\alpha \left(x^\beta f(x)\right) \in \mathscr{S}(\mathbb{R}^d) \subset L^1(\mathbb{R}^d)$ by Lemma 3.15, it follows by Lemma 3.3 that its Fourier transform belongs to $L^\infty(\mathbb{R}^d)$. Hence it is bounded, which shows that also $k^\alpha \partial^\beta \mathcal{F}[f](k)$ is bounded, so that $\mathcal{F}[f] \in \mathscr{S}(\mathbb{R}^d)$. In order to prove that $\mathcal{F}$ is bijective, assume that $\mathcal{F}[f] = \mathcal{F}[g]$ for some $f, g \in \mathscr{S}(\mathbb{R}^d)$. Since $\mathcal{F}[f] \in \mathscr{S}(\mathbb{R}^d) \subset L^1(\mathbb{R}^d)$ we can take the inverse Fourier transform and deduce from Theorem 3.12 that $f = g$ since both functions are continuous. Next let $h \in \mathscr{S}(\mathbb{R}^d)$ be given. Set $f = \mathcal{F}^{-1}[h]$. Since the inverse Fourier transform is given by $(2\pi)^{-d}\mathcal{F}[\cdot](-x)$, it follows again that $f \in \mathscr{S}(\mathbb{R}^d)$. Moreover, by Theorem 3.12 we have[20]

[20]Note that we cannot use $\mathcal{F} \circ \mathcal{F}^{-1} = \mathrm{Id}$ since the Fourier inversion theorem in L^1 only showed that $\mathcal{F}^{-1} \circ \mathcal{F} = \mathrm{Id}$ (and this only under certain assumptions). In infinite dimensions this does not automatically imply the first statement.

$$\mathcal{F}[f](k) = \mathcal{F}[\mathcal{F}^{-1}[h]](k) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{-ik\cdot x}\int_{\mathbb{R}^d} e^{i\omega\cdot x}h(\omega)\,\mathrm{d}\omega\,\mathrm{d}x$$
$$\overset{T(x)=-x}{=} \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{ik\cdot y}\int_{\mathbb{R}^d} e^{-i\omega\cdot y}h(\omega)\,\mathrm{d}\omega\,\mathrm{d}y = \mathcal{F}^{-1}[\mathcal{F}[h]](k) = h(k).$$

□

Next we discuss a geometric property of the Fourier transform with regard to the L^2-scalar product.

Proposition 3.19 (Plancherel theorem in $\mathscr{S}(\mathbb{R}^d)$)**.** *Let $f, g \in \mathscr{S}(\mathbb{R}^d)$. Setting $\langle f, g\rangle = \int_{\mathbb{R}^d}\overline{f(x)}g(x)\,\mathrm{d}x$, it holds that*

$$(2\pi)^{-d}\langle\mathcal{F}[f], \mathcal{F}[g]\rangle = \langle f, g\rangle$$

Proof. By definition we have

$$\langle\mathcal{F}[f], \mathcal{F}[g]\rangle = \int_{\mathbb{R}^d}\overline{\mathcal{F}[f](k)}\mathcal{F}[g](k)\,\mathrm{d}k = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\overline{f(x)}e^{ik\cdot x}\,\mathrm{d}x\mathcal{F}[g](k)\,\mathrm{d}k.$$

Note that the integrand belongs to $L^1(\mathbb{R}^{2d})$ since $|e^{ik\cdot x}| = 1$ and $\overline{f}, \mathcal{F}[g] \in \mathscr{S}(\mathbb{R}^d) \subset L^1(\mathbb{R}^d)$.[21] Hence we can apply Fubini's theorem to deduce that

$$(2\pi)^{-d}\langle\mathcal{F}[f], \mathcal{F}[g]\rangle = (2\pi)^{-d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} e^{ik\cdot x}\mathcal{F}[g](k)\,\mathrm{d}k\overline{f(x)}\,\mathrm{d}x = \int_{\mathbb{R}^d} g(x)\overline{f(x)}\,\mathrm{d}x,$$

where in the last equality we applied Theorem 3.18. The last term equals $\langle f, g\rangle$ by definition. □

Note that the prefactor in the above result can be avoided if we put the prefactor $\sqrt{(2\pi)^{-d}}$ in front of the Fourier transform and its inverse (for the latter instead of the prefactor $(2\pi)^{-d}$). But then one gains a prefactor for instance in the identity $\mathcal{F}[f * g] = \mathcal{F}[f]\mathcal{F}[g]$. Every convention has its advantages and disadvantages.

3.2. **The Fourier transform on $L^2(\mathbb{R}^d)$.** The Fourier transform on $\mathscr{S}(\mathbb{R}^d)$ is not always suitable for applications since it imposes strong assumptions on the functions involved. In particular, any discontinuities are ruled out by the smoothness. Hence in what follows we extend the Fourier transform to an isometry on the space of complex-valued square-integrable functions. This is achieved via an abstract argument since the classical definition of the Fourier transform is not well-defined for $f \in L^2(\mathbb{R}^d)$ (the exponential $e^{ik\cdot x}$ belongs to $L^\infty(\mathbb{R}^d)$ but it does not decay at infinity).

Theorem 3.20. *The Fourier transform $\mathcal{F}$ can be extended in a unique way to a continuous, linear and bijective operator $\mathcal{F} : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ such that*

$$(2\pi)^{-d}\langle\mathcal{F}[f], \mathcal{F}[g]\rangle = \langle f, g\rangle \qquad \forall f, g \in L^2(\mathbb{R}^d).$$

Proof. The argument is part of a more general theory about the extension of linear, continuous operators. Let $f \in L^2(\mathbb{R}^d)$. Due to Proposition 3.17 we find a sequence $(f_n)_{n\in\mathbb{N}} \subset \mathscr{S}(\mathbb{R}^d)$ such that $\|f_n - f\|_2 \to 0$ as $n \to +\infty$. In particular, $(f_n)_{n\in\mathbb{N}}$ is a Cauchy-sequence in $L^2(\mathbb{R}^d)$, which means that

$$\lim_{n\to+\infty}\sup_{m\geq n}\|f_n - f_m\|_2 = 0.$$

By Plancherel's theorem we conclude that also $(\mathcal{F}[f_n])_{n\in\mathbb{N}}$ is a Cauchy-sequence in $L^2(\mathbb{R}^d)$, since

$$\lim_{n\to+\infty}\sup_{m\geq n}\|\mathcal{F}[f_n] - \mathcal{F}[f_m]\|_2 = (2\pi)^d\lim_{n\to+\infty}\sup_{m\geq n}\|f_n - f_m\|_2 = 0.$$

Since $L^2(\mathbb{R}^d)$ is a Banach space, there exists $\mathcal{F}[f] := \lim_{n\to+\infty}\mathcal{F}[f_n] \in L^2(\mathbb{R}^d)$. Note that this is the only choice for a continuous extension, which proves uniqueness. Let us show next that this extension does not

[21] To prove this conclusion in detail one can apply Tonelli's theorem for the modulus of the integrand.

depend on the choice of the approximating sequence $(f_n)_{n\in\mathbb{N}}$. Let $(\widetilde{f}_n)_{n\in\mathbb{N}} \subset \mathscr{S}(\mathbb{R}^d)$ be another sequence such that $\|\widetilde{f}_n - f\| \to 0$. Then again due to Plancherel's theorem we have

$$\|\mathcal{F}[\widetilde{f}_n] - \mathcal{F}[f_n]\|_2 = (2\pi)^d \|\widetilde{f}_n - f_n\|_2 \leq (2\pi)^d (\|\widetilde{f}_n - f\|_2 + \|f - f_n\|_2) \to 0.$$

Thus $\mathcal{F}[\widetilde{f}_n]$ converges to the same element as $\mathcal{F}[f_n]$ as claimed. Moreover, one can show that $\mathcal{F} : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ inherits the linearity from the Fourier transform on $\mathscr{S}(\mathbb{R}^d)$. In order prove that it is continuous on $L^2(\mathbb{R}^d)$ it suffices to prove that

$$\langle \mathcal{F}[f], \mathcal{F}[g] \rangle = (2\pi)^d \langle f, g \rangle \tag{7}$$

for all $f, g \in L^2(\mathbb{R}^d)$. Indeed, setting $f = g$ in (7) implies that $\|\mathcal{F}[f]\|_2 = \sqrt{(2\pi)^d}\|f\|_2$, so that by linearity

$$\|\mathcal{F}[f_n] - \mathcal{F}[f]\|_2 = \|\mathcal{F}[f_n - f]\|_2 = \sqrt{(2\pi)^d}\|f_n - f\|_2.$$

This implies that $\mathcal{F}[f_n] \to \mathcal{F}[f]$ in $L^2(\mathbb{R}^d)$ whenever $f_n \to f$ in $L^2(\mathbb{R}^d)$. In order to prove (7), let $(f_n)_{n\in\mathbb{N}}, (g_n)_{n\in\mathbb{N}} \subset \mathscr{S}(\mathbb{R}^d)$ be sequences such that $f_n \to f$ and $g_n \to g$ in $L^2(\mathbb{R}^d)$. Then by Proposition 3.19 we have

$$\langle \mathcal{F}[f_n], \mathcal{F}[g_n] \rangle = (2\pi)^d \langle f_n, g_n \rangle.$$

Thus it suffices to pass to the limit as $n \to +\infty$. As all functions converge in $L^2(\mathbb{R}^d)$ we only treat the right hand side term since the left hand side is completely analogous. Due to Hölder's inequality

$$\begin{aligned} |\langle f_n, g_n \rangle - \langle f, g \rangle| &\leq |\langle f_n, g_n - g \rangle| + |\langle f_n - f, g \rangle| \leq \|f_n(g_n - g)\|_1 + \|(f_n - f)g\|_1 \\ &\leq \underbrace{\|f_n\|_2}_{\to \|f\|_2} \underbrace{\|g_n - g\|_2}_{\to 0} + \underbrace{\|f_n - f\|_2}_{\to 0} \|g\|_2 \to 0. \end{aligned}$$

This proves (7) and therefore it only remains to prove bijectivity. Note that by linearity and (7) we have that

$$\|\mathcal{F}[f] - \mathcal{F}[g]\|_2^2 = \|\mathcal{F}[f - g]\|_2^2 = (2\pi)^d \|f - g\|_2^2.$$

Hence $\mathcal{F}[f] = \mathcal{F}[g]$ implies that $f = g$ in $L^2(\mathbb{R}^d)$, which shows that $\mathcal{F}$ is injective. Finally, we show that $\mathcal{F}$ is surjective. Fix $h \in L^2(\mathbb{R}^d)$ and consider a sequence $(h_n)_{n\in\mathbb{N}} \subset \mathscr{S}(\mathbb{R}^d)$ be such that $h_n \to h$ in $L^2(\mathbb{R}^d)$. Since the Fourier transform is bijective on $\mathscr{S}(\mathbb{R}^d)$ we find a unique $f_n \in \mathscr{S}(\mathbb{R}^d)$ such that $\mathcal{F}[f_n] = h_n$. Again by Plancherel's identity the fact that (h_n) is a Cauchy-sequence implies that also (f_n) is a Cauchy-sequence with a limit $f \in L^2(\mathbb{R}^d)$. By construction it follows that $\mathcal{F}[f] = h$, so that $\mathcal{F}$ is surjective. □

Using approximation with Schwartz functions one can further show that for $f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ the abstract definition of the Fourier transform coincides a.e. with the classical one given by integration. More precisely, using the construction of the proof of Theorem 2.42 combined with the construction of exercise H 11.1 we obtain a sequence of Schwartz functions $(f_n)_{n\in\mathbb{N}} \subset \mathscr{S}(\mathbb{R}^d)$ that converges to f with respect to the L^1-norm and the L^2-norm. Then for fixed $k \in \mathbb{R}^d$ we have that

$$\left| \int_{\mathbb{R}^d} f(y) e^{-ik\cdot x}\, \mathrm{d}x - \int_{\mathbb{R}^d} f_n(y) e^{-ik\cdot x}\, \mathrm{d}x \right| \leq \|f - f_n\|_1 \to 0.$$

This shows in particular that $\mathcal{F}[f_n](k) \to \int_{\mathbb{R}^d} f(y) e^{-ik\cdot x}\, \mathrm{d}x$ pointwise in k and since the right hand side in the above estimate is independent of k it follows from Lebesgue's dominated convergence theorem that the convergence also holds in $L^2(B_j(0))$ for every ball $B_j(0) \subset \mathbb{R}^d$. As $\mathcal{F}[f_n] \to \mathcal{F}[f]$ in $L^2(\mathbb{R}^d)$ by construction, we also have that $\mathcal{F}[f_n] \to \mathcal{F}[f]$ in $L^2(B_j(0))$. Hence $\mathcal{F}[f](k) = \int_{\mathbb{R}^d} f(x) e^{-ik\cdot x}\, \mathrm{d}x$ for a.e. $k \in B_j(0)$. Using that $\mathbb{R}^d = \bigcup_{j\in\mathbb{N}} B_j(0)$ we obtain that the functions agree a.e. on $\mathbb{R}^d$ which is the best we can expect since an L^2-function is only defined up to null sets.

The Fourier transform on $L^2(\mathbb{R}^d)$ is not given by a Lebesgue integral when $f \notin L^1(\mathbb{R}^d)$. However, in what follows we derive a representation via an 'improper' Lebesgue integral.

Lemma 3.21. *Let $f \in L^2(\mathbb{R}^d)$. Then it holds that*

$$\mathcal{F}[f](k) = \lim_{n\to+\infty} \int_{B_n(0)} f(x) e^{-ik\cdot x}\, \mathrm{d}x,$$

where the limit is understood in the sense of $L^2(\mathbb{R}^d)$ (not pointwise).

Remark 3.22. In dimension 1 the above result also holds for a.e. $k \in \mathbb{R}$ (this is called Carlson's theorem). To the best of the author's knowledge the almost everywhere convergence in higher dimensions is an open problem.

Proof of Lemma 3.21. Fix $f \in L^2(\mathbb{R}^d)$ and consider its L^2-Fourier transform $\mathcal{F}[f] \in L^2(\mathbb{R}^d)$. Due to exercise H 12.1 it holds that $\mathbb{1}_{B_n(0)} f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, so that due to Plancherel's theorem (in the L^2-version) and the preceding arguments we have that

$$\int_{\mathbb{R}^d} \left| \mathcal{F}[f](k) - \int_{B_n(0)} f(x) e^{-ik\cdot x}\,\mathrm{d}x \right|^2 \mathrm{d}k = \|\mathcal{F}[f] - \mathcal{F}[\mathbb{1}_{B_n(0)} f]\|_2^2 = \|\mathcal{F}[f(1-\mathbb{1}_{B_n(0)})]\|_2^2$$
$$= (2\pi)^d \|f(1-\mathbb{1}_{B_n(0)})\|_2^2 = (2\pi)^d \int_{\{\|x\|\geq n\}} |f(x)|^2\,\mathrm{d}x.$$

Since $\mathbb{1}_{\{\|x\|\geq n\}}(x)|f(x)|^2$ converges pointwise to 0 as $n \to +\infty$ and is bounded by the integrable function $|f(x)|^2$, it follows from Lebesgue's dominated convergence theorem that the right hand side vanishes when $n \to +\infty$. This proves the claim. □

Concerning other properties that we proved hitherto, their validity on $L^2(\mathbb{R}^d)$ depends on whether they are inherited from Schwartz functions. For instance, as we have seen Plancherel's theorem extends to $L^2(\mathbb{R}^d)$. On the contrary, the fact that $\mathcal{F}[f]$ is bounded and continuous (cf. Lemma 3.3) cannot be true for every $f \in L^2(\mathbb{R}^d)$ since the Fourier transform is bijective and not all L^2-functions are continuous. The connection between decay at infinity and smoothness leads to the so-called Sobolev spaces that are beyond the scope of this course. We finally discuss the Fourier inversion theorem in the sense of L^2 from a more detailed perspective than mere bijectivity.

Corollary 3.23 (Structure of the inverse Fourier transform on $L^2(\mathbb{R}^d)$)**.** *Define the flipping operator $\mathcal{R} : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ by $\mathcal{R}(f)(x) = f(-x)$. Then the inverse Fourier transform $\mathcal{F}^{-1} : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ is given by the formula*

$$\mathcal{F}^{-1} = \frac{1}{(2\pi)^d} \mathcal{R} \circ \mathcal{F}.$$

In particular, also $\mathcal{F}^{-1} : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ is continuous.[22]

Proof. Fix $h \in L^2(\mathbb{R}^d)$ and choose $f \in L^2(\mathbb{R}^d)$ as the unique element such that $h = \mathcal{F}[f]$. We have to prove that $\frac{1}{(2\pi)^d}\mathcal{R}(\mathcal{F}[h]) = f$. Let $(f_n)_{n\in\mathbb{N}} \subset \mathscr{S}(\mathbb{R}^d)$ be such that $f_n \to f$ in $L^2(\mathbb{R}^d)$. Then by continuity also $\mathcal{F}[f_n] \to \mathcal{F}[f] = h$. By the Fourier inversion theorem on the Schwartz space we have pointwise for all $x \in \mathbb{R}^d$

$$f_n(x) = \mathcal{F}^{-1}[\mathcal{F}[f_n]](x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{ik\cdot x} \mathcal{F}[f_n](k)\,\mathrm{d}k = \frac{1}{(2\pi)^d} \mathcal{R}(\mathcal{F}[\mathcal{F}[f_n]])(x)$$

In $L^2(\mathbb{R}^d)$ this equality turns into $f_n = \frac{1}{(2\pi)^d} (\mathcal{R} \circ \mathcal{F})(\mathcal{F}[f_n])$. Since $\mathcal{R} \circ \mathcal{F}$ is continuous on $L^2(\mathbb{R}^d)$ as the composition of continuous functions we can pass to the L^2-limit in n and conclude the proof. □

3.3. Tempered distributions and their Fourier transform.

The final objects for which we want to define their Fourier transform are so-called tempered distributions. First we define what is a distribution.

Definition 3.24 (Distributions)**.** Let $O \subset \mathbb{R}^d$ be an open set. A distribution is a linear functional $T : C_c^\infty(O) \to \mathbb{R}$ (or $\mathbb{C}$) that is continuous with respect to the following notion of convergence:

$$\varphi_n \overset{C_c^\infty(O)}{\to} \varphi :\iff \{\exists K \subset O \text{ compact } : \operatorname{supp}(\varphi_n) \subset K \ \ \forall n \quad \& \quad \|\partial^\alpha \varphi_n - \partial^\alpha \varphi\|_\infty \to 0 \quad \forall \alpha \in \mathbb{N}_0^d\}.$$

One writes $T \in \mathcal{D}'(O)$.

Example 3.25. The following are examples of distributions $T \in \mathcal{D}'(\mathbb{R}^d)$.

[22] The continuity of the inverse of a linear mapping between Banach spaces follows also from the more general open mapping theorem in functional analysis.

(i) $T(\varphi) = \varphi(0)$ (this is also called the Dirac delta $T = \delta_0$);
(ii) $T(\varphi) = \partial^\alpha\varphi(x_0)$ for some fixed multi-index $\alpha \in \mathbb{N}_0^d$ and some fixed $x_0 \in \mathbb{R}^d$;
(iii) $T_f(\varphi) = \int_{\mathbb{R}^d} f(x)\varphi(x)\,\mathrm{d}x$, where f is integrable on every compact subset $K \subset \mathbb{R}^d$;
(iv) $T(\varphi) = \int_{\mathbb{R}^d} f(x)\partial^\alpha\varphi(x)\,\mathrm{d}x$ for some fixed multi-index $\alpha \in \mathbb{N}_0^d$ and f as in (iii).
(iv) $T(\varphi) = \int_{\mathbb{R}^d} \varphi(x)\,\mathrm{d}\mu$, where μ is a measure on $(\mathbb{R}^d, \mathcal{L}^d)$ that is finite on every compact subset of $\mathbb{R}^d$.

On distributions one can define many operations by duality. In the next definition we present three important examples.

Definition 3.26. Let $O \subset \mathbb{R}^d$ be open and $T \in \mathcal{D}'(O)$ be a distribution.

a) The partial derivative of T is defined by
$$(\partial^\alpha T)(\varphi) = (-1)^{|\alpha|}T(\partial^\alpha\varphi).$$
b) If $\psi \in C^\infty(\mathbb{R}^d)$, then the product of T with ψ is defined by
$$(\psi T)(\varphi) = T(\psi\varphi).$$
c) If $O = \mathbb{R}^d$ and $\psi \in C_c^\infty(\mathbb{R}^d)$ then the convolution of ψ with T is a function defined by
$$(\psi * T)(x) = T(\psi(x - \cdot))$$

Remark 3.27. One can show that the partial derivative and the product with a smooth function are again distributions, while the convolution is a C^∞-function on $\mathbb{R}^d$. The idea behind these definitions is that they coincide with the classical definitions in the case when T is a distribution as in Example 3.25(iii). For instance, when $f \in C^m(\mathbb{R}^d)$, then for any multi-index $\alpha \in \mathbb{N}_0^d$ with $|\alpha\| \leq m$ we have by integration by parts (no boundary terms since φ has compact support) that
$$(\partial^\alpha T_f)(\varphi) = (-1)^{|\alpha|}\int_{\mathbb{R}^d} f(x)\partial^\alpha\varphi(x)\,\mathrm{d}x = \int_{\mathbb{R}^d} \partial^\alpha f(x)\varphi(x)\,\mathrm{d}x = T_{\partial^\alpha f}(\varphi).$$
Similarly, it holds that $(\psi T_f) = (T_{\psi f})$ and $\psi * T_f(x) = \psi * f(x)$. Studying PDEs often leads to the notion of distributional derivatives of integrable functions. This means exactly the quantity $\partial^\alpha T_f$, which makes sense even when f is not differentiable. For instance, consider the heavyside-function $f = \mathbb{1}_{[0,+\infty)}$ in dimension 1. Then for any $\varphi \in C_c^\infty(\mathbb{R})$ we have
$$T_f'(\varphi) = -\int_{\mathbb{R}} f(x)\varphi'(x)\,\mathrm{d}x = -\int_0^\infty \varphi'(x)\,\mathrm{d}x = \varphi(0),$$
where we used that φ has compact support. This means that the distributional derivative of f is the Dirac delta centered in 0.

We can also speak about the convergence of distributions as defined below.

Definition 3.28. Let $O \subset \mathbb{R}^d$ be open and $(T_n)_{n\in\mathbb{N}} \subset \mathcal{D}'(O)$ be a sequence of distributions. We say that T_n converges to some $T \in \mathcal{D}'(O)$ if $T_n(\varphi) \to T(\varphi)$ for all $\varphi \in C_c^\infty(O)$.

Using convolution techniques one can then show that for every distribution $T \in D'(O)$ there exists a sequence of functions $(f_n)_{n\in\mathbb{N}} \subset C_c^\infty(O)$ such that $T_{f_n} \to T$. Moreover, the notion of approximate identity (cf. Analysis 3) can be phrased in the sense of convergence of distributions to the Dirac delta.

Can we also define a Fourier transform acting on distributions? Again we would like to do this by duality. Consider a distribution T_f with $f \in L^1(\mathbb{R}^d)$. Then the Fourier transform of T_f should be given by $T_{\mathcal{F}[f]}$ which leads to
$$T_{\mathcal{F}[f]}(\varphi) = \int_{\mathbb{R}^d} \varphi(k)\mathcal{F}[f](k)\,\mathrm{d}k = \int_{\mathbb{R}^d} \varphi(k)\int_{\mathbb{R}^d} e^{-ik\cdot x}f(x)\,\mathrm{d}x\,\mathrm{d}k \overset{\text{Fubini}}{=} \int_{\mathbb{R}^d} f(x)\mathcal{F}[\varphi](x)\,\mathrm{d}x = T_f(\mathcal{F}[\varphi]).$$
Can we extend the right hand side to any distribution? The answer is no since the Fourier transform of a $C_c^\infty(\mathbb{R}^d)$ never has compact support (cf. exercises), so that $T(\mathcal{F}[\varphi])$ is not defined in general. However, we know that the Fourier transform maps $C_c^\infty(\mathbb{R}^d)$ into $\mathscr{S}(\mathbb{R}^d)$. Hence we could define the Fourier transform on distributions that are well-defined on $\mathscr{S}(\mathbb{R}^d)$. Since $C_c^\infty(\mathbb{R}^d) \subset \mathscr{S}(\mathbb{R}^d)$, this means that those are special distributions.

Definition 3.29. A tempered distribution is a linear functional $T : \mathscr{S}(\mathbb{R}^d) \to \mathbb{R}$ (or $\mathbb{C}$) that is continuous with respect to the following type of convergence:

$$\varphi_n \overset{\mathscr{S}(\mathbb{R}^d)}{\to} \varphi :\Longleftrightarrow \ \forall \alpha, \beta \in \mathbb{N}_0^d : \ \sup_{x \in \mathbb{R}^d} |x^\alpha \partial^\beta (\varphi_n(x) - \varphi(x))| \to 0.$$

We write $T \in \mathscr{S}'(\mathbb{R}^d)$.

Note that every tempered distribution is also a distribution on $\mathbb{R}^d$ since the convergence for $C_c^\infty(\mathbb{R}^d)$-functions implies the convergence in $\mathscr{S}(\mathbb{R}^d)$. Most of the definitions made for distributions make also sense for tempered distributions. We collect them in the definition below.

Definition 3.30. Let $T, (T_n)_{n\in\mathbb{N}} \subset \mathscr{S}'(\mathbb{R}^d)$ be tempered distributions, $\alpha \in \mathbb{N}_0^d$ and $\psi \in \mathscr{S}(\mathbb{R}^d)$.

a) The partial derivative of T is defined by
$$(\partial^\alpha T)(\varphi) = (-1)^\alpha T(\partial^\alpha \varphi).$$
b) The product of T with ψ[23] is defined by
$$(\psi T)(\varphi) = T(\psi\varphi).$$
c) If $O = \mathbb{R}^d$ the convolution of ψ with T is defined as the function
$$(\psi * T)(x) = T(\psi(x - \cdot)).$$
d) We say that $T_n \to T$ in $\mathscr{S}'(\mathbb{R}^d)$ if $T_n(\varphi) \to T(\varphi)$ for all $\varphi \in \mathscr{S}(\mathbb{R}^d)$.

Note that the partial derivative and the product are again tempered distributions. Moreover, one can show that the convolution belongs to $C^\infty(\mathbb{R}^d)$ and one can control its growth and of all its derivatives.

Finally, we introduce the Fourier transform for tempered distributions which are the most general objects for which the Fourier transform is defined (note that every $L^p(\mathbb{R}^d)$ function can be interpreted as tempered distribution in the sense of Example 3.25 (iii)). However, not all cases in Example 3.25 define tempered distributions. For instance, the locally integrable function $f(x) = e^{x^2}$ grows to quickly so that its product with the Schwartz function $\varphi(x) = e^{-x^2}$ has no finite integral on $\mathbb{R}$.

Theorem 3.31. *Given $T \in \mathscr{S}'(\mathbb{R}^d)$ define its Fourier transform $\widehat{T} \in \mathscr{S}'(\mathbb{R}^d)$ by $\widehat{T}(\varphi) = T(\mathcal{F}[\varphi])$. Then the Fourier transform is a linear bijective map from $\mathscr{S}'(\mathbb{R}^d)$ to $\mathscr{S}'(\mathbb{R}^d)$.*

Proof. First we show that $\widehat{T}$ is again a tempered distribution. Since $\mathcal{F}[\varphi] \in \mathscr{S}(\mathbb{R}^d)$ for all $\varphi \in \mathscr{S}(\mathbb{R}^d)$ it is well-defined. Linearity follows from linearity of the Fourier transform and linearity of T. Thus it remains to show that it is continuous. By linearity it suffices to prove that if $(\varphi_n)_{n\in\mathbb{N}} \subset \mathscr{S}(\mathbb{R}^d)$ is a sequence such that $\varphi_n \overset{\mathscr{S}(\mathbb{R}^d)}{\to} 0$ then $\widehat{T}(\varphi_n) \to 0$. By definition and continuity of T this reduces to prove that $\mathcal{F}[\varphi_n] \overset{\mathscr{S}(\mathbb{R}^d)}{\to} 0$. This is a consequence of exercise H 12.5.

To conclude the proof, we need to show that the Fourier transform is bijective. We first prove injectivity. By linearity it suffices to prove that $\widehat{T} = 0$ implies that $T = 0$. Let $\varphi \in \mathscr{S}(\mathbb{R}^d)$ be given. Then $\mathcal{F}^{-1}[\varphi] \in \mathscr{S}(\mathbb{R}^d)$ by the Fourier inversion theorem on the Schwartz space and hence

$$0 = \widehat{T}(\mathcal{F}^{-1}[\varphi]) = T(\mathcal{F}[\mathcal{F}^{-1}[\varphi]]) = T(\varphi),$$

where we used that $\mathcal{F}^{-1}$ is also the right inverse to $\mathcal{F}$ (cf. the proof of Theorem 3.18). Since $\varphi \in \mathscr{S}(\mathbb{R}^d)$ was arbitrary, the above equation implies that $T = 0$. To prove surjectivity, let $T \in \mathscr{S}'(\mathbb{R}^d)$ and define a tempered distribution $S \in \mathscr{S}'(\mathbb{R}^d)$ by the formula $S(\varphi) = T(\mathcal{F}^{-1}[\varphi])$. Again this is a tempered distribution since the inverse Fourier transform is also linear and continuous with respect to the convergence in $\mathscr{S}(\mathbb{R}^d)$ (the argument being the same as for the Fourier transform in H 12.5). Then by definition we have

$$\widehat{S}(\varphi) = S(\mathcal{F}[\varphi]) = T(\mathcal{F}^{-1}[\mathcal{F}[\varphi]]) = T(\varphi).$$

[23]Note that the product of a Schwartz function with a $C^\infty(\mathbb{R}^d)$ function is not necessarily again a Schwartz function. Hence the product ψT is in general not defined as a tempered distribution when $\psi \in C^\infty(\mathbb{R}^d)$. One can weaken the assumption that $\psi \in \mathscr{S}(\mathbb{R}^d)$ by requiring that ψ is smooth and that ψ and all its derivatives grow at most polynomially.

Note that we obtained the formula $\widehat{T}^{-1} = T \circ \mathcal{F}^{-1}$. $\square$

Many properties of the Fourier transform on $\mathscr{S}(\mathbb{R}^d)$ transform to tempered distributions by duality. For instance, it holds that

$$\begin{aligned}\widehat{\partial^\alpha T}(\varphi) &= \partial^\alpha T(\mathcal{F}[\varphi]) = (-1)^{|\alpha|} T(\partial^\alpha \mathcal{F}[\varphi]) \overset{\text{Prop. 3.7}}{=} T(\mathcal{F}[x \mapsto i^{|\alpha|} x^\alpha \varphi(x)]) \\ &= \widehat{T}(x \mapsto i^{|\alpha|} x^\alpha \varphi(x)) = (i^{|\alpha|} k^\alpha \widehat{T})(\varphi),\end{aligned}$$

which is well-defined in the sense of Footnote 23. Compare this formula with Proposition 3.5.

This shall be enough for our short introduction to (tempered) distributions.

4. An introduction to linear operators on Banach/Hilbert spaces

In what follows we let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ and consider $\mathbb{K}$-vector spaces, i.e., the scalars are either real or complex numbers. We recall the definition of Banach and Hilbert spaces; cf. previous Analysis courses.

Definition 4.1. A normed $\mathbb{K}$-vector space $(X, \|\cdot\|)$ is called a Banach space if every Cauchy sequence with respect to the norm $\|\cdot\|$ has a limit in X. It is called a Hilbert space if it is a Banach space and the norm is given by $\|x\| = \sqrt{\langle x, x\rangle}$, where $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{K}$ is a scalar product on X.

For complex vector spaces we shall always consider scalar products to be linear in the second component.

4.1. Linear operators on normed spaces.

Next we introduce linear operators between $\mathbb{K}$-vector spaces X and Y (both with the same scalars).

Definition 4.2. Let X, Y be normed $\mathbb{K}$-vector spaces. A bounded linear operator between X and Y is a linear function $A : X \to Y$ that is continuous. We write $A \in \mathcal{L}(X;Y)$. For $A \in \mathcal{L}(X;Y)$ we define the operator-norm by

$$\|A\| = \|A\|_{\mathcal{L}(X;Y)} := \sup_{x \in X\setminus\{0\}} \frac{\|A(x)\|_Y}{\|x\|_X} = \sup_{\substack{x\in X \\ \|x\|_X = 1}} \|A(x)\|_Y .$$

When $Y = \mathbb{K}$ we write $\mathcal{L}(X;Y) = X'$. This is the so-called topological dual space of X and its elements are called linear functionals.

As shown in the next lemma the operator norm characterizes the continuity of linear functions.

Lemma 4.3. *Let X, Y be normed $\mathbb{K}$-vector spaces and $A : X \to Y$ be a linear function. Then the following statements are equivalent:*

a) $A \in \mathcal{L}(X;Y)$;
b) A *is continuous in* 0;
c) $\|A\| := \sup_{x \in X\setminus\{0\}} \frac{\|A(x)\|_Y}{\|x\|_X} = \sup_{\substack{x\in X \\ \|x\|_X = 1}} \|A(x)\|_Y < +\infty.$

Proof. Clearly a) implies b). Next we show that b) implies c). Assume by contradiction that $\|A\| = +\infty$. Then for every $n \in \mathbb{N}$ there exists $x_n \in X$ with $\|x_n\|_X = 1$ and $\|A(x_n)\|_Y \geq n$. Define $\tilde{x}_n = x_n/\sqrt{n}$. Then $\|\tilde{x}_n\|_X = 1/\sqrt{n}$, so that $\tilde{x}_n \to 0$ in X. However, by linearity of A and homogeneity of the norm

$$\lim_{n\to+\infty} \|A(\tilde{x}_n)\|_Y = \lim_{n\to+\infty} \frac{1}{\sqrt{n}} \|A(x_n)\|_Y \geq \lim_{n\to+\infty} \sqrt{n} = +\infty,$$

which contradicts the continuity of A in 0.

Finally we show that c) yields a). Let $x, x' \in X$ be such that $x \neq x'$. Then by linearity

$$\|A(x) - A(x')\|_Y = \|A(x - x')\|_Y \leq \|A\| \|x - x'\|_X .$$

Obviously this estimate also holds for $x = x'$, so that A is Lipschitz continuous. Hence $A \in \mathcal{L}(X;Y)$. $\square$

The previous lemma motivates the definition of unbounded operators as those operators for which $\|A\| = +\infty$. However, usually unbounded operators are considered to be defined on a subspace of some larger space. This is not necessary for general normed spaces since every subspace is still a normed space itself. However, as soon as we are interested in Banach or Hilbert spaces, subspaces may loose the completeness property when they are not closed. This phenomenon is particular for infinite dimensional spaces. Anyway, linear maps between finite dimensional spaces are always continuous.

Definition 4.4. Let X, Y be normed $\mathbb{K}$-vector spaces and $D(A) \subset X$ be a subspace. A linear function $A : D(A) \to Y$ is called an unbounded linear operator if

$$\|A\| := \sup_{x\in D(A)\setminus\{0\}} \frac{\|A(x)\|_Y}{\|x\|_X} = \sup_{\substack{x\in D(A)\\ \|x\|_X=1}} \|A(x)\|_Y = +\infty.$$

The space $D(A)$ is the domain of the unbounded linear operator.

We look at some examples:

- Let $X = Y = L^2(\mathbb{R}^d)$. Then the Fourier transform $\mathcal{F}$ belongs to $\mathcal{L}(X;Y)$ and due to Plancherel's theorem we have $\|\mathcal{F}\| = \sqrt{(2\pi)^d}$.
- Let $X = Y = L^p(\mathbb{R}^d)$ and for $x \in \mathbb{R}^d$ consider the shift operator $T_x(f) = f(x - \cdot))$. By the change of variables it follows that $\|T_x\| = 1$ for all $x \in \mathbb{R}^d$.
- Let $X = L^1(\mathbb{R}^d)$ and $Y = C_b(\mathbb{R}^d)$ equipped with the supremum norm. Then the Fourier transform belongs to $\mathcal{L}(X;Y)$ and by Lemma 3.3 we know that $\|\mathcal{F}\| \le 1$.
- for fixed $g \in C(\mathbb{R}^d)$ and $X = Y = L^1(\mathbb{R}^d)$ consider the multiplication operator $A(f) = gf$ with domain $D(A) = \{f \in L^1(\mathbb{R}^d) : gf \in L^1(\mathbb{R}^d)\}$. In this case it depends on g whether A is bounded. If g is bounded, then $D(A) = L^1(\mathbb{R}^d)$ and $\|A\| \le \|g\|_\infty$ by Hölder's inequality. If g is not bounded, then there exists a sequence $x_n \in \mathbb{R}^d$ such that $|g(x_n)| \ge n$ for all $n \in \mathbb{N}$. Consider then for fixed $n \in \mathbb{N}$ the sequence $f_k = \frac{1}{\lambda_d(B_{1/k}(0))}\mathbb{1}_{B_{1/k}(x_n)}$. Note that $\|f_k\|_1 = 1$ for all $k \in \mathbb{N}$. Moreover,

$$|g(x_n)| - \int_{\mathbb{R}^d} |g(x) f_k(x)|\,\mathrm{d}x = \frac{1}{\lambda_d(B_{1/k}(0))}\int_{B_{1/k}(x_n)} (|g(x_n)| - |g(x)|)\,\mathrm{d}x \le \sup_{\|x-x_n\|\le 1/k} |g(x_n) - g(x)|.$$

  By continuity of g the right hand side tends to zero when $k \to +\infty$. Hence there exists k_n such that the right hand side is bounded by 1 for all $k \ge k_n$. The sequence $f_n = f_{k_n}$ then satisfies $\|f_n\|_1 = 1$, but $\|A(f_n)\|_1 \ge n - 1$. Since n was arbitrary the operator cannot be bounded.
- the differentiation operator $A = \frac{\mathrm{d}}{\mathrm{d}x} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ with domain $D(A) = \mathscr{S}(\mathbb{R})$ is unbounded (cf. exercises)

In the next lemma we see that the set of bounded linear operators becomes a Banach space as soon as Y is a Banach space.

Lemma 4.5. *Let X, Y be normed $\mathbb{K}$-vector spaces. The space $\mathcal{L}(X;Y)$ becomes a normed $\mathbb{K}$-vector space with the operator norm $\|\cdot\|_{\mathcal{L}(X;Y)}$. When Y is a Banach space, then so is $\mathcal{L}(X;Y)$. In particular, X' is always a Banach space.*

Proof. See exercises. □

Our next result in this rather general setting of normed spaces is the bounded linear transformation theorem.

Theorem 4.6. *Let Z, Y be normed $\mathbb{K}$-vector spaces and $X \subset Z$ be a subspace that is dense in Z with respect to $\|\cdot\|_Z$. Assume that Y is a Banach space. Then every $A \in \mathcal{L}(X;Y)$ can be extended uniquely to an operator $\widetilde{A} \in \mathcal{L}(Z;Y)$.*

Proof. The proof is a minor adaption from the extension procedure of the Fourier transform from $\mathscr{S}(\mathbb{R}^d)$ to $L^2(\mathbb{R}^d)$ except that Plancherel's equality has to be replaced by the inequality $\|Ax - Ay\|_Y \le \|A\|\|x - y\|_Z$ for all $x, y \in X$, so we leave the details to the interested reader. □

Finally we introduce the notion of the adjoint operator. In general also this concept makes sense for normed spaces, but soon we will focus on the Hilbert space setting where the structure becomes easier. At this point we only define the adjoint operator for a bounded linear operator.

Definition 4.7 (Adjoint operator)**.** Let X, Y be normed $\mathbb{K}$-vector spaces with dual spaces X' and Y', respectively. For $A \in \mathcal{L}(X;Y)$ the adjoint operator $A' : Y' \to X'$ is defined by the formula

$$(A'(y'))(x) = y'(A(x)).$$

It holds that $A' \in \mathcal{L}(Y';X')$.

Proof. We first have to show that the adjoint operator is well-defined. To this end, we first check that $A'(y') \in X'$. Let $\lambda \in \mathbb{K}$ and $x_1, x_2 \in X$. Then by definition and the linearity of A and y' we have

$$(A'(y'))(\lambda x_1 + x_2) = y'(A(\lambda x_1 + x_2)) = \lambda y'(A(x_1)) + y'(A(x_2)) = \lambda(A'(y'))(x_1) + (A'(y'))(x_2),$$

which shows that $A'(y')$ is linear on X. Next we show that it is continuous. Note that

$$|(A'(y'))(x)| = |y'(A(x))| \leq \|y'\|\|A(x)\|_Y \leq \|y'\|\|A\|\|x\|_X, \tag{8}$$

which shows that $\|A'(y')\| \leq \|y'\|\|A\| < +\infty$. Hence $A'(y') \in X'$. The second property concerns the linearity on Y'. Let $\lambda \in \mathbb{K}$ and $y'_1, y'_2 \in Y'$. Then for every $x \in X$ we have

$$(A'(\lambda y'_1 + y'_2))(x) = (\lambda y'_1 + y'_2)(A(x)) = \lambda y'_1(A(x)) + y'_2(A(x)) = \lambda(A'(y'_1))(x) + (A'(y'_2))(x).$$

Thus $A'(y')$ is linear in y'. Finally, we need to show continuity with respect to y'. Note that on X' and Y' we take the operator norm. By (8) we have that

$$\sup_{\substack{y' \in Y' \\ \|y'\|_{Y'}=1}} \|A'(y')\|_{X'} = \sup_{\substack{y' \in Y' \\ \|y'\|_{Y'}=1}} \sup_{\substack{x \in X \\ \|x\|_X=1}} \|(A'(y'))(x)| \leq \|A\| < +\infty.$$

□

The final question we want to address in this subsection is whether the adjoint operator already determines the operator itself. This question reduces to the following one: Does the equality $A'_1(y') = A'_2(y')$ for all $y' \in Y'$ imply that $A_1 = A_2$. Inserting the definition of the adjoint operator we obtain that

$$y'(A_1(x)) = y'(A_2(x)) \quad \forall y' \in Y', x \in X \overset{?}{\Longrightarrow} A_1 = A_2. \tag{9}$$

This question can be answered affirmatively with the help of the Hahn-Banach theorem, another instance for which one uses the axiom of choice. For the sake of completeness we mention it here without proof.

Theorem 4.8 (Hahn-Banach theorem for normed spaces)**.** *Let X be a normed $\mathbb{K}$-vector space and $Z \subset X$ be a non-trivial subspace. Assume that $f \in Z'$. Then f can be extended to an element $\widetilde{f} \in X'$ with $\|\widetilde{f}\|_{X'} = \|f\|_{Z'}$.*

This theorem implies in particular that for every $x_0 \in X \setminus \{0\}$ there exists $x' \in X'$ with $x'(x_0) = \|x_0\|$ and $\|x'\|_{X'} = 1$ by setting $Z = \text{span}(x_0)$ and $f(\lambda x_0) = \lambda\|x_0\|$. The validity of (9) then follows from the fact that if $A_1(x) \neq A_2(x)$ for some $x \in X$ then $A_1(x) - A_2(x) \neq 0$, so that there exists $y' \in Y'$ such that $y'(A_1(x) - A_2(x)) \neq 0$, which yields a contradiction.

4.2. **Linear operators on Hilbert spaces.** We now focus on the case when X and Y are Hilbert spaces. We will prove the Riesz-representation theorem which shows that the dual space of any Hilbert space X can be identified with X. For the proof we will need a result about the existence of projections that is interesting on its own.

Proposition 4.9 (Projection onto closed, convex[24] sets)**.** *Let X be a $\mathbb{K}$-Hilbert space and $K \subset X$ be a nonempty, closed, convex set. Then for every $x \in X$ there exists a unique element $x_K \in K$ such that $\|x - x_k\| \leq \|x - y\|$ for all $y \in K$.*

[24] A subset K of a vector space X is called convex if for all $x_1, x_2 \in K$ and $t \in [0,1]$ it holds that $tx_1 + (1-t)x_2 \in K$.

Proof. By considering the non-empty, closed and convex set $x - K$ we can assume that $x = 0$ and we need to prove the existence and uniqueness of the minimizer of the problem

$$\inf_{y \in K} \|y\|.$$

First we prove existence. Note that this is a non-trivial issue since bounded and closed sets are not always compact in infinite dimensional normed spaces. Consider a minimizing sequence $(y_n)_{n\in\mathbb{N}} \subset K$ such that $\lim_{n\to+\infty} \|y_n\| = \inf_{y\in K} \|y\| =: \delta$. Then for any $m, n \in \mathbb{N}$ we have by convexity that $\frac{y_n+y_m}{2} \in K$, so that the parallelogram law (cf. Analysis 3) yields

$$\left\|\frac{y_n - y_m}{2}\right\|^2 = \frac{1}{2}\|y_m\|^2 + \frac{1}{2}\|y_n\|^2 - \left\|\frac{y_n + y_m}{2}\right\|^2 \leq \frac{1}{2}\|y_m\|^2 + \frac{1}{2}\|y_n\|^2 - \delta^2$$

Since $\|y_n\| \to \delta$ this implies that $(y_n)_{n\in\mathbb{N}}$ is a Cauchy sequence. Indeed, the right hand side tends to zero when $n, m \to +\infty$. Being X complete there exists a limit $y = \lim_{n\to+\infty} y_n$. Since K is closed by assumption it follows that $y \in K$ and by continuity of the norm we conclude that $\|y\| = \delta$. By definition of δ it follows that y is the claimed minimizer.

Next, we show uniqueness. To this end, assume that there exist two elements $y, \widetilde{y} \in K$ such that $\|y\| = \|\widetilde{y}\| = \delta$. Repeating the above estimate with y_n replaced by y and y_m replaced by $\widetilde{y}$ be obtain

$$\left\|\frac{y - \widetilde{y}}{2}\right\|^2 \leq 0.$$

Hence $y = \widetilde{y}$. □

The statement of the previous proposition remains true for so-called reflexive and strictly convex Banach spaces, but fails for general Banach spaces (even in finite dimensions). For instance, consider in $\mathbb{R}^2$ the set $K = [0, 1] \times \{0\}$. Then for the point $x = (0, 1)$ all points in $y \in K$ satisfy $\|y - x\|_\infty = 1$. Hence uniqueness fails. In infinite dimensions even the existence of (maybe many) closest points fails in general.

Before proving the Riesz representation theorem we need to study the concept of orthogonal complements in Hilbert spaces.

Definition 4.10. Let X be a $\mathbb{K}$-Hilbert space with scalar product $\langle \cdot, \cdot \rangle$.

(i) $x_1 \in X$ is called orthogonal to $x_2 \in X$ if $\langle x_1, x_2 \rangle = 0$;
(ii) Given $S \subset X$ the orthogonal complement of S is defined by

$$S^\perp = \{x \in X : \langle x, s \rangle = 0 \ \forall s \in S\}.$$

Lemma 4.11. *Let X be a $\mathbb{K}$-Hilbert space and $S \subset X$. Then $S^\perp$ is a closed subspace of X.*

Proof. Let $\lambda \in \mathbb{K}$ and $x_1, x_2 \in S^\perp$. Then for any $s \in S$ we have that

$$\langle \lambda x_1 + x_2, s \rangle = \overline{\lambda} \underbrace{\langle x_1, s \rangle}_{=0} + \underbrace{\langle x_2, s \rangle}_{=0} = 0.$$

Thus $S^\perp$ is a subspace. Next, let $(x_n)_{n\in\mathbb{N}} \subset S^\perp$ be a sequence such that $x_n \to x \in X$. The Cauchy-Schwarz inequality (cf. Analysis 3) implies that for all $s \in S$

$$|\langle x_n, s \rangle - \langle x, s \rangle| = |\langle x_n - x, s \rangle| \leq \|x_n\|\|s\| \to 0.$$

Since $\langle x_n, s \rangle = 0$ for all $n \in \mathbb{N}$ this yields that $\langle x, s \rangle = 0$. As $s \in S$ was arbitrary this implies that $x \in S^\perp$, so that $S^\perp$ is closed. □

Since subspaces are always nonempty and convex, there exists in particular a projection onto closed subspaces. This implies the following corollary.

Corollary 4.12 (Orthogonal decomposition)**.** *Let X be a $\mathbb{K}$-Hilbert space and $U \subset X$ be a closed subspace. Then every $x \in X$ can be uniquely decomposed as $x = x_1 + x_2$ with $x_1 \in U$ and $x_2 \in U^\perp$. We write $X = U \oplus U^\perp$.*

Proof. Let x_1 be the unique element in U that satisfies $\|x - x_1\| \le \|x - y\|$ for all $y \in U$. We claim that $x - x_1 \in U^\perp$. Fix $u \in U$. Then for all $\varepsilon \in \mathbb{R}$ we have by minimality that

$$\|x - x_1\|^2 \le \|x - \underbrace{(x_1 + \varepsilon u)}_{\in U}\|'2 = \|x - x_1\|^2 + \varepsilon^2\|u\|^2 - 2\varepsilon \mathrm{Re}(\langle x - x_1, u\rangle)$$

This is only possible for all $\varepsilon \in \mathbb{R}$ when $\mathrm{Re}(\langle x - x_1, u\rangle) = 0$. Since also $v = iu \in U$ (if $\mathbb{K} = \mathbb{R}$ we are already done) we further obtain that $0 = \mathrm{Re}(\langle x - x_1, v\rangle) = \mathrm{Re}(i\langle x - x_1, u\rangle) = -\mathrm{Im}(\langle x - x_1, u\rangle)$, so that finally $\langle x - x_1, u\rangle = 0$. Hence $x - x_1 \in U^\perp$.

To show uniqueness of the representation, assume that $x_1 + x_2 = y_1 + y_2$ with $x_1, y_1 \in U$ and $x_2, y_2 \in U^\perp$. Then $x_1 - y_1 = y_2 - x_2$. Taking the scalar product with $x_1 - y_1 \in U$ and $y_2 - x_2 \in U^\perp$ we obtain by orthogonality that $\|x_1 - y_1\|^2 = \|y_2 - x_2\|^2 = 0$. Thus $x_1 = y_1$ and $x_2 = y_2$ as claimed. □

Now we are in a position to prove the Riesz representation theorem about the dual space of $\mathbb{K}$-Hilbert spaces.

Theorem 4.13 (Riesz representation theorem)**.** *Let X be a $\mathbb{K}$-Hilbert space and $x' \in X'$. Then there exists a unique $y \in X$ such that*

$$x'(x) = \langle y, x\rangle \quad \forall x \in X.$$

Moreover, $\|x'\|_{X'} = \|y\|_X$.

Proof. Fix $x' \in X'$. We first prove the existence of y. Let $U := \{x \in X : x'(x) = 0\}$ be the kernel of x'. Since x' is continuous and linear it follows that U is a closed subspace. If $U = X$ then $x' = 0$ and we choose $y = 0$ to conclude. Thus we can assume that $U \ne X$, so that due to Corollary 4.12 there exists $u^\perp \in U^\perp$ such that $x'(u^\perp) \ne 0$. Up to normalization we can assume that $x'(u^\perp) = 1$. Then for all $x \in X$ it holds that $x - x'(x)u^\perp \in U$ and therefore by orthogonality

$$\langle u^\perp, x\rangle = \langle u^\perp, x - x + x'(x)u^\perp\rangle = x'(x)\|u^\perp\|^2.$$

Setting $y = \frac{u^\perp}{\|u^\perp\|^2}$ we conclude from the above equality that $\langle y, x\rangle = x'(x)$ as claimed. To prove uniqueness note that $\langle y_1, x\rangle = \langle y_2, x\rangle$ for all $x \in X$ implies that $\langle y_1 - y_2, x\rangle = 0$ for all $x \in X$. Setting $x = y_1 - y_2$ we deduce that $\|y_1 - y_2\|^2 = 0$, which proves uniqueness. Finally, note that by the Cauchy-Schwarz inequality

$$\|x'\|_{X'} = \sup_{\substack{x \in X \\ \|x\|_X = 1}} |x'(x)| = \sup_{\substack{x \in X \\ \|x\|_X = 1}} |\langle y, x\rangle| \le \|y\|_X.$$

On the other hand, as soon as $y \ne 0$ (otherwise $\|y\|_X = 0 = \|x'\|_{X'}$ holds true) we know that $y/\|y\|_X$ has norm 1, so that

$$\|x'\|_{X'} = \sup_{\substack{x \in X \\ \|x\|_X = 1}} |x'(x)| = \sup_{\substack{x \in X \\ \|x\|_X = 1}} |\langle y, x\rangle| \ge |\langle y, y/\|y\|_X\rangle| = \|y\|_X.$$

Thus $\|x'\|_{X'} = \|y\|_X$. □

By the Cauchy-Schwarz inequality, for every $y \in X$ the functional $x \mapsto \langle y, x\rangle$ belongs to X'. Hence we obtain the following corollary.

Corollary 4.14. *Let X be a $\mathbb{K}$-Hilbert space. Then the map $I : X \to X'$ defined by $y \mapsto x \mapsto \langle y, x\rangle$ is a bijective, conjugate-linear isometry.*

With the map I the spaces X and X' can be identified. We stress that this does not mean that the elements are equal which sometimes leads to confusion. For instance, one can show that $L^2(\mathbb{R}^d)$ is isometrically isomorphic to $\ell^2 = L^2(\mathbb{N})$. But in $L^2(\mathbb{R}^d)$ we have equivalence classes of square-integrable functions, while ℓ^2 contains square-summable sequences.

Next we study the adjoint operator between Hilbert spaces X and Y inserting the identification of the dual spaces. Then the adjoint operator belongs to $\mathcal{L}(Y; X)$ and is given by $A^* = I^{-1} \circ A' \circ I$, or equivalently, $I \circ A^* = A' \circ I$, which means that $A^*(y)$ is the unique element such that

$$\langle A^*(y), x\rangle = A'(\langle y, \cdot\rangle)(x) = \langle y, A(x)\rangle.$$

Often the round brackets are omitted for linear operators and we will do the same in what follows. In the definition below we define the adjoint also for unbounded operators. We will always assume that their domain $D(A)$ is dense in some Hilbert space. This is no restriction since otherwise we can consider the closure of $D(A)$ which is again a Hilbert space.

Definition 4.15. Let X, Y be $\mathbb{K}$-Hilbert spaces with scalar products $\langle\cdot,\cdot\rangle_X$ and $\langle\cdot,\cdot\rangle_Y$, respectively. Let $A : D(A) \subset X \to Y$ be a linear function, where $D(A)$ is a dense subspace of X. The adjoint operator $A^* : D(A^*) \subset Y \to X$ is defined by

$$\langle A^*y, x\rangle_X = \langle y, Ax\rangle_Y \qquad \forall x \in D(A), y \in D(A^*),$$

where $D(A^*) = \{y \in Y : D(A) \ni x \mapsto \langle y, Ax\rangle_Y \text{ is a bounded functional}\}$.

Remark 4.16. Since we assume that $D(A)$ is dense in X, for any $y \in D(A^*)$ the bounded functional $x \mapsto \langle A^*y, x\rangle$ can be extended uniquely to an element in $X' \sim X$. Hence $A^*y \in X$ is uniquely defined. Moreover, $D(A^*)$ is always a linear space and A^* is linear in y. If $A \in \mathcal{L}(X;Y)$ then $D(A) = X$ and $D(A^*) = Y$ by the Cauchy-Schwarz inequality. If we only know that $D(A^*) = Y$, then one can show that A was already bounded on the dense set $D(A)$ and thus can be extended to an element of $\mathcal{L}(X;Y)$. However, this is beyond the scope of this introduction.

We next give some definitions that can be useful in mathematical physics, for instance in quantum mechanics.

Definition 4.17. Let X be a $\mathbb{K}$-Hilbert space and let $A : D(A) \subset X \to X$ be a linear operator (not necessarily bounded) such that $D(A)$ is dense in X.

(i) We call A self-adjoint if $D(A) = D(A^*)$ and $A = A^*$;

(ii) We call A symmetric (or hermitian[25]) if

$$\langle Ax, y\rangle = \langle x, Ay\rangle \qquad \forall x, y \in D(A);$$

(iii) We say that A is closed if for every sequence $(x_n)_{n\in\mathbb{N}} \subset D(A)$ converging to some $x \in X$ and such that $Ax_n \to y \in X$ we have $x \in D(A)$ and $Ax = y$.[26]

(iv) We say that A is normal if it is closed and $A^*A = AA^*$. In particular this requires that $D(A^*A) = D(AA^*)$.

Remark 4.18. Note that for a symmetric operator it holds that $D(A) \subset D(A^*)$. If equality holds, then A is self-adjoint. The Hellinger-Toeplitz theorem states that every symmetric operator with $D(A) = X$ is already bounded. One can also show that A^* is always a closed operator, so that self-adjoint operators are normal.

We have no time to discuss why the above definitions are useful. We just mention that a version of the spectral theorem holds for normal operators. In the next lemma we collect some elementary properties of the adjoint operator. To avoid technical issues we restrict it to the case of bounded, linear operators.

Lemma 4.19. *Let X be a $\mathbb{K}$-Hilbert space and $A, B \in \mathcal{L}(X;X)$. Then the following properties hold:*

(i) $(A^*)^* = A$;

(ii) $(AB)^* = B^*A^*$;

(iii) $(A + \lambda B)^* = A^* + \overline{\lambda}B^*$ *for all* $\lambda \in \mathbb{K}$;

(iv) $\|A\| = \|A^*\|$;

(v) $\|A^*A\| = \|AA^*\| = \|A\|^2$.

Proof. (i) By definition we have

$$\langle y, (A^*)^*x\rangle = \overline{\langle (A^*)^*x, y\rangle} = \overline{\langle x, A^*y\rangle} = \langle A^*y, x\rangle = \langle y, Ax\rangle.$$

Since this holds for all $y \in X$ it follows that $(A^*)^*x = Ax$ for all $x \in X$.

[25] Mathematicians prefer the term symmetric, while in the physics literature the term hermitian prevails

[26] The concept of closed operators is also used for linear operators between Banach spaces X and Y with the obvious modification in the definition. Note that any bounded operator can be extended to a closed operator. However, not all closed operators are continuous.

(ii) It holds that

$$\langle y, (AB)^* x\rangle = \overline{\langle (AB)^* x, y\rangle} = \overline{\langle x, ABy\rangle} = \overline{\langle A^* x, By\rangle} = \overline{\langle B^* A^* x, y\rangle} = \langle y, B^* A^* x\rangle.$$

As in (i) this yields the claim.

(iii) We have that

$$\begin{aligned}\langle y, (A+\lambda B)^* x\rangle &= \overline{\langle (A+\lambda B)^* x, y\rangle} = \overline{\langle x, (A+\lambda B)y\rangle} = \overline{\langle x, Ay\rangle + \lambda\langle x, By\rangle} = \overline{\langle x, Ay\rangle} + \overline{\lambda}\,\overline{\langle x, By\rangle} \\ &= \overline{\langle A^* x, y\rangle} + \overline{\lambda}\,\overline{\langle B^* x, y\rangle} = \langle y, A^* x\rangle + \overline{\lambda}\,\langle y, B^* x\rangle = \langle y, (A^* + \overline{\lambda} B^*)x\rangle.\end{aligned}$$

(iv) For every $x \in X$ we have

$$\|A^* x\|^2 = \langle A^* x, A^* x\rangle = \langle x, AA^* x\rangle \le \|x\|\|AA^* x\| \le \|x\|\|A\|\|A^* x\|. \tag{10}$$

Thus $\|A^* x\| \le \|A\|\|x\|$, which yields that $\|A^*\| \le \|A\|$. Using (i) this further implies that $\|A\| = \|(A^*)^*\| \le \|A^*\|$, so that $\|A\| = \|A^*\|$ as claimed.

(v) Using the elementary inequality $\|AB\| \le \|A\|\|B\|$ together with (iv) we obtain that $\|AA^*\| \le \|A\|\|A^*\| = \|A\|^2$. Moreover, from (10) it follows that $\|A^* x\|^2 \le \|AA^*\|\|x\|^2$, which yields that $\|AA^*\| \ge \|A^*\|^2 = \|A\|^2$. Hence $\|AA^*\| = \|A\|^2$. Using this estimate for $\tilde{A} = A^*$ and (i) we conclude that also $\|A^* A\| = \|A^*\|^2 = \|A\|^2$. □

As an example let us consider the Fourier transform $\mathcal{F} \in \mathcal{L}(L^2(\mathbb{R}^d); L^2(\mathbb{R}^d))$. We claim that its adjoint is $(2\pi)^d$ times the inverse Fourier transform. Indeed, fix $f, g \in L^2(\mathbb{R}^d)$. Since also $\mathcal{F}^{-1}[f] \in L^2(\mathbb{R}^d)$ Plancherel's theorem implies that

$$\langle (2\pi)^d \mathcal{F}^{-1}[f], g\rangle = \langle f, \mathcal{F}[g]\rangle,$$

so that $\mathcal{F}^* = (2\pi)^d \mathcal{F}^{-1}$. In particular, $\mathcal{F}$ is a normal operator since it is continuous and as shown in a previous lecture the Fourier transform commutes with its inverse. In the exercises you will see some further examples of adjoint operators.

As a last topic of this course we introduce the basics for the spectral theory of linear operators. This topic can take a whole course on its own. Therefore we just focus on the very basics so that you are minimally prepared for further courses where linear operators appear.

Definition 4.20. Let X be a complex Hilbert space and $A : D(A) \subset X \to X$ be a linear operator that is densely defined, i.e., $D(A)$ is dense in X.

(a) The set of $\lambda \in \mathbb{C}$ such that $A - \lambda I$ has no bounded inverse is called the spectrum of A and is denoted by $\sigma(A)$;

(b) Assume from now on that A is closed[27] Then the spectrum is further divided into three sets:
   (i) The set of $\lambda \in \mathbb{C}$ such that $A - \lambda I$ is not injective is called the point spectrum of A and we denote it by $\sigma_p(A)$. It consists of the eigenvalues of A.
   (ii) The set of $\lambda \in \mathbb{C}$ such that $A - \lambda I$ is injective, not surjective, but with dense range in X, is called the continuous spectrum of A and we denote it by $\sigma_c(A)$.
   (iii) The set of $\lambda \in \mathbb{C}$ such that $A - \lambda I$ is injective and its range is not dense in X is called the residual spectrum and we denote it by $\sigma_r(A)$.

At first glance there is one case missing, namely that $A - \lambda I$ is bijective, but the inverse is not bounded. This cannot occur for closed operators. Indeed, as we show in the lemma below the inverse of a closed operator is bounded provided its domain is complete. We will use the so-called closed graph theorem which is a deep result characterizing the continuity of linear operators between Banach spaces. Due to a lack of time we cannot prove it in this course.

[27] One can show that when A is not closed then the spectrum satisfies $\sigma(A) = \mathbb{C}$. Indeed, assume that there exists $\lambda \in \mathbb{C}$ such that there exists a bounded, linear inverse $R : X \to D(A)$ of $A - \lambda I$. We show that A must be closed. Let $(x_n)_{n\in\mathbb{N}} \subset D(A)$ be such that $x_n \to x \in X$ and $Ax_n \to y \in X$. Then $x = \lim_{n\to\infty} x_n = \lim_{n\to\infty} R(A - \lambda I)x_n = R(y - \lambda x) \in D(A)$. Moreover, it follows that $(A - \lambda)x = y - \lambda x$, so that $Ax = y$ and thus A is closed. However, no satisfactory spectral theory is available for non-closed operators.

Theorem 4.21 (Closed graph theorem on Banach spaces)**.** *Let X, Y be $\mathbb{K}$-Banach spaces and $A : X \to Y$ be linear. Then A is continuous if and only if A is closed, i.e., for all sequences $(x_n)_{n\in\mathbb{N}} \subset X$ such that $x_n \to x \in X$ and $Ax_n \to y \in Y$ it follows that $Ax = y$.*[28]

Corollary 4.22 (Continuity of the inverse of closed operators)**.** *Let X, Y be $\mathbb{K}$-Hilbert spaces*[29] *and $A : D(A) \subset X \to Y$ be linear, bijective and closed. Then the inverse operator $A^{-1} : Y \to D(A)$ belongs to $\mathcal{L}(Y; X)$.*

Proof. It is a well-known fact that the inverse of a linear function is linear. We show that the inverse operator is closed. Then the statement follows from the closed graph theorem since Y is complete. Note that it does not matter whether $D(A)$ is complete. Let $(y_n)_{n\in\mathbb{N}} \subset \mathrm{Ran}(A) = D(A^{-1})$ be such that $y_n \to y \in Y$ and $A^{-1}y_n \to x \in X$. Then there exists $x_n \in D(A)$ such that $y_n = Ax_n$. Note that $x_n = A^{-1}y_n \to x \in X$ and $Ax_n = y_n \to y \in Y$. Since A is closed it follows that $x \in D(A)$ and $Ax = y$. Hence $y \in D(A^{-1})$ and $A^{-1}y = x$. Thus by definition A^{-1} is closed. □

Note that we did not use that $D(A^{-1})$ is complete to show that the inverse is closed, but this was only used to apply the closed graph theorem. We state this as the final result of the lecture.

Corollary 4.23. *Let X, Y be $\mathbb{K}$-Hilbert spaces and $A : D(A) \subset X \to Y$ be linear, injective and closed. Then the inverse operator $A^{-1} : \mathrm{Ran}(A) \subset Y \to D(A) \subset X$ is linear and closed.*

Concerning the spectrum, note that in finite dimensions we have $\sigma(A) = \sigma_p(A)$. In infinite dimensions this is not true in general. For instance, consider the right-shift $T_R : \ell^2 \to \ell^2$ defined by $T_R((a_n)_{n\in\mathbb{N}}) = (0, a_1, a_2, \ldots)$. If $T_R(a) = \lambda a$, then either $\lambda = 0$, so that $a_n = 0$ for all $n \in \mathbb{N}$. Thus T_R is injective. If $\lambda \neq 0$, then $T_R(a) = \lambda a$ yields $a_1 = 0$ and then by iteration $a_n = 0$ for all $n \geq 2$. Hence also $T_R - \lambda I$ is injective. However, T_R is not surjective since $e_1 \notin \mathrm{Ran}(T_R)$, so that $0 \in \sigma(T_R) \setminus \sigma_p(T_R)$. More precisely, the image of T_R is not dense in ℓ^2 since for each element $a \in \ell^2$ we have $\|T_R(a) - e_1\|_2 \geq 1$. Hence $0 \in \sigma_r(T_R)$. One can show that $\sigma_r(T_R) = B_1(0)$ and $\sigma_c(T_R) = \partial B_1(0)$. There is a rich theory on the spectrum of linear operators, especially for normal or self-adjoint operators. For more information I suggest to attend a course on (linear) functional analysis or to consult the literature.

We close the lecture with a notation that is usually used in quantum mechanics for some objects we introduced in this chapter.

Definition 4.24 (Dirac's Bra-ket notation)**.** Let X be a $\mathbb{K}$-Hilbert space with dual space X'. One writes vectors $x \in X$ as $|x\rangle$ (called 'ket') and bounded, linear functionals $x' \in X'$ as $\langle x'|$ (called 'bra'). The scalar product (or evaluation of x' at x according to the Riesz representation theorem) then takes the form $\langle x'|x\rangle$. Note that one writes only one vertical bar. The symbol $|x\rangle\langle x'|$ means the operator $X \ni y \mapsto x\langle x', y\rangle \in X$.

Email address: matthias.ruf@epfl.ch

[28] The closedness of a linear operator is equivalent to the closedness of its graph $G(A) = \{(x, A(x)) : x \in D(A)\}$ in $X \times Y$. Hence the name closed graph theorem.

[29] The same result holds for operators between Banach spaces X and Y.