Part II

Functional Analysis II

April 3, 2013

12 Spectral theory for bounded operators on complex Hilbert spaces

12.1 Notation

We recall some basic notation and definitions.

Definition 12.1.1. Let E be a Banach space over $\mathbb{K}$ and let $T \in \mathscr{L}(E)$ be a continuous linear mapping on E.

i) The spectrum of T is the set $\sigma(T) \subset \mathbb{K}$ defined as

$$\sigma(T) := \{\lambda \in \mathbb{K} : \lambda I - T \text{ does not have a bounded inverse}\}.$$

ii) $\lambda \in \mathbb{K}$ is called an eigenvalue, if $\ker(\lambda I - T) \neq \{0\}$, i.e. if there is an $x \neq 0$ with $Tx = \lambda x$.

iii) If λ is an eigenvalue, the elements of $\ker(\lambda I - T)$ are called *eigenvectors* of T associated with λ.

We recall the basic properties of $\sigma(T)$:

i) $\lambda \in \sigma(T) \Rightarrow |\lambda| \leq \|T\|$;

ii) $\sigma(T)$ is closed, i.e. the *resolvent set* $\varrho(T) := \mathbb{K} \setminus \sigma(T)$ is open;

iii) If $\mathbb{K} = \mathbb{C}$, then $\sigma(T) \neq \varnothing$.

We shall study linear operators on complex Hilbert spaces. Following definition recalls the basic types of such operators.

Definition 12.1.2. Let H be a Hilbert space over $\mathbb{C}$ and let $T \in \mathscr{L}(H)$.

i) The operator T is called *positive*, if $\langle Tx, x\rangle \geq 0$ for all $x \in H$. We sometimes write this as $T \geq 0$;

ii) There is a unique operator $T^* \in \mathscr{L}(H)$, such that $\langle Tx, y\rangle = \langle x, T^*y\rangle$ for all $x, y \in H$. This operator is called the *adjoint operator* to T;

iii) If $T = T^*$, T is called *self-adjoint*.

iv) If at least $TT^* = T^*T$, then T is called a *normal operator*.

Few basic properties of these classes of operators may be found in Exercises. We give one example here.

Example 12.1.3. We consider the *shift operator* S on ℓ_2, which is defined by

$$S((x_1, x_2, \dots)) = (0, x_1, x_2, \dots).$$

We show that

$$\sigma_p(S) = \varnothing, \quad \sigma_{ap}(S) = \{\lambda \in \mathbb{C} : |\lambda| = 1\}, \quad \sigma(S) = \{\lambda \in \mathbb{C} : |\lambda| \leq 1\},$$

where $\sigma_p(S)$ is the set of eigenvalues of S (the so-called *point spectrum*) and $\sigma_{ap}(S) = \{\lambda \in \mathbb{C} : \inf\{\|Sx - \lambda x\| : \|x\| = 1\} = 0\}$ is the *approximative spectrum*.

Let $Sx = \lambda x$. Then $0 = \lambda x_1$ and $x_j = \lambda x_{j+1}$ for $j \in \mathbb{N}$. If $\lambda \neq 0$, this implies $x_1 = 0$, which implies $x_2 = 0$ and so on, i.e. $x = 0$. If $\lambda = 0$, $x = 0$ follows as well. Hence, there are no eigenvalues of S.

We know, that $\sigma(S) \subset \{\lambda \in \mathbb{C} : |\lambda| \leq 1\}$ (as $\|S\| = 1$). Let on the other hand λ be a complex number with $|\lambda| \leq 1$. Then the equation

$$(S - \lambda I)z = (-\lambda z_1, z_1 - \lambda z_2, z_2 - \lambda z_3, \dots) = (1, 0, 0, \dots)$$

shows that $\lambda \neq 0$ and $z = (-1/\lambda, -1/\lambda^2, -1/\lambda^3, \dots)$, which does not lie in ℓ_2. We conclude, that there is no $z \in \ell_2$ with $(S - \lambda I)z = (1, 0, 0, \dots)$, i.e. $(S - \lambda I)$ does *not* map ℓ_2 *onto* ℓ_2, and is therefore not invertible. Therefore, $\sigma(S) = \{\lambda \in \mathbb{C} : |\lambda| \leq 1\}$.

If $|\lambda| < 1$ and $x \in \ell_2$, we have $\|Sx - \lambda x\| \geq \|Sx\| - \|\lambda x\| = (1 - |\lambda|)\|x\|$, i.e. $\lambda \notin \sigma_{ap}(S)$. Finally, if $|\lambda| = 1$, we put

$$x^n = \frac{1}{\sqrt{n}} \cdot (1, \lambda^{-1}, \lambda^{-2}, \dots, \lambda^{-(n-1)}, 0, \dots).$$

It follows, that $\|x^n\| = 1$ for all $n \in \mathbb{N}$ and

$$\|Sx^n - \lambda x^n\| = (2/n)^{1/2} \to 0$$

as $n \to \infty$, i.e. $\sigma_{ap}(S) = \{\lambda \in \mathbb{C} : |\lambda| = 1\}$.

The main aim of this section is to generalize the Spectral Theorem for compact self-adjoint operators (Theorem 11.8) to bounded self-adjoint (or even normal) operators.

12.2 Functional calculus

Through the rest of this section, we shall work only with a Hilbert space H over $\mathbb{C}$.

Definition 12.2.1. Let $T \in \mathscr{L}(H)$. Then $W(T) = \{\langle Tx, x\rangle : \|x\| = 1\}$ is called a *numerical range* of T.

Lemma 12.2.2. *Let $T \in \mathscr{L}(H)$. Then $\sigma(T) \subset \overline{W(T)}$.*

Proof. Let $\lambda \notin \overline{W(T)}$ and put $d := \operatorname{dist}(\lambda, W(T)) = \inf\{|\lambda - \mu| : \mu \in W(T)\} > 0$. Then for $\|x\| = 1$ we have

$$d \leq |\lambda - \langle Tx, x\rangle| = |\langle \lambda x - Tx, x\rangle| \leq \|(\lambda I - T)x\| \cdot \|x\| = \|(\lambda I - T)x\|.$$

Therefore, $\lambda I - T$ is injective and $(\lambda I - T)^{-1} : \operatorname{ran}(\lambda I - T) \to H$ with norm at most $1/d$. Hence, $\operatorname{ran}(\lambda I - T)$ and H are isomorphic and $\operatorname{ran}(\lambda I - T)$ is closed.

Let us assume, that there is $x_0 \in \operatorname{ran}(\lambda I - T)^{\perp}$ with $\|x_0\| = 1$. Then we obtain

$$0 = \langle (\lambda I - T)x_0, x_0\rangle = \lambda - \langle Tx_0, x_0\rangle$$

and $\lambda \in W(T)$, which is a contradiction. Hence $\operatorname{ran}(\lambda I - T) = H$ and $\lambda \in \varrho(T)$. $\square$

If T is self-adjoint, then $\langle Tx, x\rangle \in \mathbb{R}$ for every $x \in H$ and the previous lemma gives another proof of $\sigma(T) \subset \mathbb{R}$. A bit more is even true.

Corollary 12.2.3. *If $T \in \mathscr{L}(H)$ is self-adjoint, then*

$$\sigma(T) \subset [m(T), M(T)],$$

where $m(T) = \inf\{\langle Tx, x\rangle : \|x\| = 1\}$ and $M(T) = \sup\{\langle Tx, x\rangle : \|x\| = 1\}$. If $T \geq 0$ (i.e. T is positive), then $\sigma(T) \subset [0, \infty)$.

Functional calculus is a way, how to define $f(T)$ for (at least some) functions f and operators T. Surely, if $T \in \mathscr{L}(H)$ and $p(z) = \sum_{k=0}^{n} a_k z^k$ is a polynomial, we would like to have $p(T) = \sum_{k=0}^{n} a_k T^k$.

Moreover, for $H = \mathbb{C}^n$ and T self-adjoint (i.e. a hermitian matrix), we may consider the diagonalization

$$T = U^{-1}DU,$$

where U is unitary[4] and D is diagonal. Let $f(z) = \sum_{k=0}^{\infty} a_k z^k$ be holomorph on the whole $\mathbb{C}$, then (at least formally)

$$\begin{aligned} f(T) &= \sum_{k=0}^{\infty} a_k T^k = \sum_{k=0}^{\infty} a_k (U^{-1}DU)^k = \sum_{k=0}^{\infty} a_k U^{-1} D^k U = U^{-1} \left(\sum_{k=0}^{\infty} a_k D^k \right) U \\ &= U^{-1} f(D) U, \end{aligned}$$

where

$$f(D) = \begin{pmatrix} f(d_1) & 0 & 0 & \dots \\ 0 & f(d_2) & 0 & \dots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & f(d_n) \end{pmatrix}$$

is again a diagonal matrix. We see, that to define $f(T)$, f needs to be defined only on $\sigma(T)$ and "$f(\sigma(T)) = \sigma(f(T))$", the so-called *Spectral Mapping Theorem.*

Theorem 12.2.4. (Continuous Functional Calculus) *Let $T \in \mathscr{L}(H)$ be self-adjoint. Then there is a unique mapping*

$$\Phi : C(\sigma(T)) \to \mathscr{L}(H),$$

such that

i) $\Phi(z)$[5] $= T, \Phi(1) = I$*;*

ii) Φ *is an involutive homomorphism of algebras, i.e.*

a) Φ *is linear;*

b) Φ *is multiplicative:* $\Phi(fg) = \Phi(f) \circ \Phi(g)$*;*

c) Φ *is involutive:* $\Phi(f)^* = \Phi(\overline{f})$*;*

iii) Φ *is continuous.*

Φ *is called* continuous functional calculus *of* T *and we denote* $f(T) := \Phi(f)$ *for* $f \in C(\sigma(T))$.

Proof. Uniqueness: $\Phi(z^n) = T^n$ and Φ is unique on all polynomials. Furthermore, $\sigma(T) \subset [m(T), M(T)]$ is compact and polynomials are dense in $C(\sigma(T))$. [6] Due to continuity, Φ is unique on all $C(\sigma(T))$.

[4] i.e. $U^*U = UU^* = I$

[5] By z we denote of course the identity mapping $z \to z$ on $\mathbb{C}$ and by 1 the mapping $z \to 1$.

[6] We assume, that you know, that polynomials are dense in $C(I)$ for every bounded interval I.

Existence: We set $\Phi(f) = \sum_{k=0}^{n} a_k T^k$ for $f(z) = \sum_{k=0}^{n} a_k z^k$ a polynomial. If we show continuity of Φ on polynomials, then there would be a unique extension, which we denote by Φ again.

We shall use the *Spectral Mapping Theorem* for polynomials (cf. Exercise 1.4)

$$\sigma(\Phi(f)) = f(\sigma(T)) = \{f(\lambda) : \lambda \in \sigma(T)\}$$

to obtain

$$\begin{aligned}
\|\Phi(f)\|^2 &= \|\Phi(f)^*\Phi(f)\| = \|\Phi(\overline{f}f)\| \\
&= \sup\{|\lambda| : \lambda \in \sigma(\Phi(\overline{f}f))\} \\
&= \sup\{|\overline{f}f(\lambda)| : \lambda \in \sigma(T)\} \\
&= \sup\{|f(\lambda)|^2 : \lambda \in \sigma(T)\} = \|f\|_\infty^2.
\end{aligned}$$

We refer to Exercise 1.1.2 for the first equality, the second equality follows from a simple calculation for polynomials, the third one is based on Exercise 2.3 and the fact that $\Phi(\overline{f}f)$ is self-adjoint.

All the required properties are then proven by approximation. Let us assume (for example) that $p_n \rightrightarrows f$ and $q_n \rightrightarrows g$ for some polynomials p_n, q_n and $f, g \in C(\sigma(T))$. Then we get

$$\Phi(fg) \leftarrow \Phi(p_n q_n) = \Phi(p_n) \circ \Phi(q_n) \to \Phi(f) \circ \Phi(g).$$

The proof of the other properties is similar. □

Theorem 12.2.5. (Properties of the continuous functional calculus)
Let $T \in \mathscr{L}(H)$ be self-adjoint and let $f \to f(T)$ be the continuous functional calculus for $f \in C(\sigma(T))$. Then

- *i)* $\|f(T)\| = \|f\|_\infty := \sup_{\lambda \in \sigma(T)} |f(\lambda)|$.
- *ii) If $f \geq 0$ on $\sigma(T)$, then $f(T) \geq 0$, i.e. $f(T)$ is positive.*
- *iii) If $Tx = \lambda x$ for some $x \in H$, then also $f(T)x = f(\lambda)x$.*
- *iv) $\sigma(f(T)) = f(\sigma(T))$, i.e. the Spectral Mapping Theorem holds for all $f \in C(\sigma(T))$.*
- *v) $\{f(T) : f \in C(\sigma(T))\}$ is a commutative Banach algebra of operators.*
- *vi) All $f(T)$ are normal; if f is real, then $f(T)$ is self-adjoint.*

Proof. (i) was proven for polynomials already before. For general $p_n \rightrightarrows f$ we have $\|\Phi(f)\| \leftarrow \|\Phi(p_n)\| = \|p_n\|_\infty \to \|f\|_\infty$.
Let $f \geq 0$ and take $g \in C(\sigma(T))$ with $g^2 = f$ and $g \geq 0$. Then

$$\langle f(T)x, x\rangle = \langle g(T)x, g(T)^*x\rangle = \langle g(T)x, \overline{g}(T)x\rangle = \langle g(T)x, g(T)x\rangle = \|g(T)x\|^2 \geq 0.$$

(iii) is again clear for polynomials, through approximation it follows for all $f \in C(\sigma(T))$. Also (iv) was already discussed for polynomials. Let $\mu \notin f(\sigma(T))$. Then $g := (f - \mu)^{-1} \in C(\sigma(T))$ and $g(f - \mu) = (f - \mu)g = 1$. Hence, we get

$$g(T)(f(T) - \mu I) = (f(T) - \mu I)g(T) = I.$$

Hence $\mu \in \varrho(f(T))$. This shows the "$\subset$" part in (iv).
Let on the other hand $\mu = f(\lambda)$ for some $\lambda \in \sigma(T)$ and choose polynomials p_n with $\|p_n - f\|_\infty \le 1/n$. Then

$$|f(\lambda) - p_n(\lambda)| \le 1/n \quad \text{and} \quad \|f(T) - p_n(T)\| \le 1/n.$$

We know that $p_n(\lambda) \in \sigma(p_n(T))$, i.e. there exists $x_n \in H$ with $\|x_n\| = 1$ and $\|(p_n(T) - p_n(\lambda)I)x_n\| \le 1/n$. Finally, from

$$\begin{aligned}\|(f(T) - \mu I)x_n\| &\le \|(f(T) - p_n(T))x_n\| + \|(p_n(T) - p_n(\lambda)I)x_n\| + \|(p_n(\lambda) - \mu)Ix_n\| \\ &\le 1/n + 1/n + 1/n = 3/n\end{aligned}$$

we see, that $f(T) - \mu I$ is not cont. invertible, i.e. $\mu \in \sigma(f(T))$. Finally, (v) is clear now and (vi) follows from $f(T)^* = \overline{f}(T)$ and $f(T)^* f(T) = \overline{f} f(T) = f\overline{f}(T) = f(T)f(T)^*$. $\square$

Let us now have a look on the spectral decomposition theorem for compact self-adjoint operators and how it could be generalized to bounded operators. One obvious obstacle would be that the spectrum does not have to be countable any more, i.e. we will have to replace the sums by certain (appropriately interpreted) integrals. Furthermore, we have now spent some time to define $f(T)$ for (at least some) functions f on $\sigma(T)$. We therefore rewrite

$$Tx = \sum_{k=0}^{\infty} \lambda_k \langle x, x_k \rangle x_k = \sum_{k=0}^{\infty} \lambda_k E_k x,$$

where $E_k x = \langle x, x_k \rangle x_k$ is the projection of x onto the linear span of x_k. And if we put $f_k(\lambda_j) = \delta_{j,k}$, we get even

$$f_k(T)x = \sum_{n=0}^{\infty} f_k(\lambda_n) \langle x, x_n \rangle x_n = \langle x, x_k \rangle x_k = E_k x,$$

where (for simplicity) we assumed that all the λ_k's are different.

We will therefore write the spectral decomposition of a bounded self-adjoint operator as an integral with respect to a certain system of projections. Unfortunately, f_k do not have to be continuous on $\sigma(T)$. We therefore have to extend the definition of $f(T)$ to bounded measurable functions f. This shall be done through duality. For this sake, we shall need (another) Riesz Representation Theorem, and for this we shall need some notation.

Let (K, τ) be a compact topological space. We denote by $\mathcal{B}$ the smallest σ-algebra containing τ (the so-called *Borel σ-algebra*). As usual, $f : K \to \mathbb{C}$ is called *Borel measurable*, if $f^{-1}(U) \in \mathcal{B}$ for every $U \subset \mathbb{C}$ open.

The interaction between measures on $\mathcal{B}$ and the topology τ is described by the following regularity conditions.

A finite measure μ on $(K, \mathcal{B})$ is said to be a *Radon measure*, if

$$\begin{aligned}&\mu(B) = \inf\{\mu(G) : G \supset B, G \text{ open}\} \text{ for every } B \in \mathcal{B}, \\ &\mu(G) = \sup\{\mu(T) : T \subset G, T \text{ compact}\} \text{ for every open } G \in \tau.\end{aligned}$$

Signed Radon measure is a σ-additive mapping on $\mathcal{B}$ with $\mu = \mu^+ - \mu^-$, where both μ^+ and μ^- are Radon measures. Unfortunately, this representation is not unique. Indeed, if ν is another bounded positive Radon measure, we get $\mu = \mu^+ - \mu^- = (\mu^+ + \nu) - (\mu^- + \nu)$. Nevertheless, one of these representations is "minimal" (i.e. μ^+ and μ^- are singular to each other).[7]

[7]The decomposition of the measure space goes under the name *Hahn decomposition* and the decomposition of the measure is known as *Jordan* or *Hahn-Jordan decomposition*.

Using this minimal decomposition, we denote by $|\mu| = \mu^+ + \mu^-$ the *total variation* of μ. If μ is complex-valued, we furthermore decompose it into the real and complex part.

Finally, we denote by $\mathcal{M}(K)$ the space of all signed Radon measures on K with $\|\mu\|_{\mathcal{M}(K)} = |\mu|(K)$.

Theorem 12.2.6. (Yet Another Riesz Representation Theorem)

$$C(K)^* \approx \mathcal{M}(K),$$

where $\mu(f) = \int_K f d\mu$. Furthermore, positive functionals (i.e. those with $f \geq 0 \Rightarrow \varphi(f) \geq 0$) correspond to non-negative measures.

Lemma 12.2.7. *Let H be a complex Hilbert space and let $Q : H \to \mathbb{C}$ be a function. Then the following are equivalent.*

i) There exists exactly one operator $A \in \mathscr{L}(H)$ such that $Q(x) = \langle Ax, x\rangle$ for all $x \in H$.

ii) There is a constant $C > 0$ such that $|Q(x)| \leq C\|x\|^2$, $Q(x+y) + Q(x-y) = 2Q(x) + 2Q(y)$ and $Q(\lambda x) = |\lambda|^2 Q(x)$ for all $x, y \in H$ and all $\lambda \in \mathbb{C}$.

Proof. It is easy to show that $(i) \Rightarrow (ii)$. If (ii) is satisfied, we put (motivated by polarization identity)

$$\Psi(x,y) = 1/4(Q(x+y) - Q(x-y) + iQ(x+iy) - iQ(x-iy))$$

and

$$Ax = \sum_{\gamma\in\Gamma} \Psi(x, e_\gamma) e_\gamma,$$

where $(e_\gamma)_{\gamma\in\Gamma}$ is some orthonormal basis of H. The rest follows by simple calculations. Finally, let us observe that $\Psi(x,y) = \langle Ax, y\rangle$. Uniqueness of A follows from Exercises. $\square$

Remark 12.2.8. *(The main idea of Borel-measurable functional calculus)*
Let $T \in \mathscr{L}(H)$ be self-adjoint and let $x \in H$ be fixed. We consider the mapping

$$f \to \langle f(T)x, x\rangle, \quad C(\sigma(T)) \to \mathbb{C}.$$

This mapping is linear, non-negative (i.e. $f \geq 0$ implies $\langle f(T)x, x\rangle \geq 0$ for all $x \in H$). Therefore, there exists a non-negative Radon measure E^x on $\sigma(T)$, such that

$$\langle f(T)x, x\rangle = \int_{\sigma(T)} f dE^x \quad \text{for all} \quad f \in \mathcal{C}(\sigma(T)).$$

If g is a bounded Borel-measurable function on $\sigma(T)$, the integral $\displaystyle\int_{\sigma(T)} g dE^x$ converges and we shall show (using the Lemma above) that it is equal to some $\langle Gx, x\rangle$ for all $x \in H$ and some $G \in \mathscr{L}(H)$. This G will then be defined to be $g(T)$.

Let us now elaborate on the program just stated.

Theorem 12.2.9. *Let $T \in \mathscr{L}(H)$ be self-adjoint.*

i) Let $x \in H$. Then there is a non-negative Radon measure E^x, such that

$$\langle f(T)x, x\rangle = \int_{\sigma(T)} f dE^x$$

for every $f \in C(\sigma(T))$.

ii) Let g be a bounded Borel-measurable function on $\sigma(T)$. Then there is a unique $G \in \mathscr{L}(H)$ with

$$\langle Gx, x\rangle = \int_{\sigma(T)} g dE^x$$

for all $x \in H$. If g is real, G is self-adjoint, if $g \geq 0$, G is positive.

Proof. The proof is based on Lemma 12.2.7. Obviously, the operator is unique ($\langle Gx, x\rangle = \langle \tilde{G}x, x\rangle$ for all $x \in H$ implies $G = \tilde{G}$). Furthermore, $g \geq 0$ implies $\langle Gx, x\rangle \geq 0$, i.e. positivity, and g real implies $\langle Gx, x\rangle$ is real, i.e. G is self-adjoint.
So, we have to prove the existence. We put

$$Q(x) = \int_{\sigma(T)} g dE^x.$$

Then

$$|Q(x)| \leq \int_{\sigma(T)} \|g\|_\infty dE^x = \|g\|_\infty \int_{\sigma(T)} 1 dE^x = \|g\|_\infty \langle 1(T)x, x\rangle = \|g\|_\infty \|x\|^2. \tag{12.1}$$

For every $f \in C(\sigma(T))$, we obtain

$$\underbrace{\int_{\sigma(T)} f dE^{x+y}}_{\langle f(T)(x+y), x+y\rangle} + \underbrace{\int_{\sigma(T)} f dE^{x-y}}_{\langle f(T)(x-y), x-y\rangle} = \underbrace{2\int_{\sigma(T)} f dE^x}_{2\langle f(T)x, x\rangle} + \underbrace{2\int_{\sigma(T)} f dE^y}_{2\langle f(T)y, y\rangle}$$

and

$$\underbrace{\int_{\sigma(T)} f dE^{\lambda x}}_{\langle f(T)(\lambda x), \lambda x\rangle} = \underbrace{|\lambda|^2 \int_{\sigma(T)} f dE^x}_{|\lambda|^2 \cdot \langle f(T)x, x\rangle}$$

Due to the uniqueness of the Riesz Representation Theorem, we get $E^{x+y} + E^{x-y} = 2E^x + 2E^y$ and $E^{\lambda x} = |\lambda|^2 E^x$. This verifies the assumptions of Lemma 12.2.7 and finishes the proof. ☐

Yet another theorem from the measure theory we shall use is that the set of bounded Borel-measurable functions is the smallest set, which includes bounded continuous functions and is closed under pointwise limits of such functions.

Theorem 12.2.10. *Let $K \subset \mathbb{C}$ be compact. Let $\mathcal{B}^d(K)$ be the Banach space of bounded Borel-measurable functions on K equipped with the supremum norm. Let $C(K) \subset U \subset \mathcal{B}^d(K)$ be a set of functions with the following property*

$$(f_n)_{n\in\mathbb{N}} \subset U \textit{ with } f(t) := \lim_{n\to\infty} f_n(t) \textit{ existing everywhere and } \sup_{n\in\mathbb{N}} \|f_n\|_\infty < \infty \Rightarrow f \in U.$$

Then $U = \mathcal{B}^d(K)$.

Proof. The proof is based on measure theory. We refer to Werner, Lemma VII.1.5 for details. ☐

Theorem 12.2.11. (Borel measurable functional calculus) *Let $T \in \mathscr{L}(H)$ be self-adjoint. Then there exists a unique mapping $\Phi : \mathcal{B}^d(\sigma(T)) \to \mathscr{L}(H)$ with*

i) $\Phi(z) = T, \Phi(1) = I$;

ii) Φ *is an involutive homomorphism of algebras;*

iii) Φ *is continuous;*

iv) $f_n \in \mathcal{B}^d(\sigma(T))$ *with* $\sup_n \|f_n\|_\infty < \infty$ *and* $f_n(t) \to f(t)$ *for every* $t \in \sigma(T)$ *implies* $\langle \Phi(f_n)x, y\rangle \to \langle \Phi(f)x, y\rangle$ *for every* $x, y \in H$.

Proof. Uniqueness: $(i), (ii)$ and (iii) imply uniqueness on $C(\sigma(T))$ and Theorem 12.2.10 on all $\mathcal{B}^d(\sigma(T))$.
Existence: We define $\Phi(g) = G$, where G is the operator from Theorem 12.2.9. We have to verify $(i) - (iv)$. We know that

$$\langle f(T)x, x\rangle = \int_{\sigma(T)} f dE^x$$

for all $f \in C(\sigma(T))$, i.e. $\Phi(f) = f(T)$ for all $f \in C(\sigma(T))$, where $f(T)$ is the continuous functional calculus. This implies (i).

Let g be real valued. Then we obtain

$$\|\Phi(g)\| = \|G\| = \sup\{|\langle Gx, x\rangle| : \|x\| = 1\} \leq \sup\{\|g\|_\infty \cdot \|x\|^2 : \|x\| = 1\} = \|g\|_\infty,$$

where we have used (12.1), the fact, that G is self-adjoint and Corollary 11.6. If g is complex-valued, we split it first into its real and imaginary part.
The proof of (iv) follows from

$$\langle \Phi(f_n)x, x\rangle = \int_{\sigma(T)} f_n dE^x \to \int_{\sigma(T)} f dE^x = \langle \Phi(f)x, x\rangle,$$

which holds by Lebesgue Dominated Convergence Theorem, and by polarization identity.
(ii) follows from a bit tricky limit procedure, which is quite often used in measure theory. For example, to show that $\Phi(fg) = \Phi(f) \circ \Phi(g)$, we proceed as follows. If $f, g \in C(\sigma(T))$, then the result is known. If $g \in C(\sigma(T))$, we set $U := \{f \in \mathcal{B}^d(\sigma(T)) : \Phi(fg) = \Phi(f) \circ \Phi(g)\}$. Then we know that $C(\sigma(T)) \subset U$ and U is closed under pointwise limits of uniformly bounded sequences. According to Theorem 12.2.10, $U = \mathcal{B}^d(\sigma(T))$.
Finally, if $f \in \mathcal{B}^d(\sigma(T))$, we put $V := \{g \in \mathcal{B}^d(\sigma(T)) : \Phi(fg) = \Phi(f) \circ \Phi(g)\}$. We have just shown that $C(\sigma(T)) \subset V$, furthermore, V is again closed on pointwise limits of uniformly bounded functions, i.e. $V = \mathcal{B}^d(\sigma(T))$ and the statement is true for all $f, g \in \mathcal{B}^d(\sigma(T))$.
Other properties follow in the same way. $\square$

Remark 12.2.12. (iv) in Theorem 12.2.11 can be improved to $f_n(T)x \to f(T)x$ for every $x \in H$. This follows from

$$\begin{aligned}\|f_n(T)x\|^2 &= \langle f_n(T)x, f_n(T)x\rangle = \langle f_n(T)^* f_n(T)x, x\rangle \\ &= \langle (\overline{f}_n f_n)(T)x, x\rangle \to \langle (\overline{f} f)(T)x, x\rangle = \|f(T)x\|^2\end{aligned}$$

and $f_n(T)x \xrightarrow{\text{w}} f(T)x$.

Remark 12.2.13. (Analytic functional calculus) We were able to define $f(T)$ for (rather) ugly functions (i.e. all Borel-measurable functions) and very nice operators (i.e. self-adjoint operators on a Hilbert space over $\mathbb{C}$). It comes probably as no surprise that for nicer function we can define $f(T)$ also for more general classes of operators.

Let X be a Banach space and let $T \in L(X)$. Let $f(z) = \sum_{n=0}^{\infty} a_n z^n$ be a potential series with convergence radius bigger then $r(T)$. Then f is analytic on $\sigma(T)$ and we can define

$$f(T) := \sum_{n=0}^{\infty} a_n T^n.$$

One can show that

The series for $f(T)$ converges in $L(X)$ and $f(T)$ is therefore well-defined.

If $g(z) = \sum_{n=0}^{\infty} b_n z^n$ is another potential series with radius of convergence bigger then $r(T)$, then $(fg)(T) = f(T)g(T)$.

The Spectral mapping theorem $\sigma(f(T)) = f(\sigma(T))$ holds.

If T is a self-adjoint operator on a Hilbert space over $\mathbb{C}$, then this definition coincides with the continuous (and also with the Borel-measurable) functional calculus.

Remark 12.2.14. (Dunford's analytic functional calculus) The analytic functional calculus described above allows to study spectral theory also in the frame of Banach spaces. Nevertheless, the basis for such a theory is different. If f is holomorph on a neighborhood Ω of $\sigma(T)$ (i.e. on an open set containing $\sigma(T)$), then we may define

$$f(T) := \frac{1}{2\pi i} \oint_{\Gamma} f(\lambda) R_\lambda(T) d\lambda,$$

where $\Gamma \subset \Omega$ is a suitable curve, which goes around $\sigma(T)$ exactly one times, and $R_\lambda(T)$ is the resolvent mapping of T. Of course, this idea is heavily inspired by the Cauchy integral known from complex analysis.

12.3 Spectral theorem for bounded self-adjoint operators

For spectral theorem, $f(T)$ is especially important for $f = \chi_A$, where $A \subset \mathbb{R}$ is a Borel measurable set.

Lemma 12.3.1. *Let $T \in \mathscr{L}(H)$ be self-adjoint and let $A \subset \sigma(T)$ be a Borel set. Then $E_A := \chi_A(T)$ is an orthogonal projection.*

Proof. Due $\chi_A^2 = \chi_A$, we have $E_A = E_A^2$ and $E_A^* = E_A$ follows from $\overline{\chi_A} = \chi_A$ and Exercise 3.1. □

Lemma 12.3.2. *Let $T \in \mathscr{L}(H)$ be self-adjoint. Then*

i) $E_\varnothing = \chi_\varnothing(T) = 0, E_{\sigma(T)} := \chi_{\sigma(T)}(T) = I$;

ii) For pairwise disjoint sets $A_1, A_2, \dots \subset \sigma(T)$, and for $x \in H$, we have

$$\sum_{i=1}^{\infty} \chi_{A_i}(T)x = \chi_{\bigcup_{i=1}^{\infty} A_i}(T)x.$$

iii) $\chi_A(T)\chi_B(T) = \chi_{A\cap B}(T)$ *for Borel sets $A, B \subset \sigma(T)$.*

Proof. The proof follows from the properties of the Borel-measurable functional calculus. □

Remark 12.3.3. Let us note, that

$$\sum_{i=1}^{\infty} \chi_{A_i}(T) \neq \chi_{\bigcup_{i=1}^{\infty} A_i}(T),$$

i.e. the identity must be interpreted in the "weak sense".

Definition 12.3.4. Let Σ be the σ-algebra of the Borel sets of $\mathbb{R}$. A mapping

$$E : \Sigma \to \mathscr{L}(H), \quad E(A) = E_A$$

is called a *spectral measure* if all E_A are orthogonal projections and

i) $E_\varnothing = 0, E_\mathbb{R} = I$;

ii) For pairwise disjoint sets $A_1, A_2, \dots \in \Sigma$, and for $x \in H$, we have

$$\sum_{i=1}^{\infty} E_{A_i}(x) = E_{\bigcup_{i=1}^{\infty} A_i}(x).$$

A spectral measure E has *compact support*, if there is a compact set K with $E_K = I$.

Let us state some simple properties of spectral measures.

i) $E_A + E_B = E_{A\cap B} + E_{A\setminus B} + E_B = E_{A\cap B} + E_{A\cup B}$ holds for every $A, B \subset \mathbb{R}$.

ii) If $A \subset B \subset \mathbb{R}$, then $E_{B\setminus A} = E_B - E_A$ is itself a projection, i.e. $E_B \geq E_A$ and (with the help of Exercises) $E_B E_A = E_A E_B = E_A$.

iii) If A and B are disjoint, we get $E_A + E_B = E_{A\cup B}$, i.e. $E_A^2 + E_A E_B = E_A E_{A\cup B} = E_A$ due to the previous point; hence $E_A E_B = 0$.

iv) $E_A E_B = E_A(E_{A\cap B} + E_{B\setminus A}) = E_A E_{A\cap B} + E_A E_{B\setminus A} = E_{A\cap B}$ holds (due to the previous points) for all $A, B \subset \mathbb{R}$.

v) Finally, if $A, B \subset \mathbb{R}$ are disjoint, then $\langle E_A x, E_B y\rangle = \langle E_B^* E_A x, y\rangle = \langle E_B E_A x, y\rangle = \langle 0, y\rangle = 0$, i.e. E_A and E_B project onto mutually orthogonal subspaces.

"In some sense, E is a measure with values in $\mathscr{L}(H)$ instead of $[0, \infty)$."

For the sake of spectral decomposition theorem, we need to elaborate a way, how to calculate $\int f dE \in \mathscr{L}(H)$ for E a spectral measure and a complex-valued f. Of course, for $f = \chi_A$, we define $\int f dE = E(A) = E_A \in \mathscr{L}(H)$. For $f = \sum_{i=1}^{n} \alpha_i \chi_{A_i}$ a simple function, we put $\int f dE = \sum_{i=1}^{n} \alpha_i E_{A_i} \in \mathscr{L}(H)$. Finally, for a general bounded Borel-measurable function f we consider uniformly convergent sequence $f_n \rightrightarrows f$ of simple functions and define

$$\int f dE = \lim_{n\to\infty} \int f_n dE.$$

We have to check, that this definition is consistent. It is quite easy to see that the integral of a simple function f does not depend on the form in which f was written. The limit procedure is then justified by the following Lemma:

Lemma 12.3.5. *For a simple function f, we have*

$$\left\| \int f dE \right\| \leq \|f\|_\infty.$$

Proof. Let $\|x\| = 1$ and let $f = \sum_{i=1}^{n} \alpha_i \chi_{A_i}$ with mutually disjoint $A_1, \ldots, A_n$. Then

$$\begin{aligned} \left\| \left(\int f dE \right)(x) \right\|^2 &= \left\| \sum_{i=1}^{n} \alpha_i E_{A_i}(x) \right\|^2 = \sum_{i=1}^{n} |\alpha_i|^2 \cdot \|E_{A_i}(x)\|^2 \\ &\leq \max_{i=1,\ldots,n} |\alpha_i|^2 \cdot \sum_{i=1}^{n} \|E_{A_i}(x)\|^2 \leq \|f\|_\infty^2 \cdot \|x\|^2 = \|f\|_\infty^2. \end{aligned}$$

□

This lemma shows that if $(f_n)_{n\in\mathbb{N}}$ is a Cauchy-sequence in $\|\cdot\|_\infty$ (i.e. $f_n \rightrightarrows f$), then $(\int f_n dE)_n$ is a Cauchy-sequence in $\mathscr{L}(H)$, its limit exists and it does not depend on the choice of the sequence $(f_n)_{n\in\mathbb{N}}$. Therefore, $\int f dE$ is well defined.

We summarize our findings in the following theorem:

Theorem 12.3.6. (Properties of integration with respect to the spectral measure) *Let E be a spectral measure. Then $\int f dE \in \mathscr{L}(H)$ exists for every bounded Borel measurable function $f \in \mathcal{B}^d(\mathbb{R})$. The mapping $f \to \int f dE$ is linear and continuous, i.e. $\left\| \int f dE \right\| \leq \|f\|_\infty$. If f is real, then $\int f dE$ is self-adjoint. If $K \subset \mathbb{R}$ is compact with $E_K = I$, then it is enough if f is defined and bounded on K.*

Remark 12.3.7. On one hand, we have for each $T \in \mathscr{L}(H)$ self-adjoint the spectral measure $E_A := \chi_A(T)$. On the other hand, we can define for every spectral measure with compact support the operator $T := \int \lambda dE_\lambda$, i.e. we integrate the identity function $\lambda \to \lambda$ with respect to E. We shall show that these two operations are actually inverse to each other.

Theorem 12.3.8. *Let E be a spectral measure on $\mathbb{R}$ with compact support. Let $T \in \mathscr{L}(H)$ be the self-adjoint operator defined by $T = \int \lambda dE_\lambda$. Then the mapping*

$$\Psi : f \to \int_{\sigma(T)} f dE, \quad \Psi : \mathcal{B}^d(\sigma(T)) \to \mathscr{L}(H)$$

is the Borel-measurable functional calculus from Theorem 12.2.11. Especially, we have $E_{\sigma(T)} = I$.

Proof. We sketch first the proof that $E_{\sigma(T)} = I$. Let $f(\lambda) := \lambda$ and let $f_\delta \rightrightarrows f$ be simple functions $f_\delta = \sum_{i=1}^{n(\delta)} \alpha_{i,\delta} \chi_{A_{i,\delta}}$ with $|f_\delta(\lambda) - \lambda| \leq \delta$ on the support of E and $A_{1,\delta}, \ldots, A_{n(\delta),\delta}$ mutually disjoint. Let $\mu \in \varrho(T)$. As $\varrho(T)$ is open and $\left\| \int_{\mathbb{R}} f_\delta dE - T \right\| = \left\| \int_{\mathbb{R}} (f_\delta(\lambda) - \lambda) dE \right\| \leq \delta$, we conclude, that $\int_{\mathbb{R}} f_\delta dE - \mu I = \sum_{i=1}^{n(\delta)} \alpha_{i,\delta} E_{A_{i,\delta}} - \mu I$ is invertible for $0 < \delta \leq \delta_0$ (for $\delta_0 > 0$ small enough). On the other hand, the inverse of such an operator has a norm $\max\{|\alpha_{i,\delta} - \mu|^{-1} : 1 \leq i \leq n(\delta), E_{A_{i,\delta}} \neq 0\} \to \|(T - \mu I)^{-1}\|$. Hence, $E_{A_{i,\delta}} = 0$ for

$|\alpha_{i,\delta} - \mu|$ small and we conclude that $E_U = 0$ for some small neighborhood of $\mu \in U$. Hence $E_{\varrho(T)} = 0$ and $E_{\sigma(T)} = I$.
To conclude the proof, we have to verify that the mapping Ψ has all the properties described in Theorem 12.2.11. We know already, that it is linear, continuous, multiplicative (start with characteristic functions, then simple functions and then limits) and involutive. The condition (iv) from Theorem 12.2.11 reads in our case as

$$\left\langle \left(\int f_n dE \right) x, y \right\rangle \to \left\langle \left(\int f dE \right) x, y \right\rangle$$

and follows from the fact, that the mapping $g \to \langle (\int g dE)x, y\rangle$ is in $C(\sigma(T))^*$ and may be therefore represented by a measure $\nu_{x,y}$, i.e. $\langle (\int g dE)x, y\rangle = \int g d\nu_{x,y}$. The proof is then finished by Lebesgue dominated convergence theorem. □

Hence, integration with respect to a spectral measure E is a functional calculus of the self-adjoint operator $\int \lambda dE_\lambda$. The spectral theorem says, that actually every self-adjoint operator may be written in this way - i.e. there is a one-to-one correspondence between bounded self-adjoint operators and spectral measures with compact support.

Theorem 12.3.9. (Spectral theorem for bounded self-adjoint operators)
Let $T \in \mathscr{L}(H)$ be self-adjoint. Then there exists a unique spectral measure E on $\mathbb{R}$ with compact support with

$$T = \int_{\sigma(T)} \lambda dE_\lambda.$$

The mapping $f \to f(T) = \int f(\lambda) dE_\lambda$ coincides with the Borel-measurable functional calculus and

$$\langle f(T)x, y\rangle = \int_{\sigma(T)} f(\lambda) d\langle E_\lambda x, y\rangle.$$

By $\langle E_\lambda x, y\rangle$ we denote the (complexed-valued) measure $A \to \langle E_A x, y\rangle$.

Proof. Let E be the spectral measure associated to T as described above, i.e. $E : A \to \chi_A(T)$. We have to show, that $S := \int_{\sigma(T)} \lambda dE_\lambda$ is equal to T. Let f be a simple function with $\|f(\lambda) - \lambda\|_\infty = \sup_{\lambda\in\sigma(T)} |f(\lambda) - \lambda| \le \delta$. Then

$$\|T - S\| \le \|T - f(T)\| + \|f(T) - f(S)\| + \|f(S) - S\|.$$

We estimate all the three summands as follows. Let Ψ be the Borel-measurable functional calculus of T. Then $\Psi(\lambda) = T$ and $\Psi(f) = f(T)$. Hence

$$\|T - f(T)\| = \|\Psi(\lambda) - \Psi(f)\| = \|\Psi(\lambda - f(\lambda))\| \le \|f(\lambda) - \lambda\|_\infty \le \delta,$$

where we have used the boundedness of Borel-measurable functional calculus.

Next we estimate $\|f(T) - f(S)\|$. We write $f(t) = \sum_i \alpha_i \chi_{A_i}(t)$ with $A_i \subset \mathbb{R}$ disjoint. Then, by definition of E_{A_i}, we get.

$$f(T) = \sum_i \alpha_i \chi_{A_i}(T) = \sum_i \alpha_i E_{A_i}.$$

Furthermore, according to Theorem 12.3.8 applied to S, we know that

$$f(S) = \int_{\mathbb{R}} f(\lambda) dE_\lambda = \int_{\mathbb{R}} \sum_i \alpha_i \chi_{A_i}(\lambda) dE_\lambda = \sum_i \alpha_i E_{A_i}. \tag{12.2}$$

Hence, $\|f(T) - f(S)\| = 0$.

Finally, we estimate $\|f(S)-S\|$. We use again (12.2) combined with the norm estimate from Theorem 12.3.6 and obtain

$$\|f(S) - S\| = \left\| \int (f(\lambda) - \lambda) dE_\lambda \right\| \leq \|f(\lambda) - \lambda\|_\infty \leq \delta$$

Hence, $\|S - T\| \leq \delta$ for $\delta > 0$ arbitrary small, i.e. $S = T$.

The properties of the mapping $f \to f(T) := \int_{\sigma(T)} \lambda dE_\lambda$ follow from Theorem 12.3.8, once the identity $S = T$ was proven. Finally, the identity for $\langle f(T)x, y\rangle$ is a matter of definition for $f = \chi_A$, and also for simple functions. Finally, taking a limit, it follows for every bounded Borel measurable f. $\square$

Let us summarize the structure of this section into the following diagram.

$$\begin{tikzcd}
\overset{T\in\mathscr{L}(H)}{\text{self-adjoint}} \arrow[r, "\text{Weierstraß Theorem}", "\text{density of polynomials}"'] \arrow[d, equal] & \text{Cont. F.C.(Thm. 12.2.4)} \arrow[r, "\text{Riesz}"] & \text{Borel FC (Thm. 12.2.11)} \arrow[d, "E_A:=\chi_A(T)"'] \\
\overset{T\in\mathscr{L}(H)}{\text{self-adjoint}} & \int f dE \arrow[l, "T:=\int \lambda dE_\lambda"', "\text{Thm. 12.3.8}"] & \text{Spectral measure} \arrow[l, "\text{Thm. 12.3.6}"']
\end{tikzcd}$$

12.4 Few applications of spectral theorem

Theorem 12.4.1. *Let $S \in \mathscr{L}(H)$ be self-adjoint, $g : \sigma(S) \to \mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$ be Borel-measurable and bounded. Then*

$$(f \circ g)(S) = f(g(S)).$$

Proof. First, note that $f \circ g$ is again Borel-measurable. Furthermore, $g(S)$ is self-adjoint and, therefore, $f(g(S))$ can be defined.

It is enough to consider $f = \chi_A$, the rest follows as usually. Then

$$\chi_A \circ g = \chi_{g^{-1}(A)}$$

and we have to show that $\chi_{g^{-1}(A)}(S) = \chi_A(g(S))$. Let F be the spectral measure of S and E the spectral measure of $g(S)$; then we have to show that $F_{g^{-1}(A)} = E_A$ for all Borel sets A, or, equivalently,

$$\langle F_{g^{-1}(A)}x, y\rangle = \langle E_A x, y\rangle$$

for all $x, y \in H$ and all Borel sets A.

Let $x, y \in H$ be fixed and let us consider the measures

$$\nu^1_{x,y} : A \to \langle F_{g^{-1}(A)}x, y\rangle,$$
$$\nu^2_{x,y} : A \to \langle E_A x, y\rangle,$$
$$\mu_{x,y} : A \to \langle F_A x, y\rangle.$$

Hence, $\nu^1_{x,y}(A) = \mu_{x,y}(g^{-1}(A))$. We use the transformation law for the (complex-valued) measures $\nu^1_{x,y}$ and $\mu_{x,y}$

$$\int h d\nu^1_{x,y} = \int (h \circ g) d\mu_{x,y}$$

in the form

$$\int h(\lambda)d\nu^1_{x,y}(\lambda) = \int (h\circ g)d\mu_{x,y} = \int h(g(\lambda))d\langle F_\lambda x, y\rangle$$

and obtain

$$\int \lambda^n d\nu^1_{x,y}(\lambda) = \int g(\lambda)^n d\langle F_\lambda x, y\rangle = \langle g(S)^n x, y\rangle = \int \lambda^n d\langle E_\lambda x, y\rangle = \int \lambda^n d\nu^2_{x,y}(\lambda).$$

The measures $\nu^1_{x,y}$ and $\nu^2_{x,y}$ coincide on polynomials, and due to Theorems of Riesz and Weierstraß, they coincide on $C(\sigma(g(S)))$ and also as measures. □

Remark 12.4.2. We recall the notion of the *push-forward measure*. Let μ be a measure on X_1 and $f : X_1 \to X_2$ a mapping. Then $f_*(\mu)(B) = \mu(f^{-1}(B))$ for $B \subset X_2$ is called the push-forward measure of μ under f. To develop the complete theory (which includes the transformation law we have used just now) one has to take care also about the corresponding σ-algebras.

Corollary 12.4.3. *Let $T \in \mathscr{L}(H)$ be positive and let $n \in \mathbb{N}$. Then there exists a unique positive $S \in \mathscr{L}(H)$ with $S^n = T$.*

Proof. The functions $f_n : t \to t^{1/n}$ are continuous, bounded and non-negative on $\sigma(T) \subset [0,\infty)$. We set $S := f_n(T)$. Then $S \geq 0$ and from $t^{1/n}\dots t^{1/n} = t$, it follows $S^n = T$. Let $g_n(t) = t^n$, then $(f_n \circ g_n)(t) = t$ for $t \in [0,\infty) \supset \sigma(S)$ and

$$S = (f_n \circ g_n)(S) = f_n(g_n(S)) = f_n(S^n) = f_n(T)$$

shows the uniqueness of S. □

Corollary 12.4.4. (Polar decomposition)
To an operator $T \in \mathscr{L}(H)$, we define $|T| = (T^*T)^{1/2}$. Furthermore, there is a partial isometry U with $T = U|T|$.

Proof. (Proof is the same as for compact operators).
We see that $\||T|x\|^2 = \langle x, T^*Tx\rangle = \langle Tx, Tx\rangle = \|Tx\|^2$. Therefore $U(|T|x) = Tx$ is an isometry of $\operatorname{ran}|T|$ to $\operatorname{ran} T$. This can be extended to $U : \overline{\operatorname{ran}|T|} \to \overline{\operatorname{ran} T}$. Finally, we put $U = 0$ on $(\operatorname{ran}|T|)^\perp = \ker|T| = \ker T$. □

12.5 Spectral theorem for normal operators

We shall proceed similarly to the case of compact normal operators (cf. Exercises).

Let $T \in \mathscr{L}(H)$ be a normal operator. Using

$$T = \underbrace{\frac{T+T^*}{2}}_{=:T_1} + i\underbrace{\frac{T-T^*}{2i}}_{=:T_2},$$

we can decompose $T = T_1 + iT_2$, where T_1 and T_2 are self-adjoint and commuting operators.

Let E be the spectral measure associated with T_1 and F the spectral measure associated with T_2. We shall show that $E_A F_B = F_B E_A$ for all sets $A, B \subset \mathbb{R}$. We show first that $E_A T_2 = T_2 E_A$. The same argument repeated for the pair (E_A, T_2) instead of (T_1, T_2) then gives the general statement.

We observe that $T_1^n T_2 = T_2 T_1^n$, i.e. $p(T_1)T_2 = T_2 p(T_1)$ for all polynomials and, by limiting arguments, also for all $\varphi \in C(\sigma(T_1))$. Finally, we consider $\varphi_n \to \chi_A$, which implies $\varphi_n(T_1)x \to E_A(x)$ for every $x \in H$. This finally gives

$$T_2(E_A x) = T_2(\lim_{n\to\infty} \varphi_n(T_1)x) = \lim_{n\to\infty} T_2(\varphi_n(T_1)x) = \lim_{n\to\infty} \varphi_n(T_1)(T_2 x) = E_A(T_2 x).$$

If $\Omega \subset \mathbb{C}$ is a product of two sets, i.e. $\Omega = \{x + iy : x \in A \text{ and } y \in B\}$ for some $A, B \subset \mathbb{R}$, we put $G(\Omega) := E_A F_B = F_B E_A$. Using standard arguments from measure theory, G can be extended to a σ-additive mapping on the Borel sets of $\mathbb{C}$. Let us note, that it has also a compact support.

From

$$T_1 = \int_{\mathbb{R}} \lambda dE_\lambda = \int_{\mathbb{C}} \operatorname{Re} z dG_z \quad \text{and} \quad T_2 = \int_{\mathbb{R}} \lambda dF_\lambda = \int_{\mathbb{C}} \operatorname{Im} z dG_z,$$

we get

$$T = T_1 + iT_2 = \int_{\mathbb{C}} z dG_z.$$

This finishes a sketch of the proof of the following theorem:

Theorem 12.5.1. (Spectral theorem for bounded normal operators)
Let $T \in \mathscr{L}(H)$ be normal. Then there exists a unique spectral measure G on $\mathbb{C}$ with compact support with

$$T = \int_{\sigma(T)} \lambda dG_\lambda.$$

The mapping $f \to f(T) = \int f(\lambda) dG_\lambda$ defines a Borel-measurable functional calculus and

$$\langle f(T)x, y\rangle = \int_{\sigma(T)} f(\lambda) d\langle G_\lambda x, y\rangle.$$

13 Spectral theory for unbounded operators

Also in this section, H denotes a Hilbert space over $\mathbb{C}$.

13.1 Definitions and Motivation

Many important operators on Hilbert spaces appearing for example in physics are not bounded. This is especially the case for many differential operators. Therefore, one is interested in analysis of operators defined on $\mathrm{dom}(T)$, which is only a subspace of H. Of course, if $T \in \mathscr{L}(H)$, then $\mathrm{dom}(T) = H$. On the other hand, if T is closed and defined on all H, it is bounded by the closed graph theorem. Hence, an unbounded closed operator is never defined on the whole H. Many (unbounded) operators are closed (or at least closable) and, therefore, the flexibility of $\mathrm{dom}(T) \neq H$ is essential for the theory.

Definition 13.1.1. i) The linear mapping $T : \mathrm{dom}(T) \subset H \to H$ is called an operator, if $\mathrm{dom}(T)$ is a linear subspace of H. We say, that T is *densely defined* if $\overline{\mathrm{dom}(T)} = H$.

ii) *Graph* of an operator T is the set $G(T) = \{(Tx, x) : x \in \mathrm{dom}(T)\} \subset H \times H$, where $H \times H = \{(x, y) : x, y \in H\}$ is a Hilbert space with the scalar product $\langle (u, v), (x, y)\rangle_{H\times H} = \langle u, x\rangle + \langle v, y\rangle$.

iii) We call T *closed operator* if $G(T)$ is closed (in $H \times H$).

iv) We say that T is *closable* if $\overline{G(T)}$ is a graph of some linear operator T_0. This operator is then called *closure* of T. We shall denote this by $T_0 = \overline{T}$.

v) An operator $S : \mathrm{dom}(S) \subset H \to H$ is called *extension* of T if $G(T) \subset G(S)$ or, equivalently, if $\mathrm{dom}(T) \subset \mathrm{dom}(S)$ and $Sx = Tx$ for all $x \in \mathrm{dom}(T)$. We shall denote this by $T \subset S$.

vi) $T = S$ if $T \subset S$ and $S \subset T$, i.e. if $\mathrm{dom}(T) = \mathrm{dom}(S)$ and $Sx = Tx$ for all $x \in \mathrm{dom}(T)$.

The domain of T, i.e. $\mathrm{dom}(T)$, is an essential part of the definition of every operator T. For example, if S and T are two (unbounded) operators on H, we set

$$\mathrm{dom}(S + T) = \mathrm{dom}(S) \cap \mathrm{dom}(T), \quad (S + T)(x) = Sx + Tx,$$
$$\mathrm{dom}(ST) = \{x \in \mathrm{dom}(T) : Tx \in \mathrm{dom}(S)\}, \quad (ST)(x) = S(T(x)).$$

Definition 13.1.2. An operator $T : \mathrm{dom}(T) \to H$ is called *symmetric* if

$$\langle Tx, y\rangle = \langle x, Ty\rangle \quad \text{for all } x, y \in \mathrm{dom}(T).$$

Theorem 13.1.3. (Hellinger-Toeplitz) *Let T be a symmetric operator defined on whole* $\mathrm{dom}(T) = H$. *Then T is bounded, i.e. $T \in \mathscr{L}(H)$.*

Proof. Let $x_n \to 0$ and $Tx_n \to z$. We show that $z = 0$ and apply the Closed graph theorem. But this follows easily from

$$\langle z, z\rangle = \Big\langle \lim_{n\to\infty} Tx_n, z \Big\rangle = \lim_{n\to\infty} \langle Tx_n, z\rangle = \lim_{n\to\infty} \langle x_n, Tz\rangle = 0.$$

Let us mention that in this case $T = T^*$.[8] □

[8] Where have we used that $\mathrm{dom}(T) = H$?

Next we define the adjoint of a densely defined operator T.

Definition 13.1.4. Let T be a densely defined operator. We define

$$\mathrm{dom}(T^*) = \{y \in H : x \to \langle Tx, y\rangle \text{ is continuous on } \mathrm{dom}(T)\}.$$

For $y \in \mathrm{dom}(T^*)$, the mapping $x \to \langle Tx, y\rangle$ may be extended to a continuous mapping on all H and (according to the Riesz theorem) represented as $x \to \langle x, z\rangle$, $z \in H$. We denote this z by T^*y. Due to the density of $\mathrm{dom}(T)$, z is unique. T^* is called the adjoint operator to T. If $T = T^*$, then T is called self-adjoint.

We have $\langle Tx, y\rangle = \langle x, T^*y\rangle$ for all $x \in \mathrm{dom}(T)$ and all $y \in \mathrm{dom}(T^*)$.

Example 13.1.5. Let $H = L_2(0,1)$ and define

$$\begin{aligned}\mathrm{dom}(T) &= C_0^\infty(0,1) = \{f \in C^\infty(0,1) : f \text{ has compact support}\},\\ Tf &= if'.\end{aligned}$$

Since $C_0^\infty(0,1)$ is dense in $L_2(0,1)$, T is densely defined. It is easy to check, that T is symmetric:

$$\langle Tf, g\rangle = \int_0^1 if'(t)\overline{g(t)}dt = -i\int_0^1 f(t)\overline{g'(t)}dt = \int_0^1 f(t)\overline{ig'(t)}dt = \langle f, Tg\rangle.$$

However, T is not self-adjoint - the same calculation goes through also for $f, g \in C_0^1(0,1)$, which means, that $f \to \langle Tf, g\rangle$ is continuous also for $g \in C_0^1(0,1) \subset \mathrm{dom}(T^*)$.

Therefore, there comes a question: what is the closure of T and the adjoint of T?

Remark 13.1.6. In the sequel, we shall encounter the space $AC([0,1])$ of *absolutely continuous* functions on $[0,1]$. Let us recall their definition and basic facts about them. A function $f : [0,1] \to \mathbb{R}$ (or $\mathbb{C}$) is called absolutely continuous, if for every $\varepsilon > 0$ there is a $\delta > 0$, such that

$$\sum_{j=1}^m |f(b_j) - f(a_j)| < \varepsilon$$

for every $0 \le a_1 < b_1 \le a_2 < b_2 \le \dots \le a_m < b_m \le 1$ with $\sum_j (b_j - a_j) < \delta$. Obviously, every Lipschitz function on $[0,1]$ is absolutely continuous. Using the notion of *total variation*, one shows that every absolutely continuous function is a difference of two nondecreasing absolutely continuous functions. This, combined with the fact that nondecreasing functions are differentiable almost everywhere, then gives that absolutely continuous functions are also differentiable almost everywhere, i.e. it makes sense to ask if their derivative is an element of $L_1([0,1])$.

We shall also need the following properties:

(a) **(partial integration):** If $f, g \in AC([0,1])$, then $\int_0^1 fg' = [f(x)g(x)]_{x=0}^{x=1} - \int_0^1 f'g$.

(b) $f \in AC([0,1])$ if, and only if, f is differentiable almost everywhere, $f' \in L_1([0,1])$ and $f(x) = f(0) + \int_0^x f'(t)dt$.

Example 13.1.7. Let $H = L_2([0,1])$. We define

$$\mathrm{dom}(T) = \{f \in AC([0,1]) : f' \in L_2([0,1]) \text{ and } f(0) = f(1) = 0\}.$$

The set $\mathrm{dom}(T)$ is dense in H and T is therefore densely defined.

We shall show that the operator $Tf = f'$ for $f \in \mathrm{dom}(T)$ is closed and unbounded on H. Unboundedness follows by considering the sequence $f_n(x) = \frac{1}{n}\sin(\pi n x)$ and calculating the norms $\|f_n\|$ and $\|f_n'\|$.

Let us now suppose that $f_n \in \mathrm{dom}(T)$, $f_n \to f$ in H and $Tf_n \to g$ in H. We know that (absolute continuity) $f_n(x) = \int_0^x f_n'$ and we put $h(x) = \int_0^x g$. Then

$$|f_n(x) - h(x)| \le \int_0^1 |f_n' - g| \le \left(\int_0^1 |f_n' - g|^2\right)^{1/2} \to 0$$

as $n \to \infty$. Therefore, $f_n \rightrightarrows h$ on $[0,1]$ and hence also $f_n \to h$ in H, and finally $f = h$ a.e. We conclude that $f_n \rightrightarrows f$, $f \in AC([0,1])$ and $f(0) = f(1) = 0$. We therefore have $f \in \mathrm{dom}(T)$ and $Tf = g$, i.e. T is closed.

Next, we look for T^*. We shall show that

$$\mathrm{dom}(T^*) = \{h \in AC([0,1]) : h' \in L_2([0,1])\} \quad \text{and} \quad T^*h = -h'.$$

Let us suppose first, that $h \in AC([0,1])$ and $h' \in L_2([0,1])$. Then

$$\langle Tf, h\rangle = \int_0^1 f'\bar{h} = [f\bar{h}]_0^1 - \int_0^1 f\bar{h}' = -\int_0^1 f\bar{h}' = \langle f, -h'\rangle,$$

hence $f \to \langle Tf, h\rangle$ is continuous and $h \in \mathrm{dom}(T^*)$.

Let on the other hand $g \in \mathrm{dom}(T^*)$ and $h = T^*g$. For $f \in \mathrm{dom}(T)$, we get

$$\begin{aligned}\int_0^1 f'\bar{g} = \langle Tf, g\rangle = \langle f, T^*g\rangle = \langle f, h\rangle = \int_0^1 f\bar{h} &= \left[f(x)\int_0^x \bar{h}\right]_{x=0}^{x=1} - \int_0^1 f'(x)\left(\int_0^x \bar{h}\right)dx \\ &= -\int_0^1 f'(x)\left(\int_0^x \bar{h}\right)dx.\end{aligned}$$

We see, that

$$\left\langle f'(\cdot), g(x) + \int_0^x h\right\rangle = 0$$

for all $f \in \mathrm{dom}(T)$. In other words, the function $x \to g(x) + \int_0^x h$ is orthogonal to f' for all $f \in \mathrm{dom}(T)$. Therefore, this function is constant and (since $\int_0^x h$ is absolutely continuous) g is also in $AC([0,1])$ and $g' = -h \in H$. Furthermore, this shows that $T^*g = h = -g'$.

Example 13.1.8. If $H = L_2([0,1])$,

$$\mathrm{dom}(T) = \{f \in AC([0,1]) : f' \in L_2([0,1]), f(0) = f(1) = 0\} \text{ and } Tf = if'$$

we obtain in a similar way

$$\mathrm{dom}(T^*) = \{h \in AC([0,1]) : h' \in L_2([0,1])\} \text{ and } T^*h = ih'.$$

Therefore, $T \subset T^*$ and T is symmetric but not self-adjoint.

Example 13.1.9. Finally, if $H = L_2([0,1])$,

$$\mathrm{dom}(T) = \{f \in AC([0,1]) : f' \in L_2([0,1]), f(0) = f(1)\} \text{ and } Tf = if',$$

we get

$$\mathrm{dom}(T^*) = \mathrm{dom}(T) \quad \text{and} \quad T = T^*.$$

If T is densely defined and symmetric, then $T \subset T^*$ and $\mathrm{dom}(T) \subset \mathrm{dom}(T^*) \subset H$, and T^{**} may be defined.

Theorem 13.1.10. *Let $T : \mathrm{dom}(T) \to H$ be densely defined. Then*

i) T^* is closed.

ii) If T^* is densely defined, then $T \subset T^{**}$.

iii) If T^* is densely defined, then T^{**} is the closure of T.

Proof. (i): We have to show that $y_n \in \mathrm{dom}(T^*), y_n \to y \in H$ and $T^*y_n \to z$ implies $y \in \mathrm{dom}(T^*)$ and $z = T^*y$. We have

$$\langle Tx, y\rangle = \lim_{n\to\infty} \langle Tx, y_n\rangle = \lim_{n\to\infty} \langle x, T^*y_n\rangle = \langle x, z\rangle \quad \text{for every} \quad x \in \mathrm{dom}(T).$$

Hence $x \to \langle Tx, y\rangle$ is continuous, $y \in \mathrm{dom}(T^*)$ and $T^*y = z$.
(ii): If $x \in \mathrm{dom}(T)$ and $y \in \mathrm{dom}(T^*)$, then $\langle Tx, y\rangle = \langle x, T^*y\rangle$. Hence $y \to \langle x, T^*y\rangle$ is continuous and $x \in \mathrm{dom}(T^{**})$. Furthermore,

$$\langle x, T^*y\rangle = \langle T^{**}x, y\rangle \text{ for all } y \in \mathrm{dom}(T^*).$$

We conclude that $Tx = T^{**}x$ for all $x \in \mathrm{dom}(T)$, i.e. $T \subset T^{**}$.
(iii): We show that

$$\overline{G(T)} = G(T^{**}).$$

The inclusion "$\subset$" follows from (i) and (ii). Let $(v, u) \in G(T)^\perp$. Then $\langle x, u\rangle + \langle Tx, v\rangle = 0$ for all $x \in \mathrm{dom}(T)$. Then $v \in \mathrm{dom}(T^*)$ and $T^*v = -u$.
For $(T^{**}z, z) \in G(T^{**})$, we have

$$\langle (T^{**}z, z), (v, u)\rangle = \langle z, u\rangle + \langle T^{**}z, v\rangle = \langle z, u\rangle + \langle z, T^*v\rangle = \langle z, u + T^*v\rangle = 0,$$

hence $(v, u) \in G(T^{**})^\perp$ and the second inclusion follows.

□

Corollary 13.1.11. *Let $T : \mathrm{dom}(T) \to H$ be densely defined.*

i) T is symmetric if, and only if, $T \subset T^*$. Then we have $T \subset T^{**} \subset T^* = T^{***}$ and also T^{**} is symmetric.

ii) T is closed and symmetric if, and only if, $T = T^{**} \subset T^*$.

iii) T is self-adjoint if, and only if, $T = T^* = T^{**}$.

Example 13.1.12. We shall present an example of an operator with a non-densely defined adjoint. Let $H = L_2(-1, 1)$ and let

$$\mathrm{dom}(T) = \{f \in C^\infty(-1, 1) \cap L_2(-1, 1) : |f^{(j)}(0)| \le C_f 2^{-j} j! \text{ for all } j \ge 0\},$$
$$(Tf)(x) = \sum_{j=0}^{\infty} \frac{f^{(j)}(0)}{j!} x^j, \quad -1 < x < 1.$$

Roughly speaking, T sends f to its Taylor series, and the domain contains only functions for which this series converges uniformly and absolutely. We shall show that T is a densely defined operator with $\mathrm{dom}(T^*) = \{0\}$.

We proceed in the following way

i) If T is an arbitrary linear operator and $\ker(T)$ is dense, then $\mathrm{dom}(T^*) = \ker(T^*)$:
Let $y \in \mathrm{dom}(T^*)$. Then $x \to \langle Tx, y\rangle$ is a continuous mapping on $x \in \mathrm{dom}(T)$. As $\ker(T) \subset \mathrm{dom}(T) \subset H$ is a dense subset of H, this mapping is continuous and equal to 0 on a dense subset of H. Hence, $T^*y = 0$.

ii) If $\mathrm{ran}(T)$ is also dense, then $\mathrm{dom}(T^*) = \{0\}$:
Let again $y \in \mathrm{dom}(T^*)$. Then (by previous point) $T^*y = 0$ and the mapping $x \to \langle Tx, y\rangle$ is equal to zero for all $x \in \mathrm{dom}(T)$. This means, that $\langle \tilde{x}, y\rangle = 0$ for $\tilde{x}$ from a dense subset of H, hence $y = 0$.

iii) Finally, $\ker(T)$ and $\mathrm{ran}(T)$ are dense for the operator above:
Obviously, $\ker T \supset C_0^\infty((-1,1) \setminus \{0\})$, i.e. $\ker T$ contains all functions vanishing on some neighborhood of zero and this set is dense in H. Furthermore, the range of T includes all polynomials, as the constant C_f might be arbitrary large, hence also $\mathrm{ran}(T)$ is dense in H.

13.2 Spectral properties of unbounded operators

Definition 13.2.1. Let $T : \mathrm{dom}(T) \subset H \to H$ be densely defined.

i) The set

$$\varrho(T) := \{\lambda \in \mathbb{C} : \lambda I - T : \mathrm{dom}(T) \to H \text{ has a bounded inverse in } H\}^{9}$$

is called the *resolvent set* of T.

ii) The mapping $R : \varrho(T) \to \mathscr{L}(H), R_\lambda = R(\lambda) := (\lambda I - T)^{-1}$ is called the *resolvent mapping.*

iii) $\sigma(T) := \mathbb{C} \setminus \varrho(T)$ is called the *spectrum* of T.

Example 13.2.2. Let $\Omega \subset \mathbb{R}^n$ be measurable and let $f : \Omega \to \mathbb{R}$ a measurable function. We denote by

$$\begin{aligned} \mathrm{dom}(M_f) &= \{g \in L_2(\Omega) : fg \in L_2(\Omega)\}, \\ M_f(g)(t) &= f(t)g(t), \quad t \in \Omega \end{aligned}$$

the pointwise multiplication operator on the Hilbert space $H = L_2(\Omega)$.

Then it holds

i) M_f is densely defined:
Let $A_n := \{x \in \Omega : |f(x)| \le n\}$. Then $g\chi_{A_n} \in \mathrm{dom}(M_f)$ for all $n \in \mathbb{N}$ and all $g \in H$. And $g\chi_{A_n} \to g$ in H finishes the argument.

ii) M_f is self-adjoint (and, therefore, M_f is also closed):
Let $h \in \mathrm{dom}(M_f)$. Then $g \to \langle M_f g, h\rangle = \int_\Omega f(t)g(t)\overline{h(t)}dt = \langle g, M_f h\rangle$ is continuous, $h \in \mathrm{dom}(M_f^*)$ and $M_f^* h = M_f h$. If, on the other hand, $h \in \mathrm{dom}(M_f^*)$, then the mapping above is continuous, and $fh \in H$ follows.

[9] This means, that there is an operator $S \in \mathscr{L}(H)$, such that $(T - \lambda I)S = I$ and $S(T - \lambda I)$ is identity on $\mathrm{dom}(T)$.

iii) $\sigma_p(M_f) = \{z \in \mathbb{R} : f^{-1}(z)$ has positive measure$\}$ are the eigenvalues (the so-called *point spectrum*) of M_f and every eigenvalue of M_f has infinite multiplicity:
If the set $A_z := f^{-1}(z)$ has positive measure, then $\{g \in H : \operatorname{supp} g \subset A_z\}$ is an infinite-dimensional subspace of eigenvectors. If, on the other hand, $M_f g(t) = f(t)g(t) = zg(t)$ for almost every $t \in (-1,1)$, we get $f(t) = z$ almost everywhere on the support of g and the claim follows.

iv) $z \in \mathbb{C}$ is in $\sigma(M_f)$ if, and only if, for every $\varepsilon > 0$, the set $\{x \in \Omega : |f(x) - z| < \varepsilon\} = f^{-1}(\{w \in \mathbb{C} : |w - z| < \varepsilon\})$ has positive measure:
Let $z \in \mathbb{C}$ be such that $A_n := f^{-1}(\{w \in \mathbb{C} : |w - z| < 1/n\})$ has positive measure for every $n \in \mathbb{N}$ and consider $g_n \neq 0$ with $\operatorname{supp} g_n \subset A_n$. Then $g_n \in \operatorname{dom}(M_f)$ and

$$\|(M_f - zI)g_n\|^2 = \int_{-1}^{1} |(f(t) - z)g_n(t)|^2 dt = \int_{A_n} |(f(t) - z)g_n(t)|^2 dt \leq \frac{\|g_n\|^2}{n^2}$$

and $M_f - zI$ cannot be boundedly invertible, i.e. $z \in \sigma(M_f)$.
Let on the other hand $z \in \mathbb{C}$ and $\varepsilon > 0$ be such that $f^{-1}(\{w \in \mathbb{C} : |w - z| < \varepsilon\})$ has measure zero. Then $M_f - zI$ can be inverted by $M_{\frac{1}{f(t)-z}}$.

If T is not closed, then $(T - \lambda I)$ is not closed for every $\lambda \in \mathbb{C}$, hence $G(T - \lambda I)$ is not closed. On the other hand, if $(T - \lambda I)^{-1}$ would be bounded (i.e. in $\mathscr{L}(H)$), then $G((T - \lambda I)^{-1})$ would be closed and, therefore, also $G(T - \lambda I)$ would be closed.

This can be summarized by saying that $\sigma(T) = \mathbb{C}$ whenever T is not closed.

Theorem 13.2.3. *Let $T : \operatorname{dom}(T) \to H$ be densely defined. Then*

i) $\varrho(T)$ is open.

ii) The resolvent mapping is analytic and the resolvent identity

$$R_\lambda - R_\mu = (\mu - \lambda) R_\lambda R_\mu$$

holds.

iii) $\sigma(T)$ is closed.

Warning: $\sigma(T)$ can be unbounded, or empty.

Proof. (Same as for bounded operators!)

i) Let $\lambda_0 \in \varrho(T)$ and $|\lambda - \lambda_0| < \|(\lambda_0 I - T)^{-1}\|^{-1}$. Then

$$\lambda I - T = (\lambda_0 I - T) + (\lambda - \lambda_0) I = (\lambda_0 I - T) \underbrace{[I - (\lambda - \lambda_0)(\lambda_0 I - T)^{-1}]}_{[\dots]^{-1} = \sum_{n=0}^{\infty}((\lambda-\lambda_0)(\lambda_0 I - T)^{-1})^n}.$$

Hence, $\lambda I - T$ is invertible.

ii) Formal calculation gives immediately

$$(\lambda I - T)(\mu I - T)[(\lambda I - T)^{-1} - (\mu I - T)^{-1}] = (\mu - \lambda) I$$

and the similar identity for multiplication from the right side. So, only the (easy) inspection of domains is missing.

iii) follows from (i).

□

Theorem 13.2.4. *Let T be a densely defined operator. Then* $\ker(T^*) = \operatorname{ran}(T)^\perp$, *where*

$$\ker(T^*) = \{x \in \operatorname{dom}(T^*) : T^*x = 0\},$$
$$\operatorname{ran}(T) = \{Tx : x \in \operatorname{dom}(T)\}.$$

Proof. $y \in \ker(T^*)$ if, and only if, $\langle Tx, y\rangle = 0$ for all $x \in \operatorname{dom}(T)$ and if, and only if $y \perp \operatorname{ran}(T)$. □

Theorem 13.2.5. *Let T be a densely defined symmetric operator, $z \in \mathbb{C} \setminus \mathbb{R}$. Then the following statements are equivalent:*

i) T is self-adjoint.

ii) T is closed and $\ker(T^* - zI) = \ker(T^* - \bar{z}I) = \{0\}$.

iii) $\operatorname{ran}(T - zI) = \operatorname{ran}(T - \bar{z}I) = H$.

Proof. We shall use the fact, that if T is densely defined and $z \in \mathbb{C}$, then $(T - zI)^* = T^* - \bar{z}I$, cf. Exercises.

(i) implies (ii): $T = T^*$ and T^* is closed, therefore also T is closed. Let $x \in \ker(T^* - zI) = \ker(T - zI)$ (as $T = T^*$). Then

$$z\langle x, x\rangle = \langle zx, x\rangle = \langle Tx, x\rangle = \langle x, T^*x\rangle = \langle x, zx\rangle = \bar{z}\langle x, x\rangle,$$

i.e. $x = 0$. The same holds for $\ker(T^* - \bar{z}I)$.
(ii) implies (iii): By Theorem 13.2.4, we have $\operatorname{ran}(T - zI)^\perp = \ker(T^* - \bar{z}I) = \{0\}$. So, it is enough to show, that $\operatorname{ran}(T - zI)$ is closed.
Let $y_n \in \operatorname{ran}(T - zI)$, i.e. $y_n = (T - zI)x_n$ for some $x_n \in \operatorname{dom}(T)$ and we suppose that $y_n \to y$. Let $z = a + ib$ with $b \neq 0$ and let $u \in \operatorname{dom}(T)$. Then we have

$$\|(T - zI)u\|^2 = \langle (T - a - ib)u, (T - a - ib)u\rangle = \|(T - aI)u\|^2 + b^2\|u\|^2 \geq b^2\|u\|^2,$$

as $\langle -ibu, (T - aI)u\rangle + \langle (T - aI)u, -ibu\rangle = 0$ as $u \in \operatorname{dom}(T) \subset \operatorname{dom}(T^*)$. Hence $\|u\| \leq 1/|b| \cdot \|(T - zI)u\|$.
We apply this inequality for $u = x_m - x_n$ and observe, that $(x_n)_{n\in\mathbb{N}}$ is a Cauchy-sequence and, therefore, $x = \lim_{n\to\infty} x_n$ exists. Then $Tx_n = y_n + zx_n$ converges also, to $y + zx$. As T is closed, we get $x \in \operatorname{dom}(T)$, $Tx = y + zx$, or $y = (T - zI)x \in \operatorname{ran}(T - zI)$.
(iii) implies (i): Let $x \in \operatorname{dom}(T^*)$. By assumption (iii), we can find $y \in \operatorname{dom}(T)$, such that $(T - zI)y = (T^* - zI)x = (T^* - zI)y$ (as $T \subset T^*$). Hence, $x - y \in \ker(T^* - zI)$, which (see Theorem 13.2.4) is equal to $\operatorname{ran}(T - \bar{z}I)^\perp = \{0\}$. Therefore, $x = y \in \operatorname{dom}(T)$, $\operatorname{dom}(T) \subset \operatorname{dom}(T^*)$ and $T = T^*$. □

Corollary 13.2.6. *Let T be self-adjoint. Then $\sigma(T) \subset \mathbb{R}$.*

Proof. Let $z \in \mathbb{C}\backslash\mathbb{R}$. By Theorem 13.2.5 we have $\ker(T - zI) = 0$ and $\operatorname{ran}(T - zI) = H$. □

Example 13.2.7. We present an example, where the change of the domain of the operator has a big impact on its spectrum. This demonstrates once more, how important is the right choice of the domain.

We consider again the operator $f \to if'$ on $L_2(0, 1)$, on the following domains:

$$\operatorname{dom}(S) = \{f \in AC([0, 1]) : f' \in L_2(0, 1)\},$$
$$\operatorname{dom}(T) = \{f \in \operatorname{dom}(S) : f(0) = 0\}.$$

It is quite easy to prove (similar to previous arguments) that both S and T are closed.

We show that

$$\sigma(S) = \mathbb{C}, \quad \sigma(T) = \varnothing.$$

The claim on $\sigma(S)$ is very easy to confirm. Just notice that $e_z(x) = e^{-izx} \in \operatorname{dom}(S)$ for all $z \in \mathbb{C}$ and $(S - z)e_z = 0$.

To find $\sigma(T)$, fix $z \in \mathbb{C}$ and let

$$(R_z f)(x) = -ie^{-izx} \int_0^x e^{izt} f(t)dt.$$

This is well defined for all $f \in L_2(0,1)$ and $(R_z f)(x)$ is absolutely continuous function of $x \in [0,1]$. Hence $R_z f \in L_2(0,1)$ as well. An easy calculation shows that

$$(R_z f)'(x) = -iz(R_z f)(x) - if(x).$$

Hence $(R_z f)' \in L_2(0,1)$ and $(R_z f)(0) = 0$, i.e. $R_z f \in \operatorname{dom}(T)$.

Moreover $(T - zI)R_z = I$, so $\operatorname{ran}(T - zI) = H$. Similar arguments (integration by parts) show that $R_z(T - zI)f = f$ for all $f \in \operatorname{dom}(T)$. Together, $z \in \varrho(T)$ and $\sigma(T) = \varnothing$.

13.3 Spectral theorem for unbounded operators

The main aim of this section is to prove the following theorem.

Theorem 13.3.1. (Spectral theorem for unbounded operators) *Let $T : \operatorname{dom}(T) \to H$ be a self-adjoint operator. Then there exists a unique spectral measure E on Borel sets of $\sigma(T)$, such that*

$$T = \int_{\sigma(T)} \lambda dE_\lambda,$$

i.e.

$$\langle Tx, y\rangle = \int_{\sigma(T)} \lambda d\langle E_\lambda x, y\rangle \quad \textit{for all } x \in \operatorname{dom}(T) \textit{ and all } y \in H.$$

Remark 13.3.2. i) Formally, it looks very similar to the case of bounded operators.

ii) $\sigma(T)$ might be an unbounded subset of $\mathbb{R}$, therefore $\int_{\sigma(T)} \lambda dE_\lambda$ integrates an unbounded function over an unbounded subset of $\mathbb{R}$.

iii) Let E be a spectral measure on $\mathbb{R}$ and f be an unbounded (but everywhere finite) Borel-measurable function. We let

$$D_f := \{x \in H : \int_{\mathbb{R}} |f(t)|^2 d\langle E_t x, x\rangle < \infty\}. \tag{13.1}$$

Recall, that $A \to \langle E_A x, x\rangle$ is a non-negative measure on $\mathbb{R}$.

iv) Furthermore, one has to show that

$$\langle T_f x, y\rangle = \int_{\mathbb{R}} f(t) d\langle E_t x, y\rangle$$

makes sense and that $\operatorname{dom}(T_f) = D_f$.

Lemma 13.3.3. *D_f is a dense subset of H. For all $x \in D_f$ and $y \in H$ we have*

$$\int_{\mathbb{R}} |f(t)| d|\langle E_t x, y\rangle| \le \|y\| \cdot \left(\int_{\mathbb{R}} |f(t)|^2 d\langle E_t x, x\rangle\right)^{1/2}. \tag{13.2}$$

Here, the integration with respect to $d|\langle E_t x, y\rangle|$ means the integration with respect to the variation of the measure $\mu_{x,y} : A \to \langle E_A x, y\rangle$, i.e. $|\mu_{x,y}|$.

Proof. (Sketch)
First, we show that D_f is a linear set. From

$$\langle E_A(x+y), x+y\rangle = \|E_A(x+y)\|^2 \le (\|E_A x\| + \|E_A y\|)^2 \le 2(\|E_A x\|^2 + \|E_A y\|^2)$$

we deduce that $\mu_{x+y,x+y} \le 2\mu_{x,x} + 2\mu_{y,y}$, where again $\mu_{x,x}(A) = \langle E_A x, x\rangle$. It follows, that $x, y \in D_f$ implies $x + y \in D_f$. Similar calculation shows also that $cx \in D_f$ for $c \in \mathbb{C}$.

Second, we show that D_f is dense. Let $\omega_n = \{t \in \mathbb{R} : |f(t)| < n\} \subset \mathbb{R}$ and let $y \in \operatorname{ran}(E_{\omega_n})$. Then

$$\langle E_{\omega_n^c} y, y\rangle = 0 \text{ and } \int_{\mathbb{R}} |f(t)|^2 d\langle E_t y, y\rangle = \int_{\omega_n} |f(t)|^2 d\langle E_t y, y\rangle \le n^2 \|y\|^2 < \infty.$$

Therefore, $y \in D_f$. Furthermore, $E_{\omega_n} x \to x$ due to $\|x - E_{\omega_n} x\|^2 = \|E_{\omega_n^c} x\|^2 = \mu_{x,x}(\omega_n^c) \to 0$.

Finally, let us prove the inequality (13.2). First, we observe that (for $x \ne 0$)

$$|\mu_{x,y}(A)| = |\langle E_A x, y\rangle| = |\langle E_A x, E_A y\rangle| \le \|E_A x\| \cdot \|E_A y\|.$$

As

$$|\mu_{x,y}|(A) = \sup\left\{ \sum_{k=1}^n |\mu_{x,y}(A_k)| : \bigcup_{k=1}^n A_k = A, A_1, A_2, \dots, A_k \text{ disjoint} \right\},$$

we get

$$\begin{aligned}
|\mu_{x,y}|(A) &\le \sup\left\{ \sum_{k=1}^n \|E_{A_k} x\| \cdot \|E_{A_k} y\| : \bigcup_{k=1}^n A_k = A, A_1, A_2, \dots, A_k \text{ disjoint} \right\} \\
&\le \sup\left\{ \left(\sum_{k=1}^n \|E_{A_k} x\|^2\right)^{1/2} \cdot \left(\sum_{k=1}^n \|E_{A_k} y\|^2\right)^{1/2} : \bigcup_{k=1}^n A_k = A, A_1, A_2, \dots, A_k \text{ disjoint} \right\} \\
&\le \|E_A x\| \cdot \|E_A y\|.
\end{aligned}$$

For f a simple function $f = \sum_{i=1}^n \alpha_i \chi_{A_i}$ with A_i's disjoint, the inequality (13.2) now follows by the Cauchy-Schwartz inequality:

$$\begin{aligned}
\int_{\mathbb{R}} |f(t)| d|\langle E_t x, y\rangle| &= \sum_{i=1}^n |\alpha_i| \cdot |\mu_{x,y}|(A_i) \le \sum_{i=1}^n |\alpha_i| \cdot \|E_{A_i} x\| \cdot \|E_{A_i} y\| \\
&\le \left(\sum_{i=1}^n |\alpha_i|^2 \|E_{A_i} x\|^2\right)^{1/2} \cdot \left(\sum_{i=1}^n \|E_{A_i} y\|^2\right)^{1/2} \\
&\le \|y\| \cdot \left(\int_{\mathbb{R}} |f(t)|^2 d\langle E_t x, x\rangle\right)^{1/2}.
\end{aligned}$$

For a general function f, we consider a sequence of simple functions $|f_n| \nearrow |f|$ and apply Levi's Theorem (sometimes also called Lebesgue monotone convergence theorem). □

Theorem 13.3.4. *Let E be a spectral measure on $\mathbb{R}$ and let f be a Borel-measurable function on $\mathbb{R}$. Let D_f be defined by (13.1).*

i) There exists a unique linear operator T_f on H with $\operatorname{dom}(T_f) = D_f$ *and*

$$\langle T_f x, y\rangle = \int_{\mathbb{R}} f(t)d\langle E_t x, y\rangle \quad \text{for all } x \in D_f \text{ and all } y \in H.$$

ii) For all $x \in D_f$

$$\|T_f x\|^2 = \int_{\mathbb{R}} |f(t)|^2 d\langle E_t x, x\rangle.$$

iii) $T_f \circ T_g \subset T_{fg}$ and $\operatorname{dom}(T_f \circ T_g) = D_g \cap D_{fg}$.

iv) $T_f^* = T_{\bar{f}}$ and T_f is closed.

Proof. (Sketch) Uniqueness is clear, as T_f is determined by (i). To show the existence, one considers the map

$$y \to \int_{\mathbb{R}} f(t)d\langle E_t x, y\rangle, \quad x \in D_f.$$

This is a bounded (semi-)linear mapping on H and may be therefore represented by a $z \in H$ (which we call $T_f x := z$) such that

$$\langle z, y\rangle = \int_{\mathbb{R}} f(t)d\langle E_t x, y\rangle \quad \text{and} \quad \|z\| = \left\| y \to \int_{\mathbb{R}} f(t)d\langle E_t x, y\rangle \right\|_{H'} \le \left(\int_{\mathbb{R}} |f(t)|^2 d\langle E_t x, x\rangle \right)^{1/2}.$$

The linearity of T_f follows by linearity of $x \to d\langle E_t x, y\rangle$.
One part ("$\le$") of the identity for $\|T_f x\|^2$ follows by Lemma 13.3.3. The other inequality follows for f_n bounded and detailed study of the inequality in Lemma 13.3.3. The rest (as well as (iii) and (iv)) follows by truncation and limits. □

Theorem 13.3.5. *Let $f : \mathbb{R} \to \mathbb{C}$ be measurable. Then* $\operatorname{dom}(T_f) = H$ *if, and only if, f is essentially bounded w.r.t. to E.*

Proof. Let $x \in H$, then $\langle E_{\mathbb{R}} x, x\rangle \le \|x\|^2$. If f is bounded, we integrate in (13.1) a bounded function over a finite measure space, hence $x \in D_f$.
If, on the other hand, $\operatorname{dom}(T_f) = D_f = H$, then $T_f \in \mathscr{L}(H)$ by Closed Graph Theorem. Furthermore, if $\omega_n := \{t \in \mathbb{R} : |f(t)| \ge n\}$, then Theorem 13.3.4 (ii) implies that $\|T_f x\| \ge n\|x\|$ for $x \in \operatorname{ran}(E_{\omega_n})$. Therefore $E_{\omega_n} = 0$ for n large. □

Proof. (Sketch of the proof of Theorem 13.3.1).
Let $R := (T - iI)^{-1} \in \mathscr{L}(H)$. Then $RR^* = R^*R$, i.e. R is normal (and $R^* = (T + iI)^{-1} \in \mathscr{L}(H)$). According to the spectral theorem for bounded normal operators,

$$R = (T - iI)^{-1} = \int_{\sigma(R)} z dF_z,$$

for some spectral measure F on $\mathbb{C}$ (R is normal, not self-adjoint).

Let $\varphi(t) = 1/(t - i)$. We want to "change the variables" from z to t. We put $E(M) := F(\varphi(M))$. The fact that $\varphi(i)$ is not defined is not really disturbing, because we aim for $M \subset \mathbb{R}$ anyway.

This is a new spectral measure and $\int z dF_z = \int \frac{1}{t-i} dE_t$. Then one has to show that

$$T - iI = R^{-1} = \int (t - i) dE_t,$$

which implies $T = \int t dE_t$. One has to show also that E is supported on $\sigma(T)$. Finally, uniqueness follows from the fact, that it is also possible to recover F from E and F is unique, therefore also E is unique. □

14 Distributions and Fourier transform

14.1 The space $\mathscr{S}(\mathbb{R}^n)$ and the Fourier transform on $\mathscr{S}(\mathbb{R}^n)$

We shall use the usual notation from vector analysis, i.e. let $n \in \mathbb{N}$, $\alpha \in \mathbb{N}_0^n$ be a multiindex and $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, then we write

$$\begin{aligned}
x^\alpha &:= x_1^{\alpha_1} \dots x_n^{\alpha_n}, \\
|\alpha| &:= \alpha_1 + \dots + \alpha_n, \\
D^\alpha &:= \partial_1^{\alpha_1} \dots \partial_n^{\alpha_n} = \frac{\partial^{|\alpha|}}{\partial_{x_1}^{\alpha_1} \dots \partial_{x_n}^{\alpha_n}}, \\
|x| &:= \sqrt{x_1^2 + \dots + x_n^2}, \\
\alpha! &:= \alpha_1! \dots \alpha_n!.
\end{aligned}$$

Definition 14.1.1. Let $n \in \mathbb{N}$.

i) We put
$$\mathscr{S}(\mathbb{R}^n) = \{\varphi \in C^\infty(\mathbb{R}^n) : \|\varphi\|_{(k,l)} < \infty \text{ for all } k, l \in \mathbb{N}_0\},$$
where
$$\|\varphi\|_{(k,l)} = \sup_{x \in \mathbb{R}^n} (1 + |x|^2)^{k/2} \sum_{|\alpha| \le l} |D^\alpha \varphi(x)|.$$

ii) A sequence $(\varphi_j)_{j \in \mathbb{N}} \subset \mathscr{S}(\mathbb{R}^n)$ is said to converge in $\mathscr{S}(\mathbb{R}^n)$ to $\varphi \in \mathscr{S}(\mathbb{R}^n)$ if, and only if,
$$\lim_{j \to \infty} \|\varphi_j - \varphi\|_{(k,l)} = 0 \quad \text{for all } k, l \in \mathbb{N}_0.$$
This will be written as $\varphi_j \stackrel{\mathscr{S}}{\to} \varphi$.

Remark 14.1.2. The space $\mathscr{S}(\mathbb{R}^n)$ is a vector space (in the algebraic sense), but it is not a Banach space (or normed space). It can be shown, that it is actually impossible to introduce a norm $\mathscr{S}(\mathbb{R}^n)$, such that the convergence in this norm would be equivalent to convergence, which we have just described. Nevertheless, it *is* possible to equip $\mathscr{S}(\mathbb{R}^n)$ with a topology, such that the convergence in this topology and $\mathscr{S}$-convergence are equivalent. Together with this topology, $\mathscr{S}(\mathbb{R}^n)$ becomes a *topological vector space*. We refer to the Book of Rudin [4], which contains a lot of details on such spaces. On the other hand, the space is metrizable, cf. Exercises.

Theorem 14.1.3. *i) Let $(\varphi_k)_{k \in \mathbb{N}} \subset \mathscr{S}(\mathbb{R}^n)$ with $\varphi_k \stackrel{\mathscr{S}}{\to} \varphi \in \mathscr{S}(\mathbb{R}^n)$. Then $\varphi_k \to \varphi$ also in $L_p(\mathbb{R}^n)$ for every $0 < p \le \infty$.*

ii) Let $(\varphi_k)_{k \in \mathbb{N}} \subset \mathscr{S}(\mathbb{R}^n)$ with $\varphi_k \stackrel{\mathscr{S}}{\to} \varphi \in \mathscr{S}(\mathbb{R}^n)$. Then $D^\alpha \varphi_k \stackrel{\mathscr{S}}{\to} D^\alpha \varphi$ for every $\alpha \in \mathbb{N}_0^n$.

iii) Let $(\varphi_k)_{k \in \mathbb{N}} \subset \mathscr{S}(\mathbb{R}^n)$ with $\varphi_k \stackrel{\mathscr{S}}{\to} \varphi \in \mathscr{S}(\mathbb{R}^n)$. Then $x^\alpha \varphi_k \stackrel{\mathscr{S}}{\to} x^\alpha \varphi$ for every $\alpha \in \mathbb{N}_0^n$.

iv) Let $\varphi \in \mathscr{S}(\mathbb{R}^n)$ and let $\tau_h \varphi(x) = \varphi(x - h)$. Then $\tau_h \varphi \stackrel{\mathscr{S}}{\to} \varphi$ for $h \to 0$.

v) Let $\varphi, \psi \in \mathscr{S}(\mathbb{R}^n)$. Then $\varphi\psi$ and $(\varphi * \psi)(x) = \int_{\mathbb{R}^n} \varphi(x - y)\psi(y)dy$ belong both to $\mathscr{S}(\mathbb{R}^n)$ and it holds
$$D^\alpha(\varphi * \psi) = (D^\alpha \varphi) * \psi = \varphi * (D^\alpha \psi).$$

Proof. We have
$$|\varphi_j(x)-\varphi(x)| \le (1+|x|^2)^{-k/2}\|\varphi_j-\varphi\|_{(k,0)}.$$
Choosing k large enough, integrating over $x\in\mathbb{R}^n$ and using $\|\varphi_j-\varphi\|_{(k,0)}\to 0$ gives (i).

Similarly, (ii) and (iii) follows by using only the definition of $\mathscr{S}(\mathbb{R}^n)$.

To prove (iv), we have to estimate $\|\varphi(x+h)-\varphi(x)\|_{(k,l)}$. We estimate using Taylor theorem
$$|D^\alpha\varphi(x+h)-D^\alpha\varphi(x)| \le c_\alpha|h|\cdot \sup_{|\beta|=1}\sup_{y\in[x,x+h]}|D^{\alpha+\beta}\varphi(y)|$$
and obtain
$$\|\tau_h\varphi-\varphi\|_{(k,l)} \le c_{k,l}|h|\cdot\|\varphi\|_{(k,l+1)}, \quad |h|\le 1.$$

Finally, we give the proof of (v). Let $\varphi,\psi\in\mathscr{S}(\mathbb{R}^n)$. Then $(\varphi*\psi)$ is well defined, i.e. the integral converges for every $x\in\mathbb{R}^n$. The identity for derivatives of $(\varphi*\psi)$ follows by Lebesgue's dominated convergence theorem. Indeed, let $e_j=(0,\dots,0,1,0,\dots,0)$ with 1 in the jth entry and zeros everywhere else.
$$\begin{aligned}
D^{e_j}(\varphi*\psi)(x) &= \lim_{h\to 0}\frac{(\varphi*\psi)(x+he_j)-(\varphi*\psi)(x)}{h}\\
&= \lim_{h\to 0}\frac{1}{h}\int_{\mathbb{R}^n}[\varphi(x+he_j-y)-\varphi(x-y)]\psi(y)dy\\
&= \lim_{h\to 0}\int_{\mathbb{R}^n}\frac{\varphi(z+he_j)-\varphi(z)}{h}\cdot\psi(x-z)dz
\end{aligned}$$

Then
$$\frac{\varphi(z+he_j)-\varphi(z)}{h} - D^{e_j}\varphi(z)\to 0$$
for every $z\in\mathbb{R}^n$ as $h\to 0$ and since this expression is uniformly bounded by a constant depending only on φ, its integral with respect to the measure $\psi(x-z)dz$ converges to 0 as $h\to 0$. The general case follows by repeating this argument by induction.

Next we show, that $\varphi*\psi$ belongs also to $\mathscr{S}(\mathbb{R}^n)$. We have (for $N>n$ arbitrary)
$$|(\varphi*\psi)(x)| \le C_N\int_{\mathbb{R}^n}(1+|x-y|^2)^{-N/2}(1+|y|^2)^{-N/2}dy, \tag{14.1}$$
where C_N depends only on φ,ψ and $N>n$. The part of the integral in (14.1) over the set $\{y\in\mathbb{R}^n : |x-y|>|x|/2\}$ is bounded by
$$\int_{y:|x-y|>|x|/2}(1+(|x|/2)^2)^{-N/2}(1+|y|^2)^{-N/2}dy \le B_N(1+|x|^2)^{-N/2},$$
where B_N depends only on N and the dimension. On the other hand, if $|x|/2>|y-x|$, then $|y|>|x|/2$ and the part of the integral in (14.1) over the set $\{z\in\mathbb{R}^n : |y-x|<|x|/2\}$ may be estimated from above by
$$\int_{y:|x-y|<|x|/2}(1+|x-y|^2)^{-N/2}(1+(|x|/2)^2)^{-N/2}dy \le B_N(1+|x|^2)^{-N/2}.$$
Therefore, $(\varphi*\psi)(x)$ decays at infinity as $(1+|x|^2)^{-N/2}$ - and this holds for every $N>n$. Using $D^\alpha(\varphi*\psi)=(D^\alpha\varphi*\psi)$ and replacing φ by $D^\alpha\varphi$, we obtain by the same argument, that also $D^\alpha(\varphi*\psi)$ decays faster then the reciprocal of any polynomial, and $\varphi*\psi\in\mathscr{S}(\mathbb{R}^n)$ follows.

Finally, that $\varphi\psi$ belongs to $\mathscr{S}(\mathbb{R}^n)$ follows directly by the Leibniz rule.

□

Definition 14.1.4. Let $\varphi \in \mathscr{S}(\mathbb{R}^n)$. Then

$$(\mathcal{F}\varphi)(\xi) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \varphi(x) e^{-i\langle x,\xi\rangle} dx, \quad \xi \in \mathbb{R}^n$$

is called the *Fourier transform* of φ. Furthermore,

$$(\mathcal{F}^{-1}\varphi)(\xi) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \varphi(x) e^{i\langle x,\xi\rangle} dx, \quad \xi \in \mathbb{R}^n$$

is called the *inverse Fourier transform* of φ.

We will prove later on, that the terminology used is correct, especially, that $\mathcal{F}^{-1}$ really is an inverse of $\mathcal{F}$.
Furthermore, let us mention that many authors use another normalization of Fourier transform, i.e.

$$\int_{\mathbb{R}^n} f(x) e^{-2\pi i\langle x,\xi\rangle} dx.$$

Although this difference is harmless in general, it makes comparison between different books and textbooks rather inconvenient, and a lot of care is necessary.

Theorem 14.1.5. *i) Let $\varphi \in \mathscr{S}(\mathbb{R}^n)$. Then $\mathcal{F}\varphi \in \mathscr{S}(\mathbb{R}^n)$ and $\mathcal{F}^{-1}\varphi \in \mathscr{S}(\mathbb{R}^n)$.*

ii) Let $\varphi \in \mathscr{S}(\mathbb{R}^n)$. Then

$$D^\alpha(\mathcal{F}\varphi)(\xi) = (-i)^{|\alpha|}\mathcal{F}(x^\alpha\varphi)(\xi), \quad \alpha \in \mathbb{N}_0^n,\ \xi \in \mathbb{R}^n, \tag{14.2}$$

$$\xi^\alpha(\mathcal{F}\varphi)(\xi) = (-i)^{|\alpha|}\mathcal{F}(D^\alpha\varphi)(\xi), \quad \alpha \in \mathbb{N}_0^n,\ \xi \in \mathbb{R}^n. \tag{14.3}$$

iii) Let $(\varphi_j)_{j=1}^\infty \subset \mathscr{S}(\mathbb{R}^n)$ with $\varphi_j \xrightarrow{\mathscr{S}} \varphi$. Then

$$\mathcal{F}\varphi_j \xrightarrow{\mathscr{S}} \mathcal{F}\varphi \quad \text{and} \quad \mathcal{F}^{-1}\varphi_j \xrightarrow{\mathscr{S}} \mathcal{F}^{-1}\varphi.$$

Proof. If $\varphi \in \mathscr{S}(\mathbb{R}^n)$ and $\alpha \in \mathbb{N}_0^n$, then one gets from Theorem 14.1.3 that $x^\alpha\varphi \in \mathscr{S}(\mathbb{R}^n)$ and $D^\alpha\varphi \in \mathscr{S}(\mathbb{R}^n)$. Furthermore, we get by Lebesgue's dominated convergence theorem

$$\frac{\partial}{\partial\xi_l}(\mathcal{F}\varphi)(\xi) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} (-i)x_l e^{-i\langle x,\xi\rangle}\varphi(x)dx = (-i)\mathcal{F}(x_l\varphi(x))(\xi).$$

By iteration, we get (14.2). To prove (14.3), we use partial integration in x_l-direction with intervals tending to $\mathbb{R}$ to obtain

$$\begin{aligned}\xi_l(\mathcal{F}\varphi)(\xi) &= \frac{i}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \frac{\partial}{\partial x_l}(e^{-i\langle x,\xi\rangle})\varphi(x)dx = \frac{-i}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \frac{\partial\varphi}{\partial x_l}(x)e^{-i\langle x,\xi\rangle}dx \\ &= (-i)\mathcal{F}\left(\frac{\partial\varphi}{\partial x_l}\right)(\xi).\end{aligned}$$

These formulas can be easily iterated, cf.

$$\frac{\partial^2}{\partial\xi_m\partial\xi_l}(\mathcal{F}\varphi)(\xi) = (-i)\frac{\partial}{\partial\xi_m}(\mathcal{F}(x_l\varphi(x))(\xi)) = (-i)^2\mathcal{F}(x_m x_l\varphi(x))(\xi).$$

Finally, by (14.2) and (14.3), we obtain[10]

$$
\begin{aligned}
\|\mathcal{F}\varphi\|_{(k,l)} &\le c \sup_{\xi\in\mathbb{R}^n} \max_{|\beta|\le k} \max_{|\alpha|\le l} |\xi^\beta|\cdot|D^\alpha(\mathcal{F}\varphi)(\xi)|\\
&\le c \max_{|\beta|\le k} \max_{|\alpha|\le l} \|\mathcal{F}(D^\beta(x^\alpha\varphi(x)))\|_\infty\\
&\le c \max_{|\beta|\le k} \max_{|\alpha|\le l} \|D^\beta(x^\alpha\varphi(x))\|_1\\
&\le c \max_{|\beta|\le k} \max_{|\alpha|\le l} \|x^\alpha(D^\beta\varphi)(x))\|_1\\
&\le c\,\|\varphi\|_{(l+n+1,k)}, \qquad \text{for all} \quad \varphi\in\mathscr{S}(\mathbb{R}^n).
\end{aligned}
$$

□

Theorem 14.1.6. *i) Let $\varepsilon>0$ and let $\varphi\in\mathscr{S}(\mathbb{R}^n)$. Then*

$$\mathcal{F}(\varphi(\varepsilon\cdot))(\xi)=\varepsilon^{-n}\mathcal{F}(\varphi)(\xi/\varepsilon),\quad \xi\in\mathbb{R}^n.$$

ii) Let $h\in\mathbb{R}^n$ and $\varphi\in\mathscr{S}(\mathbb{R}^n)$. Then we denote by $(\tau_h\varphi)(x)=\varphi(x-h)$ the translation operator*. And the formula $\mathcal{F}(\tau_h\varphi)=e^{-i\langle h,\xi\rangle}\mathcal{F}\varphi$ holds for all $\varphi\in\mathscr{S}(\mathbb{R}^n)$.*

iii) Let $h\in\mathbb{R}^n$ and $\varphi\in\mathscr{S}(\mathbb{R}^n)$. Then we denote by $(M_h\varphi)(x)=e^{i\langle h,x\rangle}\varphi(x)$ the modulation *of φ. And the formula $\mathcal{F}(M_h\varphi)=\tau_h(\mathcal{F}\varphi)$ holds for every $\varphi\in\mathscr{S}(\mathbb{R}^n)$.*

iv) Let $\varphi,\psi\in\mathscr{S}(\mathbb{R}^n)$. Then the convolution *of φ and ψ satisfies the formula*

$$\mathcal{F}(\varphi*\psi)(\xi)=(2\pi)^{n/2}(\mathcal{F}\varphi)(\xi)\cdot(\mathcal{F}\psi)(\xi),\quad \xi\in\mathbb{R}^n.$$

v) If $\varphi\in\mathscr{S}(\mathbb{R}^n)$ and $\psi\in\mathscr{S}(\mathbb{R}^m)$, then $(\varphi\otimes\psi)(x,y)=\varphi(x)\psi(y)\in\mathscr{S}(\mathbb{R}^{n+m})$ is the tensor product *of φ and ψ. And the formula $\mathcal{F}(\varphi\otimes\psi)(\xi,\eta)=(\mathcal{F}\varphi)(\xi)(\mathcal{F}\psi)(\eta)$ holds for every $\xi\in\mathbb{R}^n$ and every $\eta\in\mathbb{R}^m$.*

vi) Let $\varphi,\psi\in\mathscr{S}(\mathbb{R}^n)$. Then $\mathcal{F}(\varphi\psi)=(2\pi)^{-n/2}\mathcal{F}(\varphi)*\mathcal{F}(\psi)$.

vii) It holds

$$\mathcal{F}(e^{-|x|^2/2})(\xi)=e^{-|\xi|^2/2},\quad \xi\in\mathbb{R}^n.$$

Proof. (i) - (v) are easy consequences of the Definitions. Due to the product structure, it is enough to show (vii) in $n=1$.
Let

$$h(s)=\frac{1}{(2\pi)^{1/2}}\int_{-\infty}^{\infty}e^{-t^2/2}e^{-its}dt,\quad s\in\mathbb{R}.$$

Then

$$h(0)^2=\frac{1}{2\pi}\int_{\mathbb{R}^2}e^{-(t^2+u^2)/2}d(t,u)=\frac{1}{2\pi}\int_0^\infty(2\pi r)\,e^{-r^2/2}dr=1,$$

i.e. $h(0)=1$, is the famous Gaussian integral evaluated using polar coordinates in $\mathbb{R}^2$. On the other hand, by integration by parts we get

$$
\begin{aligned}
h'(s)&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-t^2/2}e^{-its}(-it)dt=\frac{i}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{d}{dt}\left(e^{-t^2/2}\right)e^{-its}dt\\
&=\frac{-s}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-t^2/2}e^{-its}dt=-sh(s).
\end{aligned}
$$

[10] Let us note, that the meaning of c may change from one line to the other, and that it may depend on k and l, but not on $\varphi\in\mathscr{S}(\mathbb{R}^n)$.

Solving this equation, we observe that $h(s) = h(0)e^{-s^2/2} = e^{-s^2/2}$, which finishes the proof.
We leave the proof of (vi) open in the moment and come back to that once we establish the properties of the inverse Fourier transform. □

Obviously, similar statements can be proven for $\mathcal{F}^{-1}$ in the same way. So far we know that $\mathcal{F}\mathscr{S}(\mathbb{R}^n) \subset \mathscr{S}(\mathbb{R}^n)$ and $\mathcal{F}^{-1}\mathscr{S}(\mathbb{R}^n) \subset \mathscr{S}(\mathbb{R}^n)$.

Theorem 14.1.7. *Let $\varphi \in \mathscr{S}(\mathbb{R}^n)$. Then*

$$\varphi = \mathcal{F}^{-1}(\mathcal{F}\varphi) = \mathcal{F}(\mathcal{F}^{-1}\varphi).$$

Furthermore, $\mathcal{F}$ and $\mathcal{F}^{-1}$ both map $\mathscr{S}(\mathbb{R}^n)$ one-to-one onto itself.

Proof. When trying to evaluate $\mathcal{F}^{-1}(\mathcal{F}\varphi)$ using Fubini's theorem, one runs into non-convergent integrals. Therefore, we use the following trick. We evaluate

$$\frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} (\mathcal{F}\varphi)(\xi) e^{i\langle x,\xi\rangle} e^{-\varepsilon^2|\xi|^2/2} d\xi \to \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} (\mathcal{F}\varphi)(\xi) e^{i\langle x,\xi\rangle} d\xi = \mathcal{F}^{-1}(\mathcal{F}\varphi)(x)$$

as $\varepsilon \to 0^+$.

Let $\psi(x) = e^{-\varepsilon^2|x|^2/2}$. Then

$$\mathcal{F}\psi(y) = \varepsilon^{-n}\mathcal{F}\left(e^{-|x|^2/2}\right)(y/\varepsilon) = \varepsilon^{-n}e^{-\frac{|y|^2}{2\varepsilon^2}}.$$

Furthermore, we get by change of variables (even for arbitrary $\varphi, \psi \in \mathscr{S}(\mathbb{R}^n)$)

$$\begin{aligned}\int_{\mathbb{R}^n} (\mathcal{F}\varphi)(\xi) e^{i\langle x,\xi\rangle}\psi(\xi) d\xi &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n\times\mathbb{R}^n} \varphi(y) e^{-i\langle y-x,\xi\rangle}\psi(\xi) d(\xi, y) \\ &= \int_{\mathbb{R}^n} \varphi(y)\mathcal{F}\psi(y-x)dy = \int_{\mathbb{R}^n} \varphi(x+y)\mathcal{F}\psi(y)dy.\end{aligned} \tag{14.4}$$

This, applied to $\psi(x) = e^{-\varepsilon^2|x|^2/2}$ then gives

$$\begin{aligned}\frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} (\mathcal{F}\varphi)(\xi) e^{i\langle x,\xi\rangle} e^{-\varepsilon^2|\xi|^2/2} d\xi &= \frac{\varepsilon^{-n}}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \varphi(x+y) e^{-\frac{|y|^2}{2\varepsilon^2}} dy \\ &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \varphi(x+\varepsilon z) e^{-|z|^2/2} dz.\end{aligned}$$

By Lebesgue dominated convergence theorem, this tends to $\varphi(x)$ as $\varepsilon \to 0^+$. □

Finally, we come to the proof of (vi) in Theorem 14.1.6. Let us observe that in (14.4), we have shown actually that

$$(2\pi)^{n/2}\mathcal{F}^{-1}(\mathcal{F}\varphi \cdot \psi) = \varphi * \mathcal{F}^{-1}\psi.$$

We plug in $\psi = \mathcal{F}g$ and $\varphi = f$ and the statement follows.

14.2 The space $\mathscr{S}'(\mathbb{R}^n)$ and the Fourier transform on $\mathscr{S}'(\mathbb{R}^n)$

Definition 14.2.1. The space $\mathscr{S}'(\mathbb{R}^n)$ is the set of all continuous complex linear functionals on $\mathscr{S}(\mathbb{R}^n)$. That means, $T : \mathscr{S}(\mathbb{R}^n) \to \mathbb{C}$ belongs to $\mathscr{S}'(\mathbb{R}^n)$ if, and only if,

i) $T(\lambda\varphi + \mu\psi) = \lambda T(\varphi) + \mu T(\psi)$ for all $\lambda, \mu \in \mathbb{C}$ and $\varphi, \psi \in \mathscr{S}(\mathbb{R}^n)$;

ii) $T\varphi_j \to T\varphi$ whenever $\varphi_j \xrightarrow{\mathscr{S}} \varphi$.

The elements of the space $\mathscr{S}'(\mathbb{R}^n)$ are called *tempered distributions*.

Remark 14.2.2. $\mathscr{S}'(\mathbb{R}^n)$ is again a linear space, as we put $(\lambda T+\mu S)(\varphi) = \lambda T(\varphi)+\mu S(\varphi)$ for all $T, S \in \mathscr{S}'(\mathbb{R}^n)$ and all $\lambda, \mu \in \mathbb{C}$. Although one could again equip $\mathscr{S}'(\mathbb{R}^n)$ with a topology and turn it into a *topological vector space*, for our purposes it is enough to equip it with *weak convergence*. We say, that

$$T_j \overset{\mathscr{S}'}{\rightharpoonup} T \Leftrightarrow T_j(\varphi) \to T(\varphi) \text{ for all } \varphi \in \mathscr{S}(\mathbb{R}^n).$$

Theorem 14.2.3. *Let T be a linear functional on $\mathscr{S}(\mathbb{R}^n)$. Then $T \in \mathscr{S}'(\mathbb{R}^n)$ if, and only if*

$$|T(\varphi)| \le c\|\varphi\|_{(k,l)} \tag{14.5}$$

for some $c > 0, k, l \in \mathbb{N}_0$ and all $\varphi \in \mathscr{S}(\mathbb{R}^n)$.

Proof. If T satisfies (14.5), then $T \in \mathscr{S}'(\mathbb{R}^n)$. To prove the converse, we proceed by contradiction, i.e. for all $c > 0, k, l \in \mathbb{N}_0$, there is $\varphi_{c,k,l} \in \mathscr{S}(\mathbb{R}^n)$ with $1 = |T(\varphi_{c,k,l})| > c\|\varphi_{c,k,l}\|_{(k,l)}$. We consider the sequence $\varphi_k = \varphi_{k,k,k}$. We obtain, that $\varphi_k \xrightarrow{\mathscr{S}} 0$ and (by $T \in \mathscr{S}'(\mathbb{R}^n)$) also $T(\varphi_k) \to 0$, which is a contradiction. □

Let f be an integrable function, i.e. $f \in L_1(\mathbb{R}^n)$. Then the mapping

$$T_f : \varphi \to T_f(\varphi) = \int_{\mathbb{R}^n} f(x)\varphi(x)dx, \quad \varphi \in \mathscr{S}(\mathbb{R}^n)$$

is well-defined, linear and due to

$$\left|\int_{\mathbb{R}^n} f(x)\varphi(x)dx\right| \le \|f\|_1 \cdot \|\varphi\|_\infty = \|f\|_1 \cdot \|\varphi\|_{(0,0)}$$

is also continuous, hence $T \in \mathscr{S}'(\mathbb{R}^n)$. With a slight modification of the arguments, the same holds for also for $f \in L_p(\mathbb{R}^n)$ for all $1 \le p \le \infty$. However, it does *not* hold for $p < 1$, and *not* for $L_1^{\mathrm{loc}}(\mathbb{R}^n)$.

Every distribution, which is equal to T_f for some $f \in L_1(\mathbb{R}^n)$, is called *regular distribution*. Every distribution, which is not regular is called *singular distribution*. In this sense, we may identify $L_1(\mathbb{R}^n)$ with the set of regular distributions, which is a subset of $\mathscr{S}'(\mathbb{R}^n)$. The following lemma shows, that this identification is also one-to-one.

Lemma 14.2.4. *Let $f, g \in L_1(\mathbb{R}^n)$ and let $T_f(\varphi) = T_g(\varphi)$ for all $\varphi \in \mathscr{S}(\mathbb{R}^n)$. Then $f = g$ a.e.*

Proof. There is a lot of different proofs in the literature. We shall present one which uses the technique of *mollification*, which we shall encounter also later on. Obviously, it is enough to show that if $T_h(\varphi) = 0$ for some $h \in L_1(\mathbb{R}^n)$ and all $\varphi \in \mathscr{S}(\mathbb{R}^n)$, then $h = 0$ a.e.

Let $\omega \in \mathscr{S}(\mathbb{R}^n)$ be a smooth non-negative symmetric function with support in $B := \{y : |y| \le 1\}$ and $\int_{\mathbb{R}^n} \omega(x)dx = 1$. We define $\omega_\varepsilon(x) = \varepsilon^{-n}\omega(x/\varepsilon)$. Observe that also $\int_{\mathbb{R}^n} \omega_\varepsilon(x)dx = 1$. We observe the following facts

i) $h * \omega_\varepsilon$ is a continuous function on the whole $\mathbb{R}^n$ for each $\varepsilon > 0$.

ii) Using Fubini's theorem (and the symmetry of ω) we get quickly (as $\omega_\varepsilon * \varphi \in \mathscr{S}(\mathbb{R}^n)$)

$$0 = \int_{\mathbb{R}^n} h(x)(\omega_\varepsilon * \varphi)(x)dx = \int_{\mathbb{R}^n} \varphi(y)(h * \omega_\varepsilon)(y)dy.$$

iii) As $\omega_\varepsilon * h$ is a continuous function, it must be positive on some neighborhood of any point in which it is positive. Considering $\varphi \in \mathscr{S}(\mathbb{R}^n)$ with support in this neighborhood and the equation above, we obtain $(\omega_\varepsilon * h)(x) = 0$ for all $x \in \mathbb{R}^n$.

iv) Finally, we use that, for every $t > 0$, $h \in L_1(\mathbb{R}^n)$ may be written as $h = h_1 + h_2$, where h_1 is continuous with compact support and $\|h_2\|_1 \le t$.

v) The proof is finished by

$$\begin{aligned}\|h\|_1 &= \|h - \omega_\varepsilon * h\|_1 = \|h_1 + h_2 - \omega_\varepsilon * h_1 - \omega_\varepsilon * h_2\|_1\\ &\le \underbrace{\|h_1 - \omega_\varepsilon * h_1\|_1}_{(*)} + \underbrace{\|h_2\|_1 + \|\omega_\varepsilon * h_2\|_1}_{\le 2t},\end{aligned}$$

where

$$\begin{aligned}(*) &= \int_{\mathbb{R}^n} |h_1(x) - \omega_\varepsilon * h_1(x)|dx\\ &= \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} [h_1(x) - h_1(x-y)]\omega_\varepsilon(y)dy \right| dx\\ &\le \int_{\mathbb{R}^n} \omega_\varepsilon(y) \int_{\mathbb{R}^n} |h_1(x) - h_1(x-y)|dxdy\\ &\le \sup_{y:|y|\le\varepsilon} \|h_1(\cdot) - h_1(\cdot - y)\|_1.\end{aligned}$$

Due to the bounded support of h_1 and its uniform continuity, the last expression goes to zero as $\varepsilon \to 0$. Hence, choosing $\varepsilon, t > 0$ small enough, $\|h\|_1$ is arbitrary small and, therefore, equal to zero, i.e. $h = 0$ a.e.

□

Any finite Borel measure μ on $\mathbb{R}^n$ is a tempered distribution via

$$\mu(\varphi) = \int_{\mathbb{R}^n} \varphi(x)d\mu(x).$$

We show how one can extend the convolution also to measures. Let first $\varphi, \psi \in \mathscr{S}(\mathbb{R}^n)$ and $f \in \mathscr{S}(\mathbb{R}^n)$ as well. We denote by μ the measure on $\mathbb{R}^n$, which has density φ with respect to the Lebesgue measure. Furthermore, λ corresponds to ψ in the same way. Then

$$\begin{aligned}(\varphi * \psi)(f) &= \int_{\mathbb{R}^n} (\varphi * \psi)(x)f(x)dx = \int_{\mathbb{R}^n} f(x) \int_{\mathbb{R}^n} \varphi(x-y)\psi(y)dydx\\ &= \int_{\mathbb{R}^n\times\mathbb{R}^n} f(x)\varphi(x-y)\psi(y)dydx = \int_{\mathbb{R}^n\times\mathbb{R}^n} f(z+y)\varphi(z)\psi(y)dzdy\\ &= \int_{\mathbb{R}^n\times\mathbb{R}^n} f(z+y)d\mu(z)d\lambda(y).\end{aligned}$$

We observe that the last expression makes sense for all Borel measures μ and λ and $f \in C_0(\mathbb{R}^n)$, i.e. continuous functions which tend to zero at infinity.

Hence, for two Borel measures μ and λ on $\mathbb{R}^n$, we define $\mu * \lambda$ to be the unique measure on $\mathbb{R}^n$, such that (using the Riesz representation theorem)

$$\int_{\mathbb{R}^n} f d(\mu * \lambda) = \int_{\mathbb{R}^n \times \mathbb{R}^n} f(x+y) d\mu(x) d\lambda(y) \tag{14.6}$$

for all $f \in C_0(\mathbb{R}^n)$. Using standard arguments, this holds also for all bounded Borel-measurable functions, i.e. $f_\xi(x) = e^{i\langle x, \xi\rangle}$. Therefore, we obtain immediately the formula

$$\mathcal{F}(\mu * \lambda)(\xi) = (2\pi)^{n/2} (\mathcal{F}\mu)(\xi)(\mathcal{F}\lambda)(\xi), \quad \xi \in \mathbb{R}^n.$$

Furthermore, $\|\mu * \lambda\| \leq \|\mu\| \cdot \|\lambda\|$, where the norms denote total variation of measures. Finally, if μ is absolutely continuous with respect to the Lebesgue measure, then the same is true for $\mu * \lambda$. To see this, consider $f = \chi_E$, where the Lebesgue measure of E is zero. Then

$$0 = \int_{\mathbb{R}^n} f(x+y) d\mu(x)$$

for all $y \in \mathbb{R}^n$, and $(\mu * \lambda)(E) = 0$ follows.

Finally, (14.6) implies that

$$(\mu * \lambda)(E) = (\mu \otimes \lambda)(\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n : x + y \in E\})$$

for every Borel-measurable set $E \subset \mathbb{R}^n$.

Definition 14.2.5. Let $T \in \mathscr{S}'(\mathbb{R}^n)$, $\varphi, \psi \in \mathscr{S}(\mathbb{R}^n)$ and $\alpha \in \mathbb{N}_0^n$. Then we put

$$(D^\alpha T)(\varphi) = (-1)^{|\alpha|} T(D^\alpha \varphi),$$

$$(\mathcal{F}T)(\psi) = T(\mathcal{F}\psi), \quad (\mathcal{F}^{-1}T)(\psi) = T(\mathcal{F}^{-1}\psi)$$

and

$$(\varphi T)(\psi) = T(\varphi\psi). \tag{14.7}$$

Actually, we use (14.7) to define φT for every (infinitely-differentiable) function φ, such that $\varphi\psi \in \mathscr{S}(\mathbb{R}^n)$ for every $\psi \in \mathscr{S}(\mathbb{R}^n)$. This applies for example to the polynomials on $\mathbb{R}^n$ or to (linear combinations of) functions of the type $e^{i\langle \theta, x\rangle}$.

Using Theorem 14.2.3, one shows that $D^\alpha T, (\varphi T), \mathcal{F}T, \mathcal{F}^{-1}T \in \mathscr{S}'(\mathbb{R}^n)$.

Remark 14.2.6. Let us sketch the proof (given by Schwartz) that it is impossible to introduce a continuous multiplication of distributions. Let $\psi \in \mathscr{S}(\mathbb{R}^n)$ be compactly supported non-negative function with $\psi(0) = 1$, $\int \psi = 1$. Then $\psi_k(x) := k^n \psi(kx)$ converge to δ in $\mathscr{S}'(\mathbb{R}^n)$ as $k \to \infty$. By the definition of multiplication of distribution δ and test function ψ_k, we obtain

$$(\delta \cdot \psi_k)(\varphi) = \delta(\psi_k \varphi) = \psi_k(0)\varphi(0) = k^n \varphi(0) \to \infty$$

for all $\varphi \in \mathscr{S}(\mathbb{R})$ with $\varphi(0) > 0$. On the other hand, assuming that it is possible to define continuous multiplication in $\mathscr{S}'(\mathbb{R}^n)$ implies that $(\delta\psi_k)(\varphi) \to (\delta \cdot \delta)(\varphi) = \delta^2(\varphi)$, i.e. a contradiction.

Theorem 14.2.7. *Let $T \in \mathscr{S}'(\mathbb{R}^n)$.*

i) *Then* $\mathcal{F}(\mathcal{F}^{-1}T) = \mathcal{F}^{-1}(\mathcal{F}T) = T$. *Furthermore, both* $\mathcal{F}$ *and* $\mathcal{F}^{-1}$ *map* $\mathscr{S}'(\mathbb{R}^n)$ *one-to-one continuously onto itself.*

ii) *Let* $\alpha \in \mathbb{N}_0^n$. *Then* $x^\alpha T \in \mathscr{S}'(\mathbb{R}^n)$ *and* $D^\alpha T \in \mathscr{S}'(\mathbb{R}^n)$. *Furthermore,*

$$\mathcal{F}(D^\alpha T) = i^{|\alpha|} x^\alpha(\mathcal{F}T) \quad \text{and} \quad \mathcal{F}(x^\alpha T) = i^{|\alpha|} D^\alpha(\mathcal{F}T).$$

iii) *Let* $\varepsilon > 0$ *and let* $T \in \mathscr{S}'(\mathbb{R}^n)$. *Then*

$$T(\varepsilon\cdot)(\varphi) := T(\varepsilon^{-n}\varphi(\cdot/\varepsilon)), \quad \varphi \in \mathscr{S}(\mathbb{R}^n)$$

is the dilation of T *and*

$$\mathcal{F}(T(\varepsilon\cdot)) = \varepsilon^{-n}\mathcal{F}(T)(\cdot/\varepsilon).$$

iv) *Let* $h \in \mathbb{R}^n$ *and* $T \in \mathscr{S}'(\mathbb{R}^n)$. *Then we denote by* $(\tau_h T)(\varphi) = T(\varphi(\cdot + h))$, $\varphi \in \mathscr{S}(\mathbb{R}^n)$, *the* translation operator. *The formula* $\mathcal{F}(\tau_h T) = e^{-i\langle h,\xi\rangle}\mathcal{F}T$ *holds for all* $T \in \mathscr{S}'(\mathbb{R}^n)$.

v) *Let* $h \in \mathbb{R}^n$ *and* $T \in \mathscr{S}'(\mathbb{R}^n)$. *Then we denote by* $(M_h T)(\varphi) = T(e^{i\langle h,x\rangle}\varphi), \varphi \in \mathscr{S}(\mathbb{R}^n)$, *the* modulation *of* T. *And the formula* $\mathcal{F}(M_h T) = \tau_h(\mathcal{F}T)$ *holds for every* $T \in \mathscr{S}'(\mathbb{R}^n)$.

Proof. We obtain

$$\mathcal{F}(\mathcal{F}^{-1}T)(\varphi) = (\mathcal{F}^{-1}T)(\mathcal{F}\varphi) = T(F^{-1}(\mathcal{F}\varphi)) = T(\varphi)$$

for every $\varphi \in \mathscr{S}(\mathbb{R}^n)$, i.e. $\mathcal{F}(\mathcal{F}^{-1}T) = T$. The other identities follow in the same manner. □

14.3 Fourier transform on $L_1(\mathbb{R}^n)$ and $L_2(\mathbb{R}^n)$

There is another natural definition of Fourier transform, which turns out to be essentially equivalent to definition given above.

Definition 14.3.1. Let $f \in L_1(\mathbb{R}^n)$. Then we define

$$\mathbb{F}f(\xi) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} f(x) e^{-i\langle x,\xi\rangle} dx, \quad \xi \in \mathbb{R}^n.$$

This integral converges absolutely for every $\xi \in \mathbb{R}^n$.

Theorem 14.3.2. *Let* $f \in L_1(\mathbb{R}^n)$ *and* $\xi \in \mathbb{R}^n$.

i) *If* $g(x) = f(x)e^{i\langle\alpha,x\rangle}$, *then* $\mathbb{F}g(\xi) = \mathbb{F}f(\xi - \alpha)$.

ii) *If* $g(x) = f(x - \alpha)$, *then* $\mathbb{F}g(\xi) = \mathbb{F}f(\xi)e^{-i\langle\alpha,\xi\rangle}$.

iii) *If* $g \in L_1(\mathbb{R}^n)$ *and* $h = f * g$, *then* $\mathbb{F}h(\xi) = (2\pi)^{n/2}\mathbb{F}f(\xi) \cdot \mathbb{F}g(\xi)$.

iv) *If* $g(x) = \overline{f(-x)}$, *then* $\mathbb{F}g(\xi) = \overline{\mathbb{F}f(\xi)}$.

v) *If* $g(x) = f(x/\lambda)$, *then* $\mathbb{F}g(\xi) = \lambda^n\mathbb{F}f(\lambda\xi)$.

vi) *If* $g(x) = -ix_j f(x)$, *and* $g \in L_1(\mathbb{R}^n)$, *then* $\mathbb{F}f$ *is differentiable at* ξ *and* $\partial_j(\mathbb{F}f)(\xi) = \mathbb{F}g(\xi)$.

Proof. The proof is done by direct substitutions to the Definition, the last statement follows by Lebesgue dominated convergence theorem. □

Let us mention, that the properties of $\mathbb{F}$ and $\mathcal{F}$ are very similar. Nevertheless, the identities above are understood in pointwise sense, the properties of $\mathcal{F}$ were proven in the distributional sense.

We show that $\mathbb{F} = \mathcal{F}$ on $L_1(\mathbb{R}^n) \hookrightarrow \mathscr{S}'(\mathbb{R}^n)$ (and we shall after that use only the letter $\mathcal{F}$ for any of the Fourier transforms).

Theorem 14.3.3. *Let $f \in L_1(\mathbb{R}^n)$. Then $\mathcal{F}f$ is a regular distribution and $\mathbb{F}f = \mathcal{F}f$ in the distributional sense.*

Proof. If $f \in L_1(\mathbb{R}^n)$, then $\mathbb{F}f$ is obviously bounded, cf.

$$|\mathbb{F}f(\xi)| = \frac{1}{(2\pi)^{n/2}} \left| \int_{\mathbb{R}^n} f(x) e^{-i\langle x,\xi\rangle} dx \right| \le \frac{\|f\|_1}{(2\pi)^{n/2}}, \quad \xi \in \mathbb{R}^n.$$

Therefore, $T_{\mathbb{F}f}$ may be interpreted as an element of $\mathscr{S}'(\mathbb{R}^n)$.

On the other hand, let $f \in L_1(\mathbb{R}^n)$ and $\varphi \in \mathscr{S}(\mathbb{R}^n)$. Then

$$\begin{aligned}
(\mathcal{F}f)(\varphi) = f(\mathcal{F}\varphi) &= \int_{\mathbb{R}^n} f(x)(\mathcal{F}\varphi)(x) dx \\
&= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} f(x) \int_{\mathbb{R}^n} \varphi(y) e^{-i\langle x,y\rangle} dy dx \\
&= \int_{\mathbb{R}^n} \varphi(y) \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} f(x) e^{-i\langle x,y\rangle} dx dy \\
&= \int_{\mathbb{R}^n} \varphi(y) \mathbb{F}f(y) dy = T_{\mathbb{F}f}(\varphi),
\end{aligned}$$

i.e. $(\mathcal{F}f)(\varphi) = (\mathbb{F}f)(\varphi)$ for all $\varphi \in \mathscr{S}(\mathbb{R}^n)$ are equal in the distributional sense. □

Theorem 14.3.4. *Let $f \in L_1(\mathbb{R}^n)$. Then $\mathcal{F}f \in C_0(\mathbb{R}^n)$, i.e. $\mathcal{F}f \in C(\mathbb{R}^n)$ and $\lim_{|x|\to\infty} \mathcal{F}f(x) = 0$.*

Proof. The boundedness of $\mathcal{F}f$ was already discussed above. The continuity follows from Lebesgue convergence theorem (with $|f(x)|$ as an integrable majorant):

$$\frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} f(x) e^{-i\langle \xi+h,x\rangle} dx \to \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} f(x) e^{-i\langle \xi,x\rangle} dx \text{ as } h \to 0.$$

Finally, if $n = 1$ and $f = \chi_{[a,b]}$ with $-\infty < a < b < \infty$, we get

$$\mathcal{F}\chi_{[a,b]}(\xi) = \frac{1}{(2\pi)^{1/2}} \int_a^b e^{-i\xi t} dt = \frac{1}{(2\pi)^{1/2}} \frac{e^{-i\xi a} - e^{-i\xi b}}{i\xi} \to 0$$

if $\xi \to \pm\infty$. In the same way, if $n > 1$ and $g(x) = \prod_{j=1}^n \chi_{[a_j,b_j]}(x_j)$ on $\mathbb{R}^n$, then

$$\mathcal{F}g(\xi) = \frac{1}{(2\pi)^{n/2}} \prod_{j=1}^n \frac{e^{-i\xi_j a_j} - e^{-i\xi_j b_j}}{i\xi_j} \to 0$$

if $|\xi| \to \infty$. The same conclusion therefore holds also for finite sums of step functions of intervals. Finally, we use that every function $f \in L_1(\mathbb{R}^n)$ might be approximated by such a finite sum function h to arbitrary precision in the $L_1(\mathbb{R}^n)$-norm. We obtain

$$|\mathcal{F}f(\xi)| \le |\mathcal{F}(f-h)(\xi)| + |\mathcal{F}h(\xi)| \le \frac{\|f-h\|_1}{(2\pi)^{n/2}} + |\mathcal{F}h(\xi)|.$$

As the last summand goes to zero and the first can be made arbitrary small, we obtain the conclusion. □

The non-distributional approach to Fourier transform now proceeds as follows

i) The inversion theorem is proven for $f \in L_1(\mathbb{R}^n)$ and $\mathcal{F}f \in L_1(\mathbb{R}^n)$, and holds pointwise a.e.

ii) $\mathcal{F}f$ is now defined on $L_1(\mathbb{R}^n)$.

iii) One shows that for $f \in L_1(\mathbb{R}^n) \cap L_2(\mathbb{R}^n)$, the following *Plancherel identity* holds: $\|\mathcal{F}f\|_2 = \|f\|_2$.

iv) As $L_1(\mathbb{R}^n) \cap L_2(\mathbb{R}^n)$ is dense in $L_2(\mathbb{R}^n)$, there is a unique extension of $\mathcal{F}$ from $L_1(\mathbb{R}^n) \cap L_2(\mathbb{R}^n)$ to $L_2(\mathbb{R}^n)$, which will be denoted by $\mathcal{F}$ again.

v) It follows that $\mathcal{F}$ is a Hilbert space isomorphism of $L_2(\mathbb{R}^n)$ onto $L_2(\mathbb{R}^n)$.

Theorem 14.3.5. *(Fourier inversion theorem on $L_1(\mathbb{R}^n)$)*
Let $f \in L_1(\mathbb{R}^n)$ and $\mathcal{F}f \in L_1(\mathbb{R}^n)$. Then

$$f(x) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} (\mathcal{F}f)(\xi) e^{i\langle \xi, x\rangle} d\xi$$

for almost every $x \in \mathbb{R}^n$.

Proof. We know from Lemma 14.2.4 that if $T_h(\varphi) = 0$ for $h \in L_1(\mathbb{R}^n)$ and all $\varphi \in \mathscr{S}(\mathbb{R}^n)$, then $h = 0$ a.e. This can be easily generalised. Let $u \in L_1(\mathbb{R}^n)$ and $v \in L_\infty(\mathbb{R}^n)$. Furthermore, let $\psi(x) = e^{-|x|^2/2}$. Then $(u+v)\psi \in L_1(\mathbb{R}^n)$ and $\int_{\mathbb{R}^n}(u+v)\varphi = 0$ for all $\varphi \in \mathscr{S}(\mathbb{R}^n)$ implies also $\int_{\mathbb{R}^n}[(u+v)\psi]\varphi = 0$ for all $\varphi \in \mathscr{S}(\mathbb{R}^n)$. Hence $(u+v)\psi = 0$ a.e. and $u + v = 0$ a.e. follows.

Using Fubini's theorem we get immediately

$$\int_{\mathbb{R}^n} f(x)(\mathcal{F}\varphi)(x)dx = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n \times \mathbb{R}^n} f(x)\varphi(\xi)e^{-i\langle x,\xi\rangle} d(x,\xi) = \int_{\mathbb{R}^n} (\mathcal{F}f)(\xi)\varphi(\xi)d\xi$$

for $f \in L_1(\mathbb{R}^n)$ and $\varphi \in \mathscr{S}(\mathbb{R}^n)$ (actually, even $\varphi \in L_1(\mathbb{R}^n)$ can be allowed). The same formula holds also for $\mathcal{F}^{-1}$. Finally, we obtain

$$\int [\mathcal{F}^{-1}(\mathcal{F}f)]\varphi = \int (\mathcal{F}f)(\mathcal{F}^{-1}\varphi) = \int f(\mathcal{F}^{-1}(\mathcal{F}\varphi)) = \int f\varphi.$$

As $\mathcal{F}^{-1}(\mathcal{F}f) \in L_\infty(\mathbb{R}^n)$ and $f \in L_1(\mathbb{R}^n)$, we obtain $\mathcal{F}^{-1}(\mathcal{F}f) = f$ a.e. □

As the zero function is in $L_1(\mathbb{R}^n)$, we obtain the following Corollary.

Corollary 14.3.6. *Let $f \in L_1(\mathbb{R}^n)$ and let $\mathcal{F}f = 0$ a.e. Then $f = 0$ a.e.*

Theorem 14.3.7. *Let $f \in L_1(\mathbb{R}^n) \cap L_2(\mathbb{R}^n)$, then*

$$\|f\|_2 = \|\mathcal{F}f\|_2.$$

Proof. For $f \in L_1(\mathbb{R}^n) \cap L_2(\mathbb{R}^n)$, let $h = f * \bar{\tilde{f}}$. Then we have the following

i) $h \in L_1(\mathbb{R}^n)$;

ii) h is continuous at 0; indeed

$$h(x) - h(0) = \int_{\mathbb{R}^n} f(y)\overline{\tilde{f}(x-y)}dy - \int_{\mathbb{R}^n} f(y)\overline{\tilde{f}(-y)}dy$$
$$= \int_{\mathbb{R}^n} f(y)\overline{[f(y-x) - f(y)]}dy.$$

Writing $f = f_1 + f_2$, where f_1 is continuous with compact support and $\|f_2\|_2 \le t$, we get

$$|h(x) - h(0)| \le \int_{\mathbb{R}^n} |f(y)| \cdot |f_1(y-x) - f_1(y)|dy + \int_{\mathbb{R}^n} |f(y)| \cdot |f_2(y-x) - f_2(y)|dy.$$

The first integral tends to zero (using again the compact support of f_1 and its uniform continuity), the second is bounded by $2t\|f\|_2 < \infty$.

iii) $\mathcal{F}h = (2\pi)^{n/2}|\mathcal{F}f|^2 \ge 0$;

iv) Finally, we obtain

$$\|\mathcal{F}f\|_2^2 = (2\pi)^{-n/2}\int_{\mathbb{R}^n} (\mathcal{F}h)(\xi)d\xi = \lim_{\varepsilon\to 0}(2\pi)^{-n/2}\int_{\mathbb{R}^n} (\mathcal{F}h)(\xi)e^{-\varepsilon^2|\xi|^2/2}d\xi$$
$$= \lim_{\varepsilon\to 0}\varepsilon^{-n}(2\pi)^{-n/2}\int_{\mathbb{R}^n} h(x)e^{-|x|^2/(2\varepsilon^2)}dx$$
$$= h(0) = \int_{\mathbb{R}^n} f(x)\overline{\tilde{f}(-x)}dx = \|f\|_2^2,$$

where we used the continuity of h at zero and

$$\lim_{\varepsilon\to 0}\left|\varepsilon^{-n}(2\pi)^{-n/2}\int_{\mathbb{R}^n} h(x)e^{-|x|^2/(2\varepsilon^2)}dx - h(0)\right|$$
$$= \lim_{\varepsilon\to 0}\varepsilon^{-n}(2\pi)^{-n/2}\int_{\mathbb{R}^n} |h(x) - h(0)|e^{-|x|^2/(2\varepsilon^2)}dx$$
$$\le \lim_{\varepsilon\to 0}\varepsilon^{-n}(2\pi)^{-n/2}\int_{x:|x|\ge t} |h(x) - h(0)|e^{-|x|^2/(2\varepsilon^2)}dx$$
$$+ \lim_{\varepsilon\to 0}\varepsilon^{-n}(2\pi)^{-n/2}\int_{x:|x|\le t} |h(x) - h(0)|e^{-|x|^2/(2\varepsilon^2)}dx.$$

The first integral goes (for every $t > 0$) to zero[11], the second can be made arbitrary small by choosing $t > 0$ small.

□

14.4 Few applications of Fourier transform

14.4.1 Bochner's Theorem

We say, that a complex-valued function $\Phi : \mathbb{R}^n \to \mathbb{C}$ is *positive-definite* if the matrix $(\Phi(x_j - x_k))_{j,k=1}^N$ is positive semi-definite for every $N \in \mathbb{N}$ and every choice of points $x_1, \ldots, x_N \in \mathbb{R}^n$, i.e. if

$$\sum_{j,k=1}^N c_j\overline{c_k}\Phi(x_j - x_k) \ge 0$$

[11] The integrated functions go pointwise to zero, and as $\sup_{|x|\ge t, 0<\varepsilon<1} \varepsilon^{-n}e^{-|x|^2/2\varepsilon^2}$ is finite, the integrable majorant is $|h(x) - h(0)| \in L_1(\mathbb{R}^n)$.

for every $N \in \mathbb{N}$, every $x_1, \dots, x_N \in \mathbb{R}^n$ and every $c = (c_1, \dots, c_N) \in \mathbb{C}^N$.

Let Φ be a Fourier transform of a finite positive Borel measure ν, i.e.

$$\Phi(\xi) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{-i\langle x,\xi\rangle} d\nu(x), \quad \xi \in \mathbb{R}^n.$$

We observe, that each such function is positive-definite by

$$\begin{aligned}
\sum_{j,k=1}^N c_j \overline{c_k} \Phi(x_j - x_k) &= \sum_{j,k=1}^N c_j \overline{c_k} \cdot \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{-i\langle x_j - x_k,\xi\rangle} d\nu(\xi) \\
&= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \left(\sum_{j,k=1}^N c_j \overline{c_k} e^{-i\langle x_j - x_k,\xi\rangle} \right) d\nu(\xi) \\
&= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \left(\sum_{j,k=1}^N c_j e^{-i\langle x_j,\xi\rangle} \overline{c_k e^{-i\langle x_k,\xi\rangle}} \right) d\nu(\xi) \\
&= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \left| \sum_{j=1}^N c_j e^{-i\langle x_j,\xi\rangle} \right|^2 d\nu(\xi) \geq 0.
\end{aligned}$$

This gives immediately the easy part of the following theorem.

Theorem 14.4.1. *The Fourier transform of every positive Borel measure on $\mathbb{R}^n$ is a positive-definite function. On the other hand, if $\Phi : \mathbb{R}^n \to \mathbb{C}$ is continuous and positive-definite, then Φ is a Fourier transform of a finite positive Borel measure.*

Open problem: If $f : \mathbb{R}^n \to \mathbb{C}$ is a positive and positive-definite continuous function (i.e. $f(x)$ is real, $f(x) \geq 0$ for all $x \in \mathbb{R}^n$ and $f = \mathcal{F}\mu$ for some positive finite Borel measure μ) with $f(0) = 1$, then

$$\sum_{j,k=1}^N c_j \overline{c_k} f(x_j - x_k) \geq \frac{|c_1 + \cdots + c_N|^2}{N}$$

for all $x_1, \dots, x_N \in \mathbb{R}^n$ and all $c = (c_1, \dots, c_N) \in \mathbb{C}^N$.

14.4.2 Translation invariant spaces

A subspace M of $L_2(\mathbb{R}^n)$ is called *translation invariant* if $f \in M$ implies $f_\alpha(x) := f(x - \alpha) \in M$ for all $\alpha \in \mathbb{R}^n$.

The task we are going to investigate in this section sounds:

Describe the closed translation invariant subspaces of $L_2(\mathbb{R}^n)$.

We shall proceed as follows. We denote by $\hat{M}$ the image of M under $\mathcal{F}$. Then $\hat{M}$ is a closed subspace of $L_2(\mathbb{R}^n)$ (recall that $\mathcal{F}$ is an isometry on $L_2(\mathbb{R}^n)$). From the properties of Fourier transform, we obtain that $\hat{f} \in \hat{M}$ implies also $e^{-i\langle\alpha,x\rangle}\hat{f} \in \hat{M}$. Hence, $\hat{M}$ is invariant under multiplication with $e^{-i\langle\alpha,x\rangle}$.

Let $E \subset \mathbb{R}^n$ be any measurable subset of $\mathbb{R}^n$. Surely, if $\hat{M} = \{\varphi \in L_2(\mathbb{R}^n) : \varphi = 0 \text{ a.e. on } E\}$, then $\hat{M}$ is closed under multiplications with $e^{-i\langle\alpha,x\rangle}$. Moreover, $\hat{M}$ is closed (and $\hat{M}^\perp = \{\psi \in L_2(\mathbb{R}^n) : \psi = 0 \text{ a.e. on } \mathbb{R}^n \setminus E\}$). If M is the inverse image of $\hat{M}$, under Fourier transform, then M has exactly the desired properties.

One may now conjecture, that every closed translation invariant subspace is obtained exactly in this manner. So, to every closed translation invariant subspace M we have to construct a set $E \subset \mathbb{R}^n$, such that $f \in M$ if, and only if, $\mathcal{F}f(x) = 0$ a.e. on E. The obvious construction

$$E = \bigcap_{f \in M} E_f = \bigcap_{f \in M} \{x \in \mathbb{R}^n : \mathcal{F}f(x) = 0\}$$

runs into serious difficulties, when we realize that each E_f is defined only up to set of measure zero and that there are uncountably many $f \in M$. Hence, we lose every control about E.

So, let M be a closed translation invariant subspace of $L_2(\mathbb{R}^n)$ and let $\hat{M}$ be its image under Fourier transform. Let P denote the orthogonal projection onto $\hat{M}$. Hence, to each $f \in L_2(\mathbb{R}^n)$ there is $Pf \in \hat{M}$, such that $f - Pf \perp \hat{M}$, i.e.

$$(f - Pf) \perp Pg, \quad f, g \in L_2(\mathbb{R}^n)$$

and also

$$(f - Pf) \perp Pg(x)e^{-i\langle \alpha, x\rangle}, \quad f, g \in L_2(\mathbb{R}^n), \alpha \in \mathbb{R}^n.$$

This is equivalent to

$$\int_{\mathbb{R}^n} (f - Pf)(x)\overline{Pg(x)}e^{-i\langle \alpha, x\rangle} dx = 0, \quad f, g \in L_2(\mathbb{R}^n), \alpha \in \mathbb{R}^n,$$

i.e. that $\mathcal{F}((f - Pf)\overline{Pg}) = 0$. As both $f - Pf$ and Pg belong to $L_2(\mathbb{R}^n)$, their product is in $L_1(\mathbb{R}^n)$ and the uniqueness theorem on Fourier transform gives that $(f - Pf)\overline{Pg} = 0$ almost everywhere. This remains true also if we replace $\overline{Pg}$ by Pg, hence

$$f \cdot Pg = (Pf) \cdot (Pg), \quad f, g \in L_2(\mathbb{R}^n).$$

Interchanging f and g leads finally to

$$f \cdot Pg = Pf \cdot g, \quad f, g \in L_2(\mathbb{R}^n).$$

This may be (roughly speaking) interpreted as that $(Pf)/f$ is constant for all $f \in L_2(\mathbb{R}^n)$. To avoid the devision by zero here, we consider some strictly positive function in $L_2(\mathbb{R}^n)$, for example $g(x) = e^{-|x|^2}$ will do, and put

$$\varphi(x) := \frac{(Pg)(x)}{g(x)}, \quad x \in \mathbb{R}^n.$$

Then we get $Pf = \varphi \cdot f$ a.e. on $\mathbb{R}^n$. Now we observe that

$$\varphi^2 \cdot g = \varphi \cdot Pg = P^2 g = Pg = \varphi \cdot g,$$

i.e. $\varphi^2 = \varphi$. This means, that $\varphi = 0$ or $\varphi = 1$ a.e. and we put $E := \{x \in \mathbb{R}^n : \varphi(x) = 0\}$. Now $f \in \hat{M}$ if, and only if, $f = Pf = \varphi \cdot f$ gives that $\hat{M}$ consists of exactly those functions which vanish a.e. on E. This finishes the proof of the claim.

Let us mention that characterization of *shift* translation invariant spaces (i.e. space invariant under integer translations) was an open problem until recently.

14.4.3 Fast Fourier Transform

Let $\mathbb{Z}_N$ be the set of the N^{th} roots of unity, i.e. the set

$$\mathbb{Z}_N = \{1, e^{2\pi i/N}, e^{2\cdot 2\pi i/N} \dots, e^{(N-1)\cdot 2\pi i/N}\}.$$

Then $\mathbb{Z}_N$ with the usual complex multiplication is an Abelian (i.e. commutative) group. Furthermore, $\mathbb{Z}_N$ is isomorph to $\{0, 1, \dots, N\}$ equipped with summation modulo N. And it is also isomorph to $\mathbb{Z}/N\mathbb{Z}$, the set of equivalence classes based on the reminder when dividing by N.

Let us put

$$e_l(k) = e^{2\pi i l k/N} \quad \text{for } l = 0, 1, \dots, N-1 \text{ and } k = 0, 1, \dots, N-1.$$

Furthermore, we denote by V the vector space of complexed-valued functions on $\mathbb{Z}_N$ with

$$\langle F, G\rangle_V = \sum_{k=0}^{N-1} F(k)\overline{G(k)},$$

$$\|F\|_V^2 = \sum_{k=0}^{N-1} |F(k)|^2.$$

Through a simple calculation, one obtains immediately

Lemma 14.4.2. $\langle e_l, e_m\rangle_V = N \cdot \delta_{m,l}$.

Therefore, $e_l^* = e_l/\sqrt{N}, l = 0, 1, \dots, N-1$, is an orthonormal basis of V.

Hence, for every $F \in V$, we get

$$F = \sum_{n=0}^{N-1} \langle F, e_n^*\rangle_V e_n^*,$$

$$\|F\|^2 = \sum_{n=0}^{N-1} |\langle F, e_n^*\rangle_V|^2.$$

The n^{th} Fourier coefficient of F is defined as

$$a_n = \hat{F}(n) = \frac{1}{N}\sum_{k=0}^{N-1} F(k)e^{-2\pi i k n/N} = \frac{1}{\sqrt{N}}\langle F, e_n^*\rangle.$$

Theorem 14.4.3. *For $F \in V$ we have*

$$F(k) = \sum_{n=0}^{N-1} \langle F, e_n^*\rangle_V e_n^*(k) = \sum_{n=0}^{N-1} \sqrt{N} a_n e_n^*(k) = \sum_{n=0}^{N-1} a_n e_n(k) = \sum_{n=0}^{N-1} a_n e^{2\pi i n k/N},$$

$$\sum_{n=0}^{N-1} |a_n|^2 = \frac{1}{N}\sum_{n=0}^{N-1} |\langle F, e_n^*\rangle_V|^2 = \frac{1}{N}\|F\|^2 = \frac{1}{N}\sum_{k=0}^{N-1} |F(k)|^2.$$

Naive way how to compute $\hat{F}(0), \dots, \hat{F}(N-1)$ from $F(0), \dots, F(N-1)$ given and $\omega_N = e^{-2\pi i/N}$ given is

$$a_k^N(F) := \frac{1}{N}\sum_{r=0}^{N-1} F(r)\omega_N^{kr}.$$

It involves $N-2$ multiplications to get $\omega_N^2, \dots, \omega_N^{N-1}$ and each a_k^N needs $N+1$ multiplications and $N-1$ additions. Therefore, we need $2N^2 + N - 2 \leq 2N^2 + N$ operations.

Theorem 14.4.4. (Fast Fourier Transform) *Given $\omega_N = e^{-2\pi i/N}$ for $N = 2^n$ we need at most*

$$4 \cdot 2^n \cdot n = 4N \log_2(N) = O(N \log N)$$

operations to calculate all Fourier coefficients of F.

Proof. Let $\#(M)$ be the minimum number of operations needed to calculate all Fourier coefficients on $\mathbb{Z}_M$. We claim that

$$\#(2M) \leq 2\#(M) + 8M$$

provided $\omega_{2M} = e^{-2\pi i/(2M)}$ is given.
Using the claim, the rest follows by induction. For $N = 2^1 = 2$, we need less than 8 operations to calculate

$$a_0^N(F) = 1/2(F(1) + F(-1)), \quad a_1^N(F) = 1/2(F(1) - F(-1)).$$

If the statement is true up to $N = 2^{n-1}$, we get

$$\#(2N) \leq 2 \cdot 4 \cdot 2^{n-1}(n-1) + 8 \cdot 2^{n-1} = 8n2^{n-1} = 4n2^n.$$

Proof of the claim:
We need at most $2M$ operations to get $\omega_{2M}^2, \ldots, \omega_{2M}^{2M-1}$. Furthermore, for F defined on $\mathbb{Z}_{2M}$, we consider F_0 and F_1 defined on $\mathbb{Z}_M$, which are given by $F_0(r) = F(2r)$ and $F_1(r) = F(2r+1)$. We also assume, that we were able to calculate their Fourier coefficients (in $\mathbb{Z}_M$) in $\#(M)$ operations.

The claim is then proven by the following calculation $(0 \leq k \leq 2M-1)$[12]

$$\begin{aligned}
a_k^{2M}(F) &= \frac{1}{2M}\sum_{r=0}^{2M-1} F(r)\omega_{2M}^{kr} = \frac{1}{2}\left(\frac{1}{M}\sum_{l=0}^{M-1} F(2l)\omega_{2M}^{k(2l)} + \frac{1}{M}\sum_{m=0}^{M-1} F(2m+1)\omega_{2M}^{k(2m+1)}\right) \\
&= \frac{1}{2}\left(\frac{1}{M}\sum_{l=0}^{M-1} F_0(l)\omega_M^{kl} + \frac{1}{M}\sum_{m=0}^{M-1} F_1(m)\omega_M^{km}\omega_{2M}^{k}\right) \\
&= \frac{1}{2}\left(a_k^M(F_0) + a_k^M(F_1)\omega_{2M}^k\right).
\end{aligned}$$

□

14.4.4 Uncertainty principle

The *uncertainty principle* in its most simple form is the following inequality (sometimes called *Heisenberg-Pauli-Weyl inequality*).

Theorem 14.4.5. *If $f \in L_2(\mathbb{R})$ and $a, b \in \mathbb{R}$ are arbitrary, then*

$$\left(\int_{-\infty}^{\infty} (x-a)^2 |f(x)|^2 dx\right)^{1/2} \cdot \left(\int_{-\infty}^{\infty} (\xi - b)^2 |\mathcal{F}f(\xi)|^2 d\xi\right)^{1/2} \geq \frac{1}{2}\|f\|_2^2. \tag{14.8}$$

Let us give first a simple proof for $f \in \mathscr{S}(\mathbb{R})$. By substitution (i.e. by considering $e^{-i\langle x,b\rangle} f(x-a)$ instead of f), we may restrict ourselves to $a = b = 0$. Furthermore, we suppose that $\|f\|_2 = 1$. Then we get

$$1 = \int_{-\infty}^{\infty} |f(x)|^2 dx = -\int_{-\infty}^{\infty} x \frac{d}{dx}|f(x)|^2 dx = -\int_{-\infty}^{\infty} \left(xf'(x)\overline{f(x)} + x\overline{f'(x)}f(x)\right) dx,$$

[12] Let us observe, that $a_k^M(F_0) = a_{k-M}^M(F_0)$ if $k \geq M$.

where we have used partial integration and the identity $|f|^2 = f\overline{f}$. Hence

$$1 \le 2\int_{-\infty}^{\infty} |x|\cdot|f(x)|\cdot|f'(x)|dx \le 2\left(\int_{-\infty}^{\infty} x^2|f(x)|^2dx\right)^{1/2}\left(\int_{-\infty}^{\infty}|f'(x)|^2dx\right)^{1/2}.$$

The proof is then finished by Parseval's identity:

$$\|f'\|_2 = \|\mathcal{F}(f')\|_2 = \|\xi\cdot\mathcal{F}f(\xi)\|_2.$$

Theorem 14.4.6. *Let A and B be (possibly unbounded) self-adjoint operators on a Hilbert space H. Then*

$$\|(A-aI)f\|\cdot\|(B-bI)f\| \ge \frac{1}{2}|\langle[A,B]f,f\rangle|$$

for all $a,b\in\mathbb{R}$ and all f in the domain of AB and BA. Furthermore, $[A,B]=AB-BA$ is the commutator *of A and B.*

Proof.

$$\begin{aligned}\langle[A,B]f,f\rangle &= \langle\{(A-aI)(B-bI)-(B-bI)(A-aI)\}f,f\rangle\\ &= \langle(B-bI)f,(A-aI)f\rangle - \langle(A-aI)f,(B-bI)f\rangle\\ &= 2i\,\mathrm{Im}\langle(B-bI)f,(A-aI)f\rangle\end{aligned}$$

and the statement follows by applying the Cauchy-Schwartz inequality. □

The proof of Theorem 14.4.5 then follows by choosing

$$Xf(x) = xf(x), \quad Pf(x) = if'(x).$$

If f is the domain of PX and XP (and let us observe, that the left-hand side of (14.8) is infinity if this is not the case), we obtain

$$[X,P]f(x) = ixf'(x) - i(xf(x))' = -if(x)$$

and

$$\frac{\|f\|_2^2}{2} = \frac{1}{2}|\langle -if(x), f(x)\rangle| \le \|(X-aI)f\|_2\cdot\|(P-bI)f\|_2,$$

where

$$\|(P-bI)f\|_2 = \|\mathcal{F}(P-bI)f\|_2 = \|(\xi-b)\mathcal{F}f(\xi)\|_2$$

14.5 More on convolutions

When defining the convolution of a distribution and a test function, we have two possibilities. The first one is based on the calculation (for T_f a regular distribution and $\varphi,\psi\in\mathscr{S}(\mathbb{R}^n)$)

$$\begin{aligned}(T_f*\varphi)(\psi) &= \int_{\mathbb{R}^n}(f*\varphi)(x)\psi(x)dx = \int_{\mathbb{R}^n}\psi(x)\int_{\mathbb{R}^n}f(y)\varphi(x-y)dydx\\ &= \int_{\mathbb{R}^n}f(y)\int_{\mathbb{R}^n}\psi(x)\varphi(x-y)dxdy = T_f\left(\int_{\mathbb{R}^n}\psi(x)\varphi(x-y)dx\right),\end{aligned}$$

the other on the observation that

$$(f*\varphi)(x) = \int_{\mathbb{R}^n}f(y)\varphi(x-y)dy = T_f(\varphi(x-\cdot)) = T_f(\tau_x\tilde{\varphi}), \quad x\in\mathbb{R}^n.$$

We shall use the first formula as the definition and afterwards, we show that it coincides with the second one.

Definition 14.5.1. Let $T \in \mathscr{S}'(\mathbb{R}^n)$ and $\varphi \in \mathscr{S}(\mathbb{R}^n)$. Then we define

$$(\varphi * T)(\psi) = (T * \varphi)(\psi) = T(\tilde{\varphi} * \psi), \quad \psi \in \mathscr{S}(\mathbb{R}^n),$$

where $\tilde{\varphi}(x) = \varphi(-x)$ is the *reflexion* of φ.

Example 14.5.2. Let $T = \delta_{x_0}$. Then $\varphi * T$ is the function $x \to \varphi(x - x_0)$. This follows quickly from

$$(\varphi * T)(\psi) = \delta_{x_0}(\tilde{\varphi} * \psi) = (\tilde{\varphi} * \psi)(x_0) = \int_{\mathbb{R}^n} \psi(x)\varphi(x - x_0)dx.$$

It follows, that convolution with δ_0 is the identity operator.

Due to the impossibility of multiplication inside the space of distributions and due to the properties of the Fourier transform (see below), impossibility of convolutions of general distributions follows.

Proposition 14.5.3. *Let $T \in \mathscr{S}'(\mathbb{R}^n)$ and $\varphi \in \mathscr{S}(\mathbb{R}^n)$. Then $T * \varphi$ is a regular distribution and*

$$(T * \varphi)(x) = T(\tau_x \tilde{\varphi}), \quad x \in \mathbb{R}^n.$$

Proof. We have to show that

$$(T * \varphi)(\psi) = T(\tilde{\varphi} * \psi) = T\left(\int_{\mathbb{R}^n} \tilde{\varphi}(\cdot - y)\psi(y)dy\right) = T\left(\int_{\mathbb{R}^n} (\tau_y \tilde{\varphi})(\cdot)\psi(y)dy\right)$$

is equal to

$$\int_{\mathbb{R}^n} T(\tau_y \tilde{\varphi}(\cdot))\psi(y)dy,$$

i.e. that we may interchange the integration and the application of T. To do this, it is enough to prove the convergence of the Riemann sums of the integral under discussion in $\mathscr{S}(\mathbb{R}^n)$. We sketch the arguments but leave out the details.

For each $N \in \mathbb{N}$, we partition $[-N, N]^n$ into $(2N^2)^n$ cubes Q_m, $m = 1, \ldots, (2N^2)^n$ with centers y_m. One has to show that the functions

$$x \to \frac{1}{N^n} \sum_m (\tau_{y_m} \tilde{\varphi})(x)\psi(y_m)$$

converge in $\mathscr{S}(\mathbb{R}^n)$ to

$$x \to \int_{\mathbb{R}^n} (\tau_y \tilde{\varphi})(x)\psi(y)dy$$

as $N \to \infty$. Although not difficult, we leave out the technical details.

□

Definition 14.5.4. Let $T \in \mathscr{S}'(\mathbb{R}^n)$. Then the *support* of T is the intersection of all closed sets $K \subset \mathbb{R}^n$, such that $T(\varphi) = 0$ for all $\varphi \in \mathscr{S}(\mathbb{R}^n)$ with $\operatorname{supp} \varphi \subset \mathbb{R}^n \setminus K$.

The support of the distribution δ_{x_0} is exactly the set $\{x_0\}$.

Lemma 14.5.5. *Let $T \in \mathscr{S}'(\mathbb{R}^n)$. Then T has compact support if, and only if, there exists a constant $C > 0$ and $l, N \in \mathbb{N}$ such that*

$$|T(\varphi)| \leq C \sup_{|x| \leq N} \sum_{|\alpha| \leq l} |D^\alpha \varphi(x)|, \quad \varphi \in \mathscr{S}(\mathbb{R}^n).$$

Proof. Let the estimate be satisfied and let $\psi \in \mathscr{S}(\mathbb{R}^n)$ with $\operatorname{supp}\psi \subset \mathbb{R}^n \setminus \{x \in \mathbb{R}^n : |x| \le N\} = \{x \in \mathbb{R}^n : |x| > N\}$. Then obviously $|T(\psi)| = 0$ and we conclude that T has compact support.

Let on the other hand T have compact support, i.e. $\operatorname{supp} T \subset \{x :\in \mathbb{R}^n : |x| \le M\}$. We consider an infinitely differentiable function ψ with $\psi(x) = 1$ for $|x| \le M$ and $\psi(x) = 0$ for $|x| \ge M+1$. Using Theorem 14.2.3, we get

$$\begin{aligned}
|T(\varphi)| &= |T(\varphi\psi) + T(\varphi(1-\psi))| = |T(\varphi\psi)| \\
&\le c\,\|\varphi\psi\|_{(k,l)} = c \sup_{x\in\mathbb{R}^n} (1+|x|^2)^{k/2} \sum_{|\alpha|\le l} |D^\alpha(\varphi\psi)(x)| \\
&= c \sup_{|x|\le M+1} (1+|x|^2)^{k/2} \sum_{|\alpha|\le l} |D^\alpha(\varphi\psi)(x)| \\
&\le c' \sup_{|x|\le M+1} \sum_{|\alpha|\le l} |D^\alpha \varphi(x)|
\end{aligned}$$

and we obtain the statement with $N = M+1$. □

Theorem 14.5.6. *Let $T \in \mathscr{S}'(\mathbb{R}^n)$ and $\varphi \in \mathscr{S}(\mathbb{R}^n)$. Then*

i) $T * \varphi$ is a C^∞-function,

ii) $D^\alpha(T*\varphi) = (D^\alpha T) * \varphi = T * (D^\alpha \varphi)$ for all $\alpha \in \mathbb{N}_0^n$.

iii) For all multiindices $\alpha \in \mathbb{N}_0^n$ there exist constants $C_\alpha, k_\alpha > 0$, such that

$$|D^\alpha(T*\varphi)(x)| \le C_\alpha (1+|x|^2)^{k_\alpha/2}, \quad x \in \mathbb{R}^n. \tag{14.9}$$

iv) If T has compact support, then $T * \varphi$ is in $\mathscr{S}(\mathbb{R}^n)$.

Proof. We already know that $T*\varphi$ is a regular distribution given pointwise by $(T*\varphi)(x) := T(\tau_x\tilde{\varphi})$, $x \in \mathbb{R}^n$. Next, we show that this is a C^∞-function. Let $e_j = (0,\dots,0,1,0,\dots,0)$ be the canonical unit vector in $\mathbb{R}^n$ with jth coordinate equal to 1 and zeros otherwise. We shall use (cf. Exercises) that the test functions

$$\frac{\tau_{he_j}\psi - \psi}{-h}$$

converge to $D^{e_j}\psi$ in $\mathscr{S}(\mathbb{R}^n)$ for all $\psi \in \mathscr{S}(\mathbb{R}^n)$ as $h \to 0$. This gives for $\psi = \tau_x\tilde{\varphi}$

$$\frac{\tau_{x+he_j}\tilde{\varphi} - \tau_x\tilde{\varphi}}{h} \xrightarrow{\mathscr{S}} (-1)\cdot D^{e_j}(\tau_x\tilde{\varphi}) = (-1)\cdot \tau_x(D^{e_j}\tilde{\varphi})$$

and, after applying T,

$$\frac{(T*\varphi)(x+he_j) - (T*\varphi)(x)}{h} = T\left(\frac{\tau_{x+he_j}\tilde{\varphi} - \tau_x\tilde{\varphi}}{h}\right) \to T(-\tau_x(D^{e_j}\tilde{\varphi})).$$

We combine this with $T*(D^{e_j}\varphi)(x) = T(\tau_x\widetilde{(D^{e_j}\varphi)}) = T(-\tau_x(D^{e_j}\tilde{\varphi})) = T(-D^{e_j}(\tau_x\tilde{\varphi})) = (D^{e_j}T)*\varphi(x)$ and finish the proof of (ii) for $\alpha = e_j$. Iterating this identity, we obtain that $T*\varphi$ is infinitely differentiable and that (ii) holds for every $\alpha \in \mathbb{N}_0^n$.

The proof of (14.9) is based on Theorem 14.2.3. We obtain

$$
\begin{aligned}
|D^\alpha(T*\varphi)(x)| &= |T*(D^\alpha\varphi)(x)| = |T(\tau_x(\widetilde{D^\alpha\varphi}))| \\
&\le c\,\|\tau_x(\widetilde{D^\alpha\varphi})\|_{(k,l)} \le c \sup_{y\in\mathbb{R}^n}(1+|y|^2)^{k/2}\sum_{|\beta|\le l}|D^\beta\tau_x(\widetilde{D^\alpha\varphi})(y)| \\
&\le c \sup_{y\in\mathbb{R}^n}(1+|y|^2)^{k/2}\sum_{|\beta|\le l}|\tau_x(D^{\alpha+\beta}\tilde{\varphi})(y)| \\
&= c \sup_{y\in\mathbb{R}^n}(1+|y|^2)^{k/2}\sum_{|\beta|\le l}|(D^{\alpha+\beta}\tilde{\varphi})(x+y)| \\
&\le c \sup_{z\in\mathbb{R}^n}(1+|z-x|^2)^{k/2}\sum_{|\beta|\le l}|(D^{\alpha+\beta}\tilde{\varphi})(z)| \\
&\le c\,(1+|x|^2)^{k/2}\cdot \sup_{z\in\mathbb{R}^n}(1+|z|^2)^{k/2}\sum_{|\beta|\le l}|(D^{\alpha+\beta}\tilde{\varphi})(z)| \le c\,(1+|x|^2)^{k/2}\|\tilde{\varphi}\|_{(k,|\alpha|+l)}.
\end{aligned}
$$

Finally, to show that $T*\varphi\in\mathscr{S}(\mathbb{R}^n)$ for T with compact support, we use Lemma 14.5.5. We get for arbitrary $k\in\mathbb{N}$

$$
\begin{aligned}
|(T*\varphi)(x)| = |T(\varphi(x-\cdot))| &\le C \sup_{|y|\le N}\sum_{|\alpha|\le l}|D^\alpha_y\varphi(x-y)| \\
&\le C'_\varphi \sup_{|y|\le N}(1+|x-y|^2)^{-k/2} \le C''_\varphi\,(1+|x|^2)^{-k/2}
\end{aligned}
$$

for $|x|\ge 2N$. Hence $(T*\varphi)(x)$ decays faster then any polynomial. The same argument applied to $D^\alpha(T*\varphi)=T*(D^\alpha\varphi)$ yields the same for all the derivatives of $T*\varphi$. Hence $T*\varphi\in\mathscr{S}(\mathbb{R}^n)$. □

Finally, we show that products and convolutions of distributions behave under Fourier transform exactly as we expect them to do.

Theorem 14.5.7. *Let $T\in\mathscr{S}'(\mathbb{R}^n)$ and $\varphi\in\mathscr{S}(\mathbb{R}^n)$. Then*

$$
\begin{aligned}
\mathcal{F}(\varphi T) &= (2\pi)^{-n/2}\mathcal{F}(\varphi)*\mathcal{F}(T), \\
\mathcal{F}(\varphi*T) &= (2\pi)^{n/2}\mathcal{F}(\varphi)\mathcal{F}(T).
\end{aligned}
$$

Proof. We obtain for every $\psi\in\mathscr{S}(\mathbb{R}^n)$

$$
\mathcal{F}(\varphi T)(\psi) = (\varphi T)(\mathcal{F}\psi) = T(\varphi\mathcal{F}(\psi))
$$

and

$$
\mathcal{F}(\varphi)*\mathcal{F}(T)(\psi) = \mathcal{F}T(\widetilde{\mathcal{F}\varphi}*\psi) = T(\mathcal{F}(\widetilde{\mathcal{F}\varphi}*\psi)) = (2\pi)^{n/2}T(\varphi\mathcal{F}(\psi)).
$$

The other identity follows in the same manner. □

14.6 Paley-Wiener Theorem

Paley-Wiener's Theorem is in general any statement connecting the decay of a function or a distribution with the smoothness of its Fourier transform (and vice versa). Although proven by Paley and Wiener for square-integrable functions first, it was adapted to test functions and distributions by Schwartz (and Hörmander). We give one important example of such a theorem below.

Another remarkable fact about Paley-Wiener theorems is that they connect the theory of distributions with the theory of functions of several complex variables. Therefore, we first (re-)introduce some notation from complex analysis.

Definition 14.6.1. Let $\Omega \subset \mathbb{C}^n$ be an open set and let f be a complex-valued function on Ω. Then f is said to be *holomorphic in* Ω, if it is holomorphic in each variable separately. A function f of n complex variables is said to be *entire* if it is holomorphic on whole $\mathbb{C}^n$.

For any $z = (z_1, \ldots, z_n) \in \mathbb{C}^n$, we shall write $x = \operatorname{Re} z = (\operatorname{Re} z_1, \ldots, \operatorname{Re} z_n) \in \mathbb{R}^n$ and $y = \operatorname{Im} z \in \mathbb{R}^n$ for real and imaginary part of z.

Let T be a tempered distribution on $\mathbb{R}^n$, i.e. $T \in \mathscr{S}'(\mathbb{R}^n)$, with compact support. Then the Fourier transform of T was defined in Definition 14.2.5. However, the definition $\mathcal{F}f(\xi) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} f(x) e^{-i\langle x,\xi\rangle} dx$, $f \in \mathscr{S}(\mathbb{R}^n)$, suggests that we could define also

$$(\mathcal{F}T)(\xi) := T\left(\frac{e^{-i\langle\cdot,\xi\rangle}}{(2\pi)^{n/2}}\right), \tag{14.10}$$

whenever the right-hand side makes sense. As $e^{-i\langle x,\cdot\rangle}$ is not in $\mathscr{S}(\mathbb{R}^n)$, some care is of course necessary. If T has compact support, then $T(e^{-i\langle x,\cdot\rangle})$ might be defined as $T(e^{-i\langle x,\cdot\rangle}\varphi(\cdot))$ for any $\varphi \in \mathscr{S}(\mathbb{R}^n)$ with $\varphi = 1$ on $\operatorname{supp} T$. Obviously, this "definition" does not depend on the choice of φ, i.e. $T(e^{-i\langle x,\cdot\rangle}\varphi(\cdot)) = T(e^{-i\langle x,\cdot\rangle}\psi(\cdot))$ if both $\varphi, \psi \in \mathscr{S}(\mathbb{R}^n)$ with $\varphi = \psi = 1$ on $\operatorname{supp} T$.

First, using $\psi \in \mathscr{S}(\mathbb{R}^n)$ with $\psi(x) = 1$ for $x \in \operatorname{supp} T$, we get immediately that

$$\mathcal{F}T = \mathcal{F}(\psi T) = (2\pi)^{-n/2} \mathcal{F}\psi * \mathcal{F}T$$

is a C^∞-function (cf. Theorem 14.5.6). Furthermore, with $\Phi \in \mathscr{S}(\mathbb{R}^n)$ with $\mathcal{F}\Phi = \psi$, we get

$$\begin{aligned}(2\pi)^{n/2}\mathcal{F}T(\xi) &= (\mathcal{F}T * \mathcal{F}\psi)(\xi) = (\mathcal{F}T * \tilde{\Phi})(\xi) = \mathcal{F}T(\tau_\xi \Phi) \\ &= T(\mathcal{F}(\tau_\xi \Phi)) = T(e^{-i\langle\cdot,\xi\rangle}\psi),\end{aligned}$$

which we defined as $T(e^{-i\langle\cdot,\xi\rangle})$. Therefore, also (14.10) is justified.

Let us observe that (14.10) makes sense also for $z \in \mathbb{C}^n$ instead of $\xi \in \mathbb{R}^n$. This mapping (which is an extension of Fourier transform to complex arguments) is usually called *Fourier-Laplace transform* and is defined by

$$(\mathcal{F}T)(z) := T\left(\frac{e^{-i\langle\cdot,z\rangle}}{(2\pi)^{n/2}}\right), \quad z \in \mathbb{C}^n. \tag{14.11}$$

Note, that we denote it by $\mathcal{F}$ again.

Its properties are given in the following version of (Schwartz-)Paley-Wiener's theorem.

Theorem 14.6.2. *Let $T \in \mathscr{S}'(\mathbb{R}^n)$ have compact support. Then its Fourier-Laplace transform is an entire function with polynomial growth on $\mathbb{R}^n$.*

Proof. First, we observe that

$$D^{e_j}(\mathcal{F}T)(z) = \lim_{h\to 0} \frac{\mathcal{F}T(z + he_j) - \mathcal{F}T(z)}{h} = \frac{1}{(2\pi)^{n/2}} \lim_{h\to 0} T\left(\frac{e^{-i\langle x,z+he_j\rangle} - e^{-i\langle x,z\rangle}}{h}\right).$$

The fact, that $\mathcal{F}T$ is an entire function, now follows from the compact support of T and the convergence of the sequence

$$x \to \frac{e^{-i\langle x,z+he_j\rangle} - e^{-i\langle x,z\rangle}}{h}, \quad x \in \mathbb{R}^n,$$

to $-ix_j e^{-i\langle x,z\rangle}$ in $C^\infty(\mathbb{R}^n)$. This gives $D^{e_j}(\mathcal{F}T)(z) = T(-ix_j e^{-i\langle x,z\rangle})$. This argument might be iterated and we obtain $D^\alpha(\mathcal{F}T)(z) = T((-ix_j)^\alpha e^{-i\langle x,z\rangle})$ for all $\alpha \in \mathbb{N}_0^n$. Moreover, the polynomial growth on $\mathbb{R}^n$ follows from (14.9). $\square$

14.7 Fundamental solution

Theorem 14.7.1. *Let $T \in \mathscr{S}'(\mathbb{R}^n)$ be supported at the singleton $\{x_0\}$. Then there exists a number $k \in \mathbb{N}_0$ and complex numbers a_α for $|\alpha| \le k$, such that*

$$T = \sum_{|\alpha|\le k} a_\alpha D^\alpha \delta_{x_0}.$$

Proof. We may assume, that $x_0 = 0$. We shall use Lemma 14.5.5. We assume that

$$|T(\varphi)| \le C \sup_{|x|\le N} \sum_{|\alpha|\le l} |D^\alpha \varphi(x)|$$

for all $\varphi \in \mathscr{S}(\mathbb{R}^n)$. First we show, that if $\varphi \in \mathscr{S}(\mathbb{R}^n)$ satisfies

$$(D^\alpha \varphi)(0) = 0, \quad |\alpha| \le l, \tag{14.12}$$

then $T\varphi = 0$. We fix one such a $\varphi \in \mathscr{S}(\mathbb{R}^n)$ and we chose $\psi \in C^\infty(\mathbb{R}^n)$ with $\psi = 0$ on $\{x \in \mathbb{R}^n : |x| \le 1\}$ and $\psi = 1$ on $\{x \in \mathbb{R}^n : |x| \ge 2\}$. We put $\psi_\varepsilon(x) := \psi(x/\varepsilon)$ for $\varepsilon > 0$.

$$\begin{aligned}|T(\varphi)| &= |T(\varphi(\psi_\varepsilon + (1-\psi_\varepsilon)))| \le |T(\varphi\psi_\varepsilon)| + |T(\varphi(1-\psi_\varepsilon))|\\ &\le 0 + C \sup_{|x|\le N} \sum_{|\alpha|\le l} |D^\alpha(\varphi(1-\psi_\varepsilon))(x)|\\ &\le C \sup_{|x|\le 2\varepsilon} \sum_{|\alpha|\le l} |D^\alpha(\varphi(1-\psi_\varepsilon))(x)|\end{aligned}$$

and it is rather easy to see (using (14.12)), that the last expression tends to zero as $\varepsilon \to 0$.

Now, let $\varphi \in \mathscr{S}(\mathbb{R}^n)$ be arbitrary and let $\eta \in \mathscr{S}(\mathbb{R}^n)$ be a function that is equal to 1 in a neighborhood of origin. Then (using Taylor's polynomial)

$$\varphi(x) = \eta(x)\left(\sum_{|\alpha|\le l} \frac{D^\alpha\varphi(0)}{\alpha!}x^\alpha + \psi(x)\right) + (1-\eta(x))\varphi(x),$$

where $\psi \in \mathscr{S}(\mathbb{R}^n)$ satisfies $\psi(x) = O(|x|^{l+1})$ as $|x| \to 0$.

Now $T(\eta\psi) = 0$ due to the first part of the proof and $T((1-\eta)\varphi) = 0$ due to the support property of T. Hence

$$T(\varphi) = \sum_{|\alpha|\le l} \frac{D^\alpha\varphi(0)}{\alpha!} T(\eta(x)x^\alpha) = \sum_{|\alpha|\le l} a_\alpha (D^\alpha\delta_0)(\varphi)$$

for $a_\alpha = (-1)^{|\alpha|} T(\eta(x)x^\alpha)/\alpha!$. □

Corollary 14.7.2. *Let $T \in \mathscr{S}'(\mathbb{R}^n)$ with $\mathcal{F}T$ supported at the singleton $\{\xi_0\}$. Then T is a linear combination of functions $\xi^\alpha e^{i\langle\xi,\xi_0\rangle}$. In particular, if $\mathcal{F}T$ is supported at the origin, then T is a polynomial.*

The *Laplacian* is the partial differential operator

$$\Delta(T) = \sum_{j=1}^n \frac{\partial^2 T}{\partial x_j^2}, \quad T \in \mathscr{S}'(\mathbb{R}^n).$$

Solutions of the *Laplace equation* $\Delta(T) = 0$ are called *harmonic* distributions.

Corollary 14.7.3. *Let $T \in \mathscr{S}'(\mathbb{R}^n)$ be a harmonic distribution. Then T is a polynomial.*

Proof. Taking the Fourier transform, we observe that $|\xi|^2 \cdot \mathcal{F}T = 0$. Hence, $\mathcal{F}T$ is supported at the origin and, therefore, must be a polynomial. $\square$

Liouville's classical theorem says that every bounded harmonic function must be constant. We observe that it follows directly from Corollary 14.7.3.

A distribution $T \in \mathscr{S}'(\mathbb{R}^n)$ is called a *fundamental solution* of a partial differential operator with constant coefficients L if $L(T) = \delta_0$.

Theorem 14.7.4. *For $n \geq 3$, we have*

$$\Delta(|x|^{2-n}) = -(n-2)\omega_{n-1}\delta_0 = -(n-2)\frac{2\pi^{n/2}}{\Gamma(n/2)}\delta_0$$

and for $n = 2$,

$$\Delta(\log|x|) = 2\pi\delta_0.$$

Here, ω_{n-1} stands for the $(n-1)$-dimensional measure of the unit sphere in $\mathbb{R}^n$.

Proof. We use the Green's identity

$$\int_\Omega v\Delta u - u\Delta v dx = \int_{\partial\Omega}\left(v\frac{\partial u}{\partial\nu} - u\frac{\partial v}{\partial\nu}\right) ds,$$

where Ω is a domain in $\mathbb{R}^n$ with smooth boundary and $\partial/\partial\nu$ denotes the derivative with respect to the outer normal vector to $\partial\Omega$.

Let $\Omega_\varepsilon = \{x \in \mathbb{R}^n : \varepsilon < |x| < 1/\varepsilon\}$. Furthermore, we choose (for $n \geq 3$) $v(x) = |x|^{2-n}$ and $u = \varphi \in \mathscr{S}(\mathbb{R}^n)$. This leads to

$$\Delta(|x|^{2-n})(\varphi) = v(\Delta\varphi) = \int_{\mathbb{R}^n} v(x)\Delta\varphi(x)dx = \lim_{\varepsilon\to 0}\int_{\Omega_\varepsilon} |x|^{2-n}\Delta\varphi(x)dx$$

and

$$\int_{\Omega_\varepsilon} |x|^{2-n}\Delta\varphi(x)dx = \int_{\Omega_\varepsilon} \Delta(|x|^{2-n})\varphi(x)dx - \int_{\partial\Omega_\varepsilon}\left(\varphi(x)\frac{\partial|x|^{2-n}}{\partial\nu} - |x|^{2-n}\frac{\partial\varphi(x)}{\partial\nu}\right) ds.$$

The first integral is equal to zero as $\Delta(|x|^{2-n}) = 0$ for $x \neq 0$.

The second integral splits into two parts, integral over $\{x \in \mathbb{R}^n : |x| = \varepsilon\}$ and $\{x \in \mathbb{R}^n : |x| = 1/\varepsilon\}$. We observe quickly that

$$\int_{\{x\in\mathbb{R}^n:|x|=1/\varepsilon\}}\left(\varphi(x)\frac{\partial|x|^{2-n}}{\partial\nu} - |x|^{2-n}\frac{\partial\varphi(x)}{\partial\nu}\right) ds \to 0$$

as $\varepsilon \to 0$. This is due to

$$\begin{aligned}\int_{\{x\in\mathbb{R}^n:|x|=1/\varepsilon\}}&\left(\varphi(x)\frac{\partial|x|^{2-n}}{\partial\nu} - |x|^{2-n}\frac{\partial\varphi(x)}{\partial\nu}\right) ds\\ &= \int_{\{x\in\mathbb{R}^n:|x|=1/\varepsilon\}}\left(\varphi(x)(2-n)\varepsilon^{n-1} - \varepsilon^{n-2}\frac{\partial\varphi(x)}{\partial\nu}\right) ds\end{aligned}$$

and the rapid decay of $\varphi \in \mathscr{S}(\mathbb{R}^n)$ and all its derivatives at infinity.

As for the integral over $\{x \in \mathbb{R}^n : |x| = \varepsilon\}$, it is equal to

$$(2-n)\varepsilon^{1-n}\int_{\{x\in\mathbb{R}^n:|x|=\varepsilon\}}\varphi(x)ds+\varepsilon^{2-n}\int_{\{x\in\mathbb{R}^n:|x|=\varepsilon\}}\frac{\partial\varphi(x)}{\partial\nu}ds.$$

Finally, the first integral tends to $(2-n)\varphi(0)\omega_{n-1}$, where ω_{n-1} is the $(n-1)$-dimensional measure of the unit sphere in $\mathbb{R}^n$, and the second integral tends to zero.

Let us note, that the formula $\omega_{n-1} = 2\pi^{n/2}/\Gamma(n/2)$ follows easily by the following trick. Using polar coordinates in $\mathbb{R}^n$ and the substitution $t := r^2$, we get

$$(\sqrt{\pi})^n = \int_{\mathbb{R}^n} e^{-|x|^2}dx = \omega_{n-1}\int_0^\infty e^{-r^2}r^{n-1}dr = \frac{\omega_{n-1}}{2}\int_0^\infty e^{-t}t^{n/2-1}dt = \frac{\omega_{n-1}}{2}\Gamma(n/2).$$

The proof for $n = 2$ is very similar. $\square$

Remark 14.7.5. Due to the previous Corollary, the fundamental solution of Laplacian is not unique. We may add to any of the fundamental solutions above a harmonic polynomial and the Laplace operator of such a distribution would not change.

Although one could develop a full theory of fundamental solutions in the spirit of the book of Rudin (with Theorem of Malgrange-Ehrenpreis ensuring the existence of a fundamental solution for any partial differential operator with constant coefficients as the main highlight), we shall give a couple of examples of this technique instead.

Let us assume that $L(T) = \delta_0$, i.e. that T is a fundamental solution to L. Then the solution of the equation $L(u) = f$ can be obtained by convolution.

$$L(T * f) = (LT) * f = \delta_0 * f = f.$$

Of course, the interpretation of $T * f$ can cause problems if $T \in \mathscr{S}'(\mathbb{R}^n)$ and f is not smooth enough, i.e. $f \notin \mathscr{S}(\mathbb{R}^n)$.

Remark 14.7.6. Using previous theorem and the above presented arguments, one may immediately write down the solution (or, better said, all the solutions) to the equation $\Delta T = \varphi$, at least for $\varphi \in \mathscr{S}(\mathbb{R}^n)$. The obvious disadvantage is of course, that the usual physical problems are defined only on some *domain* of $\mathbb{R}^n$. We shall deal with this problem later on when considering distributions on domains.

We shall now give another application of the technique of fundamental solution. Let us consider the *heat equation*:

$$\begin{aligned} u_t(x,t) - ku_{xx}(x,t) = 0, \quad x \in \mathbb{R},\ 0 < t < \infty, \\ u(x,0) = g(x). \end{aligned}$$

The fundamental solution of these equation will be obtained by choosing $g = \delta_0$.

Let us sketch the way to the fundamental solution. We apply the Fourier transform to this equations in x variable (i.e. for every t fixed). This gives

$$\mathcal{F}_x[u_t(x,t) - ku_{xx}(x,t)] = 0, \quad \mathcal{F}_x[u(x,0)] = \mathcal{F}_x[\delta_0],$$

i.e.

$$v_t(\xi,t) + k\xi^2 v(\xi,t) = 0, \quad v(\xi,0) = (2\pi)^{-1/2},$$

where now $v(\xi,t) = \mathcal{F}_x(u(x,t))(\xi)$. We solve these equations for each $\xi \in \mathbb{R}$ separately. We fix $\xi \in \mathbb{R}$ and put $z(t) = v(\xi,t)$ and obtain $z'(t) + k\xi^2 z(t) = 0$ and $z(0) = (2\pi)^{-1/2}$. This leads to $z(t) = (2\pi)^{-1/2}e^{-k\xi^2 t}$ and $v(\xi,t) = (2\pi)^{-1/2}e^{-k\xi^2 t}$. If we denote $h(s) = e^{-s^2/2}$, i.e. $\mathcal{F}h = h$, we get $v(\xi,t) = (2\pi)^{-1/2}h(\sqrt{2kt}\xi)$ and finally the fundamental solution to the one-dimensional heat equation $\Phi(x,t) = (\mathcal{F}_\xi^{-1}v(\xi,t))(x) = (2\pi)^{-1/2}(2kt)^{-1/2}h(x/\sqrt{2kt}) = \dfrac{1}{\sqrt{4\pi kt}}\exp\left(-\dfrac{x^2}{4kt}\right)$.

We now claim that the solution to the heat equation with a general right hand side g is then given as the convolution of Φ with g in the x variable, i.e.

$$u(x,t) = \int_{\mathbb{R}} \Phi(x-y,t)g(y)dy, \quad -\infty < x < \infty,\ 0 < t < \infty.$$

Indeed, at least formally, we obtain

$$u_t(x,t) - ku_{xx}(x,t) = \int_{\mathbb{R}} \left[\Phi_t(x-y,t) - k\Phi_{xx}(x-y,t)\right] g(y)dy = 0$$

and

$$u(x,0) = \int_{\mathbb{R}} \Phi(x-y,0)g(y)dy = \int_{\mathbb{R}} \delta_0(x-y)g(y)dy = g(x).$$

Finally, let us mention that the fundamental solution to the n-dimensional heat equation is obtained simply as a tensor product of n one-dimensional fundamental solutions, i.e.

$$\Phi(x,t) = \frac{1}{(4\pi kt)^{n/2}}\exp\left(-\frac{|x|^2}{4kt}\right), \quad x \in \mathbb{R}^n,\ 0 < t < \infty.$$

14.8 Spaces $\mathscr{D}(\Omega)$ and $\mathscr{D}'(\Omega)$

Unfortunately, the spaces $\mathscr{S}(\mathbb{R}^n)$ and $\mathscr{S}'(\mathbb{R}^n)$ together with Fourier transform $\mathcal{F}$ are not well suited to deal with partial differential equations on domains. That is the reason, why there is a way (and historically, it was actually introduced before the Schwartz space and its dual), which allows to deal with distributions on domains.

Definition 14.8.1. Let $\Omega \subset \mathbb{R}^n$ be an open set.

i) We define the space of test functions

$$\mathscr{D}(\Omega) = \{\varphi \in C^\infty(\Omega) :\ \text{supp}\ \varphi \text{ is a compact subset of } \Omega\}.$$

ii) A sequence $(\varphi_j)_{j\in\mathbb{N}} \subset \mathscr{D}(\Omega)$ is said to converge to $\varphi \in \mathscr{D}(\Omega)$ in $\mathscr{D}(\Omega)$, if there is a compact set $K \subset \Omega$, such that

$$\text{supp}\ \varphi_j \subset K, \quad \text{for all } j \in \mathbb{N}$$

and

$$D^\alpha\varphi_j \rightrightarrows D^\alpha\varphi \quad \text{for all } \alpha \in \mathbb{N}_0^n.$$

We write $\varphi_j \xrightarrow{\mathscr{D}} \varphi$ to denote the convergence in $\mathscr{D}(\Omega)$.

iii) $\mathscr{D}'(\Omega)$ is the collection of all complex-valued linear continuous functionals over $\mathscr{D}(\Omega)$, i.e. $T : \mathscr{D}(\Omega) \to \mathbb{C}$ belongs to $\mathscr{D}'(\Omega)$ if, and only if,

$$T(\lambda_1\varphi_1 + \lambda_2\varphi_2) = \lambda_1 T(\varphi_1) + \lambda_2 T(\varphi_2) \text{ for all } \lambda_1,\lambda_2 \in \mathbb{C} \text{ and } \varphi_1,\varphi_2 \in \mathscr{D}(\Omega),$$

$$T(\varphi_j) \to T(\varphi) \text{ whenever } \varphi_j \xrightarrow{\mathscr{D}} \varphi.$$

The elements of $\mathscr{D}'(\Omega)$ are called *distributions*.

iv) The space $\mathscr{D}'(\Omega)$ is turned into a vector space by setting

$$(\lambda_1 T_1 + \lambda_2 T_2)(\varphi) = \lambda_1 T_1(\varphi) + \lambda_2 T_2(\varphi) \text{ for all } \lambda_1, \lambda_2 \in \mathbb{C}, T_1, T_2 \in \mathscr{D}'(\Omega) \text{ and } \varphi \in \mathscr{D}(\Omega).$$

v) Finally, we also equip the space $\mathscr{D}'(\Omega)$ with the notion of convergence. Namely, we say that T_j converges to T in $\mathscr{D}'(\Omega)$, $T_j \overset{\mathscr{D}'}{\rightharpoonup} T$, if

$$T_j(\varphi) \to T(\varphi) \text{ for all } \varphi \in \mathscr{D}(\Omega).$$

The spaces $\mathscr{D}(\Omega)$ are much more flexible due to the inclusion of the domain into their definition. Unfortunately, a lot of the algebraic structure is lost, for example Fourier transforms and convolutions do not have an easy counterpart on $\mathscr{D}(\Omega)$. Also modulations and translations are only of a limited use on Ω.

Nevertheless, one can define the derivative of a distribution $T \in \mathscr{D}'(\Omega)$ by $(D^\alpha T)(\varphi) = (-1)^{|\alpha|} T(D^\alpha \varphi)$ and its product with a smooth function $\psi \in C^\infty(\Omega)$ as $(\psi T)(\varphi) = T(\psi\varphi)$.

If $\Omega = \mathbb{R}^n$, we obtain two spaces of test function, and two spaces of distributions. Obviously, $\mathscr{D}(\mathbb{R}^n) \subset \mathscr{S}(\mathbb{R}^n)$ (including a continuous embedding) and, consequently, $\mathscr{S}'(\mathbb{R}^n) \subset \mathscr{D}'(\mathbb{R}^n)$, again also in the sense of continuous embedding.

The space $\mathscr{D}(\Omega)$ allows very well to explain what is meant by *weak solutions* to PDE's. Let us for example consider the equation

$$\frac{\partial u(t,x)}{\partial t} + \frac{\partial u(t,x)}{\partial x} = 0.$$

In the classical setting this means that we are looking for a function $u \in C^1(\mathbb{R}^2)$, such that the equation holds pointwise. In the distributional sense, we require that

$$\left(\frac{\partial u(t,x)}{\partial t} + \frac{\partial u(t,x)}{\partial x} \right)(\varphi) = 0$$

for all $\varphi \in \mathscr{D}(\mathbb{R}^2)$. But this is equivalent to

$$u\left(\frac{\partial \varphi(t,x)}{\partial t} + \frac{\partial \varphi(t,x)}{\partial x} \right) = \int_{\mathbb{R}^2} u(t,x)\left(\varphi_t(t,x) + \varphi_x(t,x)\right) d(t,x) = 0.$$

Whenever u is smooth enough, these two notions coincide. But there are functions (i.e. $u(t,x) = |t-x|$), which do satisfy the weak formulation, but can not satisfy the strong formulation due to their lack of differentiability.

15 Introduction to harmonic analysis

15.1 Approximation of identity

Theorem 15.1.1. *Let $\Omega \subset \mathbb{R}^n$ be a domain. The set of continuous functions with compact support contained in Ω is dense in $L_p(\Omega)$, $1 \le p < \infty$.*

Proof. We shall need two facts from measure theory.

i) Lebesgue measure λ in $\mathbb{R}^n$ is regular, i.e. $\lambda(A) = \inf\{\lambda(G) : G \supset A, G \,\text{open}\}$.

ii) The space of step functions, i.e. $\operatorname{span}\{\chi_A : A \subset \Omega, A \text{ measurable}\}$, is dense in $L_p(\Omega)$ for every $1 \le p < \infty$.

We first consider open sets $\Omega_j \subset \Omega$, $j \in \mathbb{N}$, such that $\Omega_j \subset \overline{\Omega_j} \subset \Omega_{j+1} \subset \Omega$ and $\bigcup_{j=1}^\infty \Omega_j = \Omega$.[13] Let us take $f \in L_p(\Omega)$. Then $f\chi_{\Omega_j} \to f$ in $L_p(\Omega)$ and we may restrict ourselves to $f \in L_p(\Omega)$ with compact support in Ω. Due to the second property of the Lebesgue measure, this function may be approximated by a step function $\sum_{k=1}^K \varrho_k \chi_{A_k}$ with $A_k \subset \operatorname{supp} f$. So, it is enough to approximate characteristic functions χ_B with $\overline{B}$ compact in Ω. Using the first property of the Lebesgue measure, we may restrict ourselves to bounded open sets $G \subset \overline{G} \subset \Omega$. Then the sequence of functions $x \to \max(0, 1 - k \operatorname{dist}(x, G))$ gives the desired approximation. □

Lemma 15.1.2. *Let $f \in L_p(\mathbb{R}^n), 1 \le p < \infty$. Then $f(\cdot + h) \to f(\cdot)$ in $L_p(\mathbb{R}^n)$ if $h \to 0$.*

Proof. If f is continuous with compact support, then the result follows by uniform continuity of f and the Lebesgue dominated convergence theorem. If $f \in L_p(\mathbb{R}^n)$, we may find for every $t > 0$ a continuous function g with compact support such that $\|f - g\|_p < t$. Then

$$\begin{aligned}\|f(\cdot + h) - f(\cdot)\|_p &\le \|f(\cdot + h) - g(\cdot + h)\|_p + \|g(\cdot + h) - g(\cdot)\|_p + \|g(\cdot) - f(\cdot)\|_p \\ &\le 2t + \|g(\cdot + h) - g(\cdot)\|_p\end{aligned}$$

and the conclusion follows. □

Theorem 15.1.3. *The family of functions $(K_\varepsilon)_{\varepsilon>0} \subset L_1(\mathbb{R}^n)$ is called the* approximation of identity*, if*

(K1) $\int_{\mathbb{R}^n} |K_\varepsilon(x)| dx \le C < \infty$ *for all* $\varepsilon > 0$,

(K2) $\int_{\mathbb{R}^n} K_\varepsilon(x) dx = 1$ *for all* $\varepsilon > 0$,

(K3) $\lim_{\varepsilon \to 0^+} \int_{|x|>\delta} |K_\varepsilon(x)| dx = 0$ *for all* $\delta > 0$.

Then

i) If $K \in L_1(\mathbb{R}^n)$ with $\int_{\mathbb{R}^n} K(x)dx = 1$, then $K_\varepsilon(x) = \varepsilon^{-n} K(x/\varepsilon)$ is an approximation of identity.

ii) If $(K_\varepsilon)_{\varepsilon>0}$ is an approximation of the identity, then

$$\lim_{\varepsilon \to 0^+} \|K_\varepsilon * f - f\|_p = 0$$

for every $1 \le p < \infty$ and $f \in L_p(\mathbb{R}^n)$.

[13] For example the sets $\Omega_j := \{x \in \Omega : |x| < j \text{ and } \operatorname{dist}(x, \partial\Omega) > 1/j\}$ will do.

Proof. (i) Let $K \in L_1(\mathbb{R}^n)$ with $\int_{\mathbb{R}^n} K(x)dx = 1$. Then we get immediately

$$\varepsilon^{-n}\int_{\mathbb{R}^n} K(x/\varepsilon)dx = \int_{\mathbb{R}^n} K(x)dx = 1 \quad \text{and} \quad \varepsilon^{-n}\int_{\mathbb{R}^n} |K(x/\varepsilon)|dx = \int_{\mathbb{R}^n} |K(x)|dx = \|K\|_1 < \infty.$$

As for the third point, we have

$$\int_{|x|>\delta} |K_\varepsilon(x)|dx = \int_{|y|>\delta/\varepsilon} |K(y)|dy \to 0$$

as $\varepsilon \to 0^+$, due to the Lebesgue dominated convergence theorem.

(ii) We calculate for $p > 1$ and its conjugated index p' with $1/p + 1/p' = 1$ using Hölder's inequality (if $p = 1$, the calculation becomes slightly simpler)

$$\begin{aligned}
\|K_\varepsilon * f - f\|_p^p &= \int_{\mathbb{R}^n} |(K_\varepsilon * f)(x) - f(x)|^p dx = \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} K_\varepsilon(y) f(x-y)dy - f(x)\right|^p dx \\
&= \int_{\mathbb{R}^n} \left|\int_{\mathbb{R}^n} K_\varepsilon(y)[f(x-y) - f(x)]dy\right|^p dx \\
&\le \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} |K_\varepsilon(y)|^{1/p+1/p'} \cdot |f(x-y) - f(x)|dy\right)^p dx \\
&\le \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} |K_\varepsilon(y)| \cdot |f(x-y) - f(x)|^p dy \cdot \left(\int_{\mathbb{R}^n} |K_\varepsilon(y)|dy\right)^{p/p'} dx \\
&\le C^{p/p'} \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} |K_\varepsilon(y)| \cdot |f(x-y) - f(x)|^p dydx \\
&= C^{p/p'} \int_{\mathbb{R}^n} |K_\varepsilon(y)| \cdot \|f(\cdot - y) - f(\cdot)\|_p^p dy \\
&\le C^{p/p'} \left\{ \int_{|y|\le\delta} |K_\varepsilon(y)| \cdot \|f(\cdot - y) - f(\cdot)\|_p^p dy + 2^p \|f\|_p^p \int_{|y|>\delta} |K_\varepsilon(y)|dy \right\}
\end{aligned}$$

for every $\delta > 0$. Using $(K3)$ and previous Lemma, we obtain the conclusion of the theorem. □

Definition 15.1.4. Let $\Omega \subset \mathbb{R}^n$ be a domain. Then $C_c^\infty(\Omega)$ denotes the set of infinitely-differentiable functions compactly supported in Ω.

It was shown in exercises that this class is actually non-empty.

Theorem 15.1.5. *$C_c^\infty(\Omega)$ is dense in $L_p(\Omega)$ for every $1 \le p < \infty$ and every domain $\Omega \subset \mathbb{R}^n$.*

Proof. Let $f \in L_p(\Omega)$. First, we approximate f by a continuous and compactly supported g (i.e. $\|f - g\|_p \le h$) and then consider the functions $\omega_\varepsilon * g$, where $\omega_\varepsilon(x) = \varepsilon^{-n}\omega(x/\varepsilon)$ and $\omega \in C_c^\infty(\mathbb{R}^n)$ has compact support and $\int \omega = 1$. It follows that $\omega_\varepsilon * g \in C_c^\infty(\mathbb{R}^n)$ (the support property is clear, the differentiability was proven in a much more general setting in Theorem 14.5.6). Together with the formula $\|\omega_\varepsilon * g - g\|_p \to 0$, the conclusion follows. □

15.2 Maximal operator

For $x \in \mathbb{R}^n$ and $r > 0$, we denote by $B(x, r)$ the ball in $\mathbb{R}^n$ with center at x and radius r, i.e. $B(x, r) = \{y \in \mathbb{R}^n : |x - y| < r\}$.

Let f be a locally integrable function on $\mathbb{R}^n$. Then we define the *Hardy-Littlewood maximal operator* of f by

$$Mf(x) = \sup_{r>0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y)|dy.$$

Here stands $|B(x,r)|$ for the Lebesgue measure of $B(x,r)$.

Let us note, that maximal operator is not linear, but is sub-linear, i.e.

$$\begin{aligned} M(f+g)(x) &= \sup_{r>0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y)+g(y)|dy \\ &\le \sup_{r>0} \frac{1}{|B(x,r)|} \left(\int_{B(x,r)} |f(y)|dy + \int_{B(x,r)} |g(y)|dy \right) \\ &\le (Mf)(x) + (Mg)(x). \end{aligned}$$

We shall study the mapping properties of the operator M in the frame of Lebesgue spaces $L_p(\mathbb{R}^n)$, $1 \le p \le \infty$. If $f \in L_\infty(\mathbb{R}^n)$, then

$$\frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y)|dy \le \frac{\|f\|_\infty}{|B(x,r)|} \int_{B(x,r)} 1dy = \|f\|_\infty$$

holds for every $x \in \mathbb{R}^n$ and every $r > 0$ and $\|Mf\|_\infty \le \|f\|_\infty$ follows. To deal with other p's, we need some more notation first.

Let $f \in L_1(\mathbb{R}^n)$ and let $\alpha > 0$. Then

$$\begin{aligned} \alpha \cdot |\{x \in \mathbb{R}^n : |f(x)| > \alpha\}| &= \int_{\{x\in\mathbb{R}^n:|f(x)|>\alpha\}} \alpha dy \\ &\le \int_{\{x\in\mathbb{R}^n:|f(x)|>\alpha\}} |f(y)|dy \le \int_{\mathbb{R}^n} |f(y)|dy = \|f\|_1. \end{aligned}$$

The set of measurable functions f on $\mathbb{R}^n$ with

$$\|f\|_{1,w} := \sup_{\alpha>0} \alpha \cdot |\{x \in \mathbb{R}^n : |f(x)| > \alpha\}| < \infty$$

is denoted by $L_{1,w}(\mathbb{R}^n)$ and called *weak* L_1. We have just shown that $L_1(\mathbb{R}^n) \hookrightarrow L_{1,w}(\mathbb{R}^n)$. Let us mention that $\|\cdot\|_{1,w}$ is not a norm, but it still satisfies

$$\begin{aligned} \|f+g\|_{1,w} &= \sup_{\alpha>0} \alpha \cdot |\{x \in \mathbb{R}^n : |f(x)+g(x)| > \alpha\}| \\ &\le \sup_{\alpha>0} \alpha \Big(|\{x \in \mathbb{R}^n : |f(x)| > \alpha/2\}| + |\{x \in \mathbb{R}^n : |g(x)| > \alpha/2\}| \Big) \\ &\le 2 \sup_{\alpha>0} \alpha/2 \cdot \Big(|\{x \in \mathbb{R}^n : |f(x)| > \alpha/2\}| + |\{x \in \mathbb{R}^n : |g(x)| > \alpha/2\}| \Big) \\ &\le 2(\|f\|_{1,w} + \|g\|_{1,w}). \end{aligned}$$

Finally, we observe, that the function $x \to \dfrac{1}{\|x\|^n} \in L_{1,w}(\mathbb{R}^n) \setminus L_1(\mathbb{R}^n)$.

The main aim of this section is to prove the following theorem.

Theorem 15.2.1. *Let f be a measurable function on $\mathbb{R}^n$. Then*

i) If $f \in L_p(\mathbb{R}^n)$, $1 \le p \le \infty$, then the function Mf is finite almost everywhere.

ii) If $f \in L_1(\mathbb{R}^n)$, then $Mf \in L_{1,w}(\mathbb{R}^n)$ and

$$\|Mf\|_{1,w} \le A\|f\|_1,$$

where A is a constant which depends only on the dimension (i.e. $A = 5^n$ will do).

iii) If $f \in L_p(\mathbb{R}^n)$ with $1 < p \le \infty$, then $Mf \in L_p(\mathbb{R}^n)$ and

$$\|Mf\|_p \le A_p\|f\|_p,$$

where A_p depends only on p and dimension n.

The proof is based on the following *covering lemma.*

Lemma 15.2.2. *Let E be a measurable subset of $\mathbb{R}^n$, which is covered by the union of a family of balls (B^j) with uniformly bounded diameter. Then from this family we can select a disjoint subsequence, $B_1, B_2, B_3, \dots$, such that*

$$\sum_k |B_k| \ge C|E|.$$

Here C is a positive constant that depends only on the dimension n; $C = 5^{-n}$ will do.

Proof. We describe first the choice of $B_1, B_2, \dots$. We choose B_1 so that it is essentially as large as possible, i.e.

$$\mathrm{diam}(B_1) \ge \frac{1}{2}\sup_j\ \mathrm{diam}(B^j).$$

The choice of B_1 is not unique, but that shall not hurt us.

If $B_1, B_2, \dots, B_k$ were already chosen, we take again B_{k+1} disjoint with $B_1, \dots, B_k$ and again nearly as large as possible, i.e.

$$\mathrm{diam}(B_{k+1}) > \frac{1}{2}\sup\{\mathrm{diam}(B^j) : B^j \text{ disjoint with } B_1, \dots, B_k\}.$$

In this way, we get a sequence $B_1, B_2, \dots, B_k, \dots$ of balls. It can be also finite, if there were no balls B^j disjoint with $B_1, B_2, \dots, B_k$.

If $\sum_k |B_k| = \infty$, then the conclusion of lemma is satisfied and we are done. If $\sum_k |B_k| < \infty$, we argue as follows.

We denote by B_k^* the ball with the same center as B_k and diameter five times as large. We claim that

$$\bigcup_k B_k^* \supset E,$$

which then immediately gives that $|E| \le \sum_k |B_k^*| = 5^n \sum_k |B_k|$.

We shall show that $\bigcup_k B_k^* \supset B^j$ for every j. This is clear if B^j is one of the balls in the preselected sequence. If it is not the case, we obtain $\mathrm{diam}(B_k) \to 0$ (as $\sum_k |B_k| < \infty$) and we choose the first k with $\mathrm{diam}(B_{k+1}) < \frac{1}{2}\mathrm{diam}(B^j)$. That means, that B^j must intersect one of the balls $B_1, \dots, B_k$, say B_{k_0}. Obvious geometric arguments (based on the inequality $\mathrm{diam}(B_{k_0}) \ge 1/2 \cdot \mathrm{diam}(B^j)$) then give that $B^j \subset B_{k_0}^*$. This finishes the proof. $\square$

Proof. (of Theorem 15.2.1). Let $\alpha > 0$ and let us consider $E_\alpha := \{x \in \mathbb{R}^n : Mf(x) > \alpha\}$. From the definition of M we obtain, that for every $x \in E_\alpha$, there is a ball B_x centered at x such that

$$\int_{B_x} |f(y)|dy > \alpha|B_x|.$$

Hence, we get $|B_x| < \|f\|_1/\alpha$ and $\cup_{x\in E_\alpha} B_x \supset E_\alpha$ and the balls $(B_x)_{x\in E_\alpha}$ satisfy the assumptions of Lemma 15.2.2. Using this covering lemma, we get a sequence of disjoint balls $(B_k)_k$, such that

$$\sum_k |B_k| \geq C|E_\alpha|.$$

We therefore obtain

$$\|f\|_1 \geq \int_{\cup_k B_k} |f(y)|dy = \sum_k \int_{B_k} |f| > \alpha \sum_k |B_k| \geq \alpha C|E_\alpha|,$$

which may be rewritten as $\sup_{\alpha>0} \alpha \cdot |E_\alpha| = \|Mf\|_{1,w} \leq \frac{1}{C}\|f\|_1$.

This proves the first assertion of the theorem for $p = 1$ and the second assertion.

We now consider $1 < p < \infty$. The proof follows from the information on the endpoints, i.e. from

$$\|Mf\|_{1,w} \leq C\|f\|_1 \quad \text{and} \quad \|Mf\|_\infty \leq \|f\|_\infty.$$

Actually, the *Theorem of Marcinkiewicz* says, that *every* operator with these two properties is then automaticaly bounded on all $L_p(\mathbb{R}^n)$, $1 < p < \infty$. Furthermore, this is a corner stone of the so-called *interpolation theory*.

We shall restrict ourself to the necessary minimum needed to prove the theorem. This is based on the technique of splitting the function into *good* and *bad* part, idea elaborated in detail by Calderón and Zygmund.

Let $\alpha > 0$ and put $f_1(x) := f(x)$ if $|f(x)| > \alpha/2$ and $f_1(x) := 0$ otherwise. Due to $|f(x)| \leq |f_1(x)| + \alpha/2$, we have also $Mf(x) \leq Mf_1(x) + \alpha/2$ and also

$$\{x \in \mathbb{R}^n : Mf(x) > \alpha\} \subset \{x \in \mathbb{R}^n : Mf_1(x) > \alpha/2\}.$$

Due to the second part of the theorem

$$\begin{aligned}|E_\alpha| = |\{x \in \mathbb{R}^n : Mf(x) > \alpha\}| &\leq |\{x \in \mathbb{R}^n : Mf_1(x) > \alpha/2\}| \leq \frac{2A}{\alpha}\|f_1\|_1\\ &\leq \frac{2A}{\alpha}\int_{\{x\in\mathbb{R}^n:|f(x)|>\alpha/2\}} |f(y)|dy.\end{aligned}$$

We use the information of the size of the level sets of Mf to estimate the L_p-norm of Mf.

$$\begin{aligned}\|Mf\|_p^p &= \int_{\mathbb{R}^n} Mf(x)^p dx = \int_0^\infty |\{x \in \mathbb{R}^n : Mf(x)^p > \alpha\}|d\alpha\\ &= \int_0^\infty |\{x \in \mathbb{R}^n : Mf(x) > \alpha^{1/p}\}|d\alpha\\ &= p\int_0^\infty \beta^{p-1}|\{x \in \mathbb{R}^n : Mf(x) > \beta\}|d\beta\\ &\leq p\int_0^\infty \beta^{p-1}\left(\frac{2A}{\beta}\int_{\{x\in\mathbb{R}^n:|f(x)|>\beta/2\}} |f(y)|dy\right)d\beta\\ &= 2Ap\int_{\mathbb{R}^n} |f(y)|\int_0^{2|f(y)|} \beta^{p-2}d\beta dy = \frac{2Ap}{p-1}\int_{\mathbb{R}^n} |f(y)|\cdot|2f(y)|^{p-1}dy\\ &= \frac{2^pAp}{p-1}\int_{\mathbb{R}^n} |f(y)|^p dy,\end{aligned}$$

where we used Fubini's theorem and the substitution $\beta := \alpha^{1/p}$ with $d\alpha = p\beta^{p-1}d\beta$. This gives the first and the third statement of the theorem with

$$A_p = 2\left(\frac{5^n p}{p-1}\right)^{1/p}, \quad 1 < p < \infty.$$

$\square$

Corollary 15.2.3. (Lebesgue's differentiation theorem)
Let f be locally integrable on $\mathbb{R}^n$. Then

$$\lim_{r\to 0}\frac{1}{|B(x,r)|}\int_{B(x,r)} f(y)dy = f(x)$$

holds for almost every $x \in \mathbb{R}^n$.

Proof. We may cover the set of "bad" points

$$\left\{x \in \mathbb{R}^n : \lim_{r\to 0}\frac{1}{|B(x,r)|}\int_{B(x,r)} f(y)dy \text{ does not exist or } \lim_{r\to 0}\frac{1}{|B(x,r)|}\int_{B(x,r)} f(y)dy \neq f(x)\right\}$$

by

$$\bigcup_{k=1}^{\infty}\left\{x \in \mathbb{R}^n : \limsup_{r\to 0}\frac{1}{|B(x,r)|}\int_{B(x,r)} f(y)dy - f(x) > \frac{1}{k}\right\}$$

united with

$$\bigcup_{k=1}^{\infty}\left\{x \in \mathbb{R}^n : \liminf_{r\to 0}\frac{1}{|B(x,r)|}\int_{B(x,r)} f(y)dy - f(x) < -\frac{1}{k}\right\}.$$

It is therefore enough to show, that each of these sets has measure zero. Let us fix $k \in \mathbb{N}$ and decompose $f = g + h$, where $g \in C(\mathbb{R}^n)$ and $\|h\|_1 \le t$, $t > 0$.

Obviously,

$$\lim_{r\to 0}\frac{1}{|B(x,r)|}\int_{B(x,r)} g(y)dy = g(x), \quad x \in \mathbb{R}^n,$$

which implies

$$\begin{aligned}
&\left\{x \in \mathbb{R}^n : \limsup_{r\to 0}\frac{1}{|B(x,r)|}\int_{B(x,r)} f(y)dy - f(x) > \frac{1}{k}\right\}\\
&\quad= \left\{x \in \mathbb{R}^n : \limsup_{r\to 0}\frac{1}{|B(x,r)|}\int_{B(x,r)} h(y)dy - h(x) > \frac{1}{k}\right\}\\
&\quad\subset \left\{x \in \mathbb{R}^n : Mh(x) > \frac{1}{2k}\right\} \cup \left\{x \in \mathbb{R}^n : |h(x)| > \frac{1}{2k}\right\}.
\end{aligned}$$

The measure of the first set is smaller than $2kA\|h\|_1$ and the measure of the second is smaller than $2k\|h\|_1$. As $\|h\|_1$ might be chosen arbitrary small, the measure of the original set is zero. The same argument works for $\liminf$ instead of $\limsup$ as well.

$\square$

Theorem 15.2.4. *Let φ be a function which is non-negative, radial, decreasing (as function on $(0,\infty)$) and integrable. Then*

$$\sup_{t>0} |\varphi_t * f(x)| \leq \|\varphi\|_1 Mf(x),$$

where again $\varphi_t(x) = t^{-n}\varphi(x/t), x \in \mathbb{R}^n$.

Proof. Let in addition φ be a simple function. Then it can be written as

$$\varphi(x) = \sum_{j=1}^{k} a_j \chi_{B_{r_j}}(x),$$

with $a_j > 0$ and $r_j > 0$. Then

$$\varphi * f(x) = \sum_{j=1}^{k} a_j |B_{r_j}| \frac{1}{|B_{r_j}|} \chi_{B_{r_j}} * f(x) \leq \|\varphi\|_1 Mf(x),$$

since $\|\varphi\|_1 = \sum_j a_j |B_{r_j}|$. As any normalized dilation of φ satisfies the same assumptions and has the same integral, it satisfies also the same inequality. Finally, any function satisfying the hypotheses of the Theorem can be approximated monotonically from below by a sequence of simple radial functions. This finishes the proof. □

15.3 Calderón-Zygmund decomposition

The aim of this is to present the decomposition method of Calderón and Zygmund. In its simple form, it was already used in the proof of boudnedness of the maximal operator M.

Theorem 15.3.1. *Given a function f, which is integrable and non-negative, and given a positive number λ, there exists a sequence*[14] *(Q_j) of disjoint dyadic cubes such that*

i) $f(x) \leq \lambda$ *for almost every* $x \notin \bigcup_j Q_j$;

ii) $\left|\bigcup_j Q_j\right| \leq \frac{1}{\lambda}\|f\|_1$;

iii) $\lambda < \frac{1}{|Q_j|}\int_{Q_j} f \leq 2^n \lambda$.

Proof. We denote by $\mathcal{Q}_k$ the collection of dyadic cubes with side length 2^{-k}, $k \in \mathbb{Z}$. Furthermore, we define

$$E_k f(x) = \sum_{Q \in \mathcal{Q}_k} \left(\frac{1}{|Q|}\int_Q f\right) \chi_Q(x).$$

We define also

$$\Omega_k := \{x \in \mathbb{R}^n : E_k f(x) > \lambda \text{ and } E_j f(x) \leq \lambda \text{ if } j < k\}$$

That is, $x \in \Omega_k$ if k is the first index with $E_k f(x) > \lambda$. As the integrability of f implies $E_k f(x) \to 0$ for $k \to -\infty$, such k always exists. The sets Ω_k are clearly disjoint and each can be written as the union of cubes in $\mathcal{Q}_k$. Together, these cubes form the system (Q_j).

[14] possibly finite, or even empty

This gives the third statement of the theorem. The first follows by Lebesgue differentiation theorem: indeed, $E_k f(x) \le \lambda$ for all $k \in \mathbb{Z}$ implies $f(x) \le \lambda$ at almost every such point. The second follows just by

$$\Big|\bigcup_j Q_j\Big| = \sum_j |Q_j| \le \frac{1}{\lambda}\sum_j \int_{Q_j} f \le \frac{1}{\lambda}\|f\|_1.$$

$\square$

15.4 Interpolation theorems

We shall present two basic interpolation theorems, the Riesz-Thorin interpolation theorem and Marcinkiewicz interpolation theorem.

The operator T mapping measurable functions to measurable functions is called sub-linear, if

$$\begin{aligned} |T(f_0+f_1)(x)| &\le |Tf_0(x)| + |Tf_1(x)|,\\ |T(\lambda f)(x)| &= |\lambda|\cdot|Tf(x)|, \quad \lambda\in\mathbb{C}. \end{aligned}$$

Let (X,μ) and (Y,ν) be measure spaces and let T be a sub-linear operator mapping $L_p(X,\mu)$ into a space of measurable functions on (Y,ν). We say that T is strong type (p,q) if it is bounded from $L_p(X,\mu)$ into $L_q(Y,\nu)$. We say, that it is of weak type (p,q), $q<\infty$, if

$$\|Tf\|_{q,w} := \sup_{\lambda>0} \lambda\cdot \nu^{1/q}(\{y\in Y : |Tf(y)|>\lambda\}) \le C\|f\|_p, \quad f\in L_p(X,\mu).$$

Theorem 15.4.1. *(Marcinkiewicz interpolation theorem)*
Let (X,μ) and (Y,ν) be measure spaces, $1\le p_0<p_1\le\infty$, and let T be a sublinear operator from $L_{p_0}(X,\mu)+L_{p_1}(X,\mu)$ to the measurable functions on Y that is weak (p_0,p_0) type and weak (p_1,p_1) type. Then T is strong (p,p) for $p_0<p<p_1$.

Proof. Let $\lambda>0$ be given and let $f\in L_p(X,\mu)$. Then we decompose f into $f=f_0+f_1$ with

$$\begin{aligned} f_0 &= f\chi_{\{x:|f(x)|>\lambda\}},\\ f_1 &= f\chi_{\{x:|f(x)|\le\lambda\}}. \end{aligned}$$

The case $p_1=\infty$ appeared implicitly already in the proof of the boundedness of Hardy-Littlewood maximal operator, so we suppose that $p_1<\infty$. Then we have

$$\nu(\{y\in Y:|Tf(y)|>\lambda\}) \le \nu(\{y\in Y:|Tf_0(y)|>\lambda/2\}) + \nu(\{y\in Y:|Tf_1(y)|>\lambda/2\})$$

and

$$\nu(\{y\in Y:|Tf_i(y)|>\lambda/2\}) \le \left(\frac{2A_i}{\lambda}\|f_i\|_{p_i}\right)^{p_i}, \quad i=0,1.$$

We combine them to get

$$\begin{aligned}
\|Tf\|_p^p &= p\int_0^\infty \lambda^{p-1}\nu(\{x : |Tf(x)| > \lambda\})d\lambda \\
&\le p\int_0^\infty \lambda^{p-1-p_0}(2A_0)^{p_0}\int_{x:|f(x)|>\lambda}|f(x)|^{p_0}d\mu d\lambda \\
&\quad + p\int_0^\infty \lambda^{p-1-p_1}(2A_1)^{p_1}\int_{x:|f(x)|\le\lambda}|f(x)|^{p_1}d\mu d\lambda \\
&= p(2A_0)^{p_0}\int_X |f(x)|^{p_0}\int_0^{|f(x)|}\lambda^{p-1-p_0}d\lambda d\mu \\
&\quad + p(2A_1)^{p_1}\int_X |f(x)|^{p_1}\int_{|f(x)|}^\infty \lambda^{p-1-p_1}d\lambda d\mu \\
&= C\|f\|_p^p.
\end{aligned}$$

□

Following theorem belongs also to the classical heart of interpolation theory. As we shall not need it in the sequel, we state it without proof.

Theorem (Riesz-Thorin Interpolation)
Let $1 \le p_0, p_1, q_0, q_1 < \infty$, and for $0 < \theta < 1$ define p and q by

$$\frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1} \quad \text{and} \quad \frac{1}{q} = \frac{1-\theta}{q_0} + \frac{\theta}{q_1}.$$

If T is a linear operator from $L_{p_0} + L_{p_1}$ to $L_{q_0} + L_{q_1}$, such that

$$\|Tf\|_{q_0} \le M_0\|f\|_{p_0} \quad \text{for } f \in L_{p_0}$$

and

$$\|Tf\|_{q_1} \le M_1\|f\|_{p_1} \quad \text{for } f \in L_{p_1},$$

then

$$\|Tf\|_q \le M_0^{1-\theta}M_1^{\theta}\|f\|_p \quad \text{for } f \in L_p.$$

15.5 Hilbert Transform

The *Hilbert transform* of a measurable function f on $\mathbb{R}$ is defined as

$$Hf(x) = \frac{1}{\pi}\int_{\mathbb{R}}\frac{f(x-y)}{y}dy.$$

As the integral does not converge absolutely, it has to be interpreted in an appropriate limiting sense, which uses its *cancelation property*, i.e.

$$Hf(x) = \lim_{\varepsilon\to 0^+}\frac{1}{\pi}\int_{y:|y|>\varepsilon}\frac{f(x-y)}{y}dy.$$

With this definition, $Hf(x)$ makes sense for all smooth functions, especially for $f \in \mathscr{S}(\mathbb{R})$. [15]

[15] Observe, that the restriction to $n = 1$ is both natural and essential for the cancelation property. Furthermore, from now on, we shall denote the Fourier transform of a function f also by the more usual $\widehat{f}$.

If we define the distribution

$$\text{p.v.}\frac{1}{x}(\varphi) := \lim_{\varepsilon\to 0^+}\int_{x:|x|>\varepsilon}\frac{\varphi(x)}{x}dx, \quad \varphi\in\mathscr{S}(\mathbb{R}),$$

then $Hf := \frac{1}{\pi}\text{p.v.}\frac{1}{x} * f$. This formula suggests that we look for the Fourier transform of Hf. Therefore, we regularize $1/x$. This can be done using complex distributions $\frac{1}{\pi(x\pm i\varepsilon)}$ and letting $\varepsilon\to 0$ or (and that is what we shall do) by defining

$$Q_t(x) := \frac{1}{\pi}\cdot\frac{x}{t^2+x^2}.$$

Obviously, $\lim_{t\to 0} Q_t(x) = \frac{1}{\pi x}$ holds pointwise and, as we shall show below, also in $\mathscr{S}'(\mathbb{R})$. As

$$\begin{aligned}
\mathcal{F}^{-1}(\operatorname{sgn}(x)e^{-a|x|})(\xi) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\operatorname{sgn}(x)e^{-a|x|}\cdot e^{ix\xi}dx\\
&= \frac{1}{\sqrt{2\pi}}\left\{-\int_{-\infty}^{0}e^{x(a+i\xi)}dx + \int_0^\infty e^{x(-a+i\xi)}dx\right\}\\
&= \frac{1}{\sqrt{2\pi}}\left\{\frac{-1}{a+i\xi}+\frac{1}{a-i\xi}\right\} = \sqrt{\frac{2}{\pi}}\cdot\frac{i\xi}{a^2+\xi^2},
\end{aligned}$$

we obtain

$$\widehat{Q}_t(\xi) = \frac{-i}{\sqrt{2\pi}}\operatorname{sgn}(\xi)e^{-t|\xi|},$$

we have also $\lim_{t\to 0}\widehat{Q}_t(\xi) = -i\operatorname{sgn}(\xi)/\sqrt{2\pi}$. As this convergence is uniform on compact sets, it holds also in $\mathscr{S}'(\mathbb{R})$. Finally, due to the continuity of Fourier transform, we obtain

$$\left(\frac{1}{\pi}\text{p.v.}\frac{1}{x}\right)^\wedge = [\lim_{t\to 0}Q_t]^\wedge = \lim_{t\to 0}\widehat{Q}_t = \frac{-i\operatorname{sgn}(\cdot)}{\sqrt{2\pi}}.$$

Theorem 15.5.1. *In $\mathscr{S}'(\mathbb{R})$,*

$$\lim_{t\to 0}Q_t(x) = \frac{1}{\pi}\text{p.v.}\frac{1}{x}.$$

Proof. For each $\varepsilon>0$, the functions $\psi_\varepsilon(x) = x^{-1}\chi_{\{y:|y|>\varepsilon\}}(x)$ are bounded and define tempered distributions with $\lim_{\varepsilon\to 0^+}\psi_\varepsilon = \text{p.v.}\frac{1}{x}$. Therefore, it is enough to show that

$$\lim_{t\to 0}(\pi Q_t - \psi_t) = 0$$

in $\mathscr{S}'(\mathbb{R})$. This follows by

$$\begin{aligned}
(\pi Q_t - \psi_t)(\varphi) &= \int_{\mathbb{R}}\frac{x\varphi(x)}{t^2+x^2}dx - \int_{x:|x|>t}\frac{\varphi(x)}{x}dx\\
&= \int_{x:|x|<t}\frac{x\varphi(x)}{t^2+x^2}dx + \int_{x:|x|>t}\left(\frac{x}{t^2+x^2}-\frac{1}{x}\right)\varphi(x)dx\\
&= \int_{x:|x|<1}\frac{x\varphi(xt)}{1+x^2}dx - \int_{x:|x|>1}\frac{\varphi(tx)}{x(1+x^2)}dx
\end{aligned}$$

for $\varphi\in\mathscr{S}(\mathbb{R})$. As $t\to 0$, we apply Lebesgue dominated convergence theorem and use the symmetry of the integrands to conclude, that the limit is zero. $\square$

$$
\begin{array}{ccc}
Q_t & \xrightarrow[t\to 0]{\mathscr{S}'(\mathbb{R})} & \frac{1}{\pi}\text{p.v.}\frac{1}{x} \\
\mathcal{F}\Big\downarrow & & \mathcal{F}\Big\downarrow \\
\frac{-i}{\sqrt{2\pi}}\operatorname{sgn}(\xi)e^{-t|\xi|} & \xrightarrow[t\to 0]{\mathscr{S}'(\mathbb{R})} & \frac{-i\operatorname{sgn}(\cdot)}{\sqrt{2\pi}},
\end{array}
$$

Summarizing, one defines the Hilbert transform Hf for $f \in \mathscr{S}(\mathbb{R}^n)$ by any of these formulas:

$$
\begin{aligned}
Hf &= \frac{1}{\pi}\text{p.v.}\frac{1}{x} * f, \\
Hf &= \lim_{t\to 0} Q_t * f, \\
(Hf)^\wedge(\xi) &= -i\operatorname{sgn}(\xi)\hat{f}(\xi).
\end{aligned}
$$

Using the third expression, we can extend the definition of H to $L_2(\mathbb{R})$ and it holds

$$
\begin{aligned}
\|Hf\|_2 &= \|(Hf)^\wedge\|_2 = \|-i\operatorname{sgn}(\cdot)\hat{f}\|_2 = \|\hat{f}\|_2 = \|f\|_2, \\
H(Hf) &= (-i\operatorname{sgn}(\cdot)(Hf)^\wedge)^\vee = ((-i\operatorname{sgn}(\cdot))^2\hat{f})^\vee = (-\hat{f})^\vee = -f, \\
\langle Hf, Hg\rangle &= \langle f, g\rangle, \quad \text{by polarization}, \\
\int Hf\cdot g &= -\int f\cdot Hg,
\end{aligned}
$$

where the last identity follows from

$$
\langle Hf, g\rangle = \langle (Hf)^\wedge, \hat{g}\rangle = \langle -i\operatorname{sgn}(\cdot)\hat{f}, \hat{g}\rangle = \langle \hat{f}, \operatorname{sgn}(\cdot)\hat{g}\rangle = \langle \hat{f}, -(Hg)^\wedge\rangle = -\langle f, Hg\rangle
$$

and the simple fact that $\overline{Hg} = H\overline{g}$.

Theorem 15.5.2. *For $f \in \mathscr{S}(\mathbb{R})$, the following is true.*

(i) (Kolmogorov) H is weak type $(1,1)$,

$$
\|Hf\|_{1,w} \le C\|f\|_1, \quad \textit{i.e.} \quad |\{x\in\mathbb{R} : |Hf(x)| > \lambda\}| \le \frac{C}{\lambda}\|f\|_1, \quad \lambda > 0, f \in L_1(\mathbb{R}^n).
$$

(ii) (M. Riesz) H is of strong type (p,p) for every $1 < p < \infty$, i.e.

$$
\|Hf\|_p \le C_p\|f\|_p.
$$

Proof. Step 1.: We show the weak type $(1,1)$ by exploiting Theorem 15.3.1. Let f be non-negative and let $\lambda > 0$, then Theorem 15.3.1 gives a sequence of disjoint intervals (I_j), such that

$$
\begin{aligned}
f(x) \le \lambda \text{ for a.e. } x \notin \Omega = \bigcup_j I_j, \\
|\Omega| \le \frac{1}{\lambda}\|f\|_1, \\
\lambda < \frac{1}{|I_j|}\int_{I_j} f \le 2\lambda.
\end{aligned}
$$

Given this decomposition of $\mathbb{R}$, we decompose f into "good" and "bad" part defined by

$$g(x) = \begin{cases} f(x), & x \notin \Omega, \\ \dfrac{1}{|I_j|}\displaystyle\int_{I_j} f, & x \in I_j, \end{cases} \qquad b(x) = \sum_j b_j(x) = \sum_j \left(f(x) - \frac{1}{|I_j|}\int_{I_j} f \right) \chi_{I_j}(x).$$

Then $g(x) \le 2\lambda$ almost everywhere, and b_j is supported on I_j and has zero integral. Since $Hf = Hg + Hb$, we have

$$|\{x \in \mathbb{R} : |Hf(x)| > \lambda\}| \le |\{x \in \mathbb{R} : |Hg(x)| > \lambda/2\}| + |\{x \in \mathbb{R} : |Hb(x)| > \lambda/2\}|.$$

We estimate the first term using the L_2-boundedness of H by

$$|\{x \in \mathbb{R} : |Hg(x)| > \lambda/2\}| \le \frac{4}{\lambda^2}\int_{\mathbb{R}} |Hg(x)|^2 dx = \frac{4}{\lambda^2}\int_{\mathbb{R}} g(x)^2 dx \le \frac{8}{\lambda}\int_{\mathbb{R}} g(x)dx = \frac{8}{\lambda}\int_{\mathbb{R}} f(x)dx.$$

Let $2I_j$ be the interval with the same center as I_j and twice the length. Let $\Omega^* = \bigcup_j 2I_j$. Then $|\Omega^*| \le 2|\Omega|$ and

$$\begin{aligned} |\{x \in \mathbb{R} : |Hb(x)| > \lambda/2\}| &\le |\Omega^*| + |\{x \notin \Omega^* : |Hb(x)| > \lambda/2\}| \\ &\le \frac{2}{\lambda}\|f\|_1 + \frac{2}{\lambda}\int_{\mathbb{R}\setminus\Omega^*} |Hb(x)|dx. \end{aligned}$$

As $|Hb(x)| \le \sum_j |Hb_j(x)|$, it is enough to show that

$$\sum_j \int_{\mathbb{R}\setminus 2I_j} |Hb_j(x)|dx \le C\|f\|_1.$$

Denote the center of I_j by c_j and use that b_j has zero integral to get

$$\begin{aligned} \int_{\mathbb{R}\setminus 2I_j} |Hb_j(x)|dx &= \int_{\mathbb{R}\setminus 2I_j} \left| \int_{I_j} \frac{b_j(y)}{x-y} dy \right| dx \\ &= \int_{\mathbb{R}\setminus 2I_j} \left| \int_{I_j} b_j(y) \left(\frac{1}{x-y} - \frac{1}{x-c_j} \right) dy \right| dx \\ &\le \int_{I_j} |b_j(y)| \left(\int_{\mathbb{R}\setminus 2I_j} \frac{|y-c_j|}{|x-y|\cdot|x-c_j|} dx \right) dy \\ &\le \int_{I_j} |b_j(y)| \left(\int_{\mathbb{R}\setminus 2I_j} \frac{|I_j|}{|x-c_j|^2} dx \right) dy \end{aligned}$$

The last inequality follows from $|y-c_j| < |I_j|/2$ and $|x-y| > |x-c_j|/2$. The inner integral equals 2, so

$$\sum_j \int_{\mathbb{R}\setminus 2I_j} |Hb_j(x)|dx \le 2\sum_j \int_{I_j} |b_j(y)|dy \le 4\|f\|_1.$$

Our proof of the weak $(1,1)$ inequality is for non-negative f, but this is sufficient since an arbitrary real function can be decomposed into its positive and negative parts, and a complex function into its real and imaginary parts.

Step 2.: Since H is weak type $(1,1)$ and strong type $(2,2)$, it is also strong type (p,p) for $1<p<2$. If $p>2$, we apply duality, i.e.

$$\begin{aligned}\|Hf\|_p &= \sup\left\{\left|\int_{\mathbb{R}} Hf\cdot g\right| : \|g\|_{p'}\le 1\right\}\\ &= \sup\left\{\left|\int_{\mathbb{R}} f\cdot Hg\right| : \|g\|_{p'}\le 1\right\}\\ &\le \|f\|_p\cdot \sup\left\{\|Hg\|_{p'} : \|g\|_{p'}\le 1\right\}\le C_{p'}\|f\|_p.\end{aligned}$$

□

The strong (p,p) inequality is false if $p=1$ or $p=\infty$; this can easily be seen if we let $f=\chi_{[0,1]}$. Then

$$Hf(x)=\frac{1}{\pi}\log\left|\frac{x}{x-1}\right|,$$

and Hf is neither integrable nor bounded.

15.6 BMO

Hilbert transform acts rather badly on $L_\infty(\mathbb{R})$. Not only is H unbounded on $L_\infty(\mathbb{R})$, it can not be easily defined on a dense subset of $L_\infty(\mathbb{R})$. The definition

$$Hf(y)=\frac{1}{\pi}\int_{\mathbb{R}}\frac{f(x)}{y-x}dx$$

runs for $f\in L_\infty(\mathbb{R})$ into troubles for x near y and x near infinity. If we look onto differences, the situation changes to

$$Hf(y)-Hf(y')=\frac{1}{\pi}\int_{\mathbb{R}} f(x)\left(\frac{1}{y-x}-\frac{1}{y'-x}\right)dx.$$

This improves the situation for x near infinity, as $1/(x-y)-1/(x-y')=O(1/x^2)$ in this case.

We say that f and g are equivalent modulo a constant if $f(x)=g(x)+C$ for some (complex) constant C and almost every $x\in\mathbb{R}$. Given $f\in L_\infty(\mathbb{R})$ and $y\in\mathbb{R}$. Then we take an open interval $B\subset\mathbb{R}$ with center at zero containing y. Then $f\chi_B\in L_2(\mathbb{R})$ and we define $Hf(y)$ to be

$$Hf(y):=H(f\chi_B)(y)+\frac{1}{\pi}\int_{\mathbb{R}\setminus B} f(x)\left(\frac{1}{y-x}+\frac{1}{x}\right)dx.$$

The first term is defined by the L_2 definition of H, the integral in the second term converges absolutely. The definition depends on B, but choosing another interval $B'\supset B$ with the center at the origin leads to a difference

$$\begin{aligned}&H(f\chi_B)(y)+\frac{1}{\pi}\int_{\mathbb{R}\setminus B} f(x)\left(\frac{1}{y-x}+\frac{1}{x}\right)dx-H(f\chi_{B'})(y)-\frac{1}{\pi}\int_{\mathbb{R}\setminus B'} f(x)\left(\frac{1}{y-x}+\frac{1}{x}\right)dx\\ &=H(f\chi_B-f\chi_{B'})(y)+\frac{1}{\pi}\int_{B'\setminus B} f(x)\left(\frac{1}{y-x}+\frac{1}{x}\right)dx\\ &=-H(f\chi_{B'\setminus B})(y)+\frac{1}{\pi}\int_{B'\setminus B} f(x)\left(\frac{1}{y-x}+\frac{1}{x}\right)dx=\frac{1}{\pi}\int_{B'\setminus B}\frac{f(x)}{x}dx,\end{aligned}$$

which does not depend on y. This defines Hf modulo constant for $f\in L_\infty(\mathbb{R})$. Of course, such a definition does not allow to measure Hf in the usual norms, as L_p. Instead of that, we need a space of functions defined modulo constants.

Definition 15.6.1. (Bounded mean oscillation). Let $f : \mathbb{R}^n \to \mathbb{C}$ be a function defined modulo a constant. The BMO (or Bounded Mean Oscillation) norm of f is defined

$$\|f|BMO(\mathbb{R}^n)\| := \sup_B \frac{1}{|B|}\int_B \left| f - \frac{1}{|B|}\int_B f \right|$$

where B ranges over all balls. Note that if one shifts f by a constant, the BMO norm is unchanged, so this norm is well-defined for functions defined modulo constants. We denote by $BMO(\mathbb{R}^n)$ the space of all functions with finite BMO norm.

Example 15.6.2. Let $f(x) = \operatorname{sgn}(x)$. Let $|y| < a/2$. We take $B = (-a, a)$ and apply the definition of Hf as presented above. This gives (for $y > 0$)

$$\begin{aligned}
\pi Hf(y) &= \text{p.v.}\int_{-a}^{a} \frac{\operatorname{sgn}(x)}{y-x}dx + \int_{(-\infty,-a)\cup(a,\infty)} \operatorname{sgn}(x)\left(\frac{1}{y-x} + \frac{1}{x}\right)dx \\
&= \lim_{\varepsilon\to 0^+}\left(\int_{(-a,a)\setminus(y-\varepsilon,y+\varepsilon)} \frac{\operatorname{sgn}(x)}{y-x}dx\right) \underbrace{-\int_{-\infty}^{-a}\left(\frac{1}{x} - \frac{1}{x-y}\right)dx}_{-\ln\frac{a}{a+y}} + \underbrace{\int_a^\infty \left(\frac{1}{x} - \frac{1}{x-y}\right)dx}_{=\ln\frac{x}{x-y}\Big|_{x=a}^{x=\infty} = -\ln\frac{a}{a-y}} \\
&= \lim_{\varepsilon\to 0^+}\left(\int_{-a}^{0} \frac{1}{x-y}dx - \int_0^{y-\varepsilon}\frac{1}{x-y}dx - \int_{y+\varepsilon}^{a}\frac{1}{x-y}\right) - \ln\frac{a}{a+y} - \ln\frac{a}{a-y} \\
&= \lim_{\varepsilon\to 0^+}\left(\ln\frac{y}{a+y} + \ln\frac{y}{\varepsilon} - \ln\frac{a-y}{\varepsilon}\right) - \ln\frac{a}{a+y} - \ln\frac{a}{a-y} \\
&= 2\ln y - 2\ln a.
\end{aligned}$$

If $y < 0$, similar calculation applies as well. Hence, $\pi Hf(y) = 2\ln|y| - 2\ln a$. Hence, ignoring the constant,

$$H(\operatorname{sgn} x) = \frac{2}{\pi}\ln|x|.$$

Let us observe, that

$$\frac{1}{|B|}\int_B \left| f - \frac{1}{|B|}\int_B f \right| \approx \inf_{c\in\mathbb{R}} \frac{1}{|B|}\int_B |f-c|$$

holds for every ball B with universal constants. Indeed, the left-hand side is obviously larger than the right hand side. On the other hand, we get as well

$$\begin{aligned}
\frac{1}{|B|}\int_B \left| f - \frac{1}{|B|}\int_B f \right| &\le \frac{1}{|B|}\int_B |f-c| + \frac{1}{|B|}\int_B \left| c - \frac{1}{|B|}\int_B f \right| \\
&= \frac{1}{|B|}\int_B |f-c| + \frac{1}{|B|}\int_B \left| \frac{1}{|B|}\int_B (c-f) \right| \\
&\le \frac{1}{|B|}\int_B |f-c| + \frac{1}{|B|}\int_B \left(\frac{1}{|B|}\int_B |c-f| \right) \\
&= \frac{2}{|B|}\int_B |f-c|\,.
\end{aligned}$$

Theorem 15.6.3. *(H maps $L_\infty(\mathbb{R})$ into $BMO(\mathbb{R})$)*
Let $f \in L_\infty(\mathbb{R})$. Then

$$\|Hf|BMO(\mathbb{R})\| \lesssim \|f\|_\infty.$$

Proof. Due to the observation above, it is enough to show that for every ball B, there is a constant c_B such that

$$\frac{1}{|B|}\int_B |Hf - c_B| \lesssim \|f\|_\infty.$$

We split $Hf = H(f\chi_{2B}) + H(f\chi_{\mathbb{R}\setminus 2B})$, where $2B$ is a ball with the same center as B, but twice the radius. First we get

$$\frac{1}{|B|}\int_B |H(f\chi_{2B})| \le \frac{1}{|B|}\left(\int_B |H(f\chi_{2B})|^2\right)^{1/2}\cdot\left(\int_B 1\right)^{1/2} \le \frac{1}{\sqrt{|B|}}\left(\int_{2B}|f|^2\right)^{1/2} \le \frac{\|f\|_\infty}{\sqrt{|B|}}\sqrt{|2B|} \lesssim \|f\|_\infty.$$

This deals with the "local" part of Hf. For the "global" part, observe that for $x\in B$ we have (modulo constant)

$$H(f\chi_{\mathbb{R}\setminus 2B})(x) = \int_{y:y\notin 2B} f(y)\left(\frac{1}{x-y} - \frac{1}{\gamma-y}\right)dy,$$

where γ is the center of B.

$$\begin{aligned}
&\frac{1}{|B|}\int_B\left|\int_{\mathbb{R}\setminus 2B} f(y)\left(\frac{1}{x-y}-\frac{1}{\gamma-y}\right)dy\right|dx\\
&\quad\le \frac{1}{|B|}\int_B\int_{\mathbb{R}\setminus 2B}\left|f(y)\left(\frac{1}{x-y}-\frac{1}{\gamma-y}\right)\right|dydx\\
&\quad\le \frac{\|f\|_\infty}{|B|}\int_B\int_{\mathbb{R}\setminus 2B}\left|\frac{1}{x-y}-\frac{1}{\gamma-y}\right|dydx
\end{aligned}$$

By shifting, we may assume, that $\gamma = 0, B = (-a,a)$ and $2B = (-2a, 2a)$. We estimate

$$\frac{1}{2a}\int_0^a\int_{2a}^\infty \frac{x}{|y(x-y)|}dydx \le \frac{1}{2a}\int_0^a xdx\int_{2a}^\infty\frac{1}{|y(y-a)|}dy \le \frac{a}{4}\int_{2a}^\infty\frac{1}{(y-a)^2}dy \le c$$

(and similarly for the remaining parts). Altogether, this gives

$$\frac{1}{|B|}\int_B |H(f\chi_{\mathbb{R}\setminus 2B})| \lesssim \|f\|_\infty.$$

Adding the two facts, we obtain the claim. □

15.7 Singular integrals

Hilbert transform is the most important example of the so-called *singular integrals*. These are convolutions with kernel K, which might have a singularity at origin.

Theorem 15.7.1. (Calderón-Zygmund) *Let $K \in \mathscr{S}'(\mathbb{R}^n)$ be a tempered distribution, which is associated to a locally integrable function on $\mathbb{R}^n\setminus\{0\}$ and satisfies*

(CZ1) $\quad |\hat{K}(\xi)| \le A, \quad \xi\in\mathbb{R}^n,$

(CZ2) $\quad \displaystyle\int_{\|x\|_2\ge 2\|y\|_2} |K(x-y)-K(x)|dx \le B, \quad y\in\mathbb{R}^n.$

Then, for $1<p<\infty$,

$$\|K*f\|_p \le C_p\|f\|_p \quad \textit{and} \quad \|K*f\|_{1,w}\le C\|f\|_1.$$

The proof copies very much the proof of the same statement for the Hilbert transform, and we leave out the details.

The condition (CZ2) is sometimes called *Hörmander condition.* By the help of mean value theorem, it is satisfied for example if

$$\|\nabla K(x)\|_2 \leq \frac{C}{\|x\|_2^{n+1}}, \quad x \neq 0.$$

An important and non-trivial generalisation of the theory of singular integrals is given by considering the vector-valued analogues. By this, we mean the following.

- $H, \tilde{H}$ are (complex) Hilbert spaces.
- For $0 < p < \infty$, $L_p(\mathbb{R}^n \to H)$ is the set of measurable functions $f : \mathbb{R}^n \to H$, such that $\int_{\mathbb{R}^n} \|f(x)\|_H^p dx < \infty$.
- Let $K : \mathbb{R}^n \to \mathscr{L}(H, \tilde{H})$. Then $Tf(x) = \displaystyle\int_{\mathbb{R}^n} K(y)f(x-y)dy$ takes values in $\tilde{H}$.
- Under same (just appropriately interpreted) conditions as above, T is bounded from $L_p(\mathbb{R}^n \to H)$ into $L_p(\mathbb{R}^n \to \tilde{H})$. Especially, the gradient condition above is still valid in this case.

15.8 Khintchine inequality

We denote by

$$r_n(t) := \text{sign}\, \sin(2^n \pi t), \quad t \in [0,1], n \in \mathbb{N}_0.$$

the *Rademacher functions.*

The system $(r_n)_{n=0}^\infty$ forms an orthonormal system in $L_2(0,1)$, but it is not a basis (consider i.e. the function $f(t) = 1 - 2t$).

Theorem 15.8.1. *Let $p \in [1,\infty)$. Then there are positive constants A_p and B_p such that*

$$A_p \left(\sum_{n=1}^m |a_n|^2 \right)^{1/2} \leq \left(\int_0^1 \left| \sum_{n=1}^m a_n r_n(t) \right|^p dt \right)^{1/p} \leq B_p \left(\sum_{n=1}^m |a_n|^2 \right)^{1/2}$$

holds for every $m \in \mathbb{N}$ and every sequence of real numbers $a_1, \ldots, a_m$.

Proof. By A_p and B_p we denote the best possible constants (which are actually known, but we shall derive slightly weaker estimates). Furthermore, orthogonality of Rademacher functions gives immediately $A_2 = B_2 = 1$. Finally, due to monotonicity of the L_p-norms, we have $A_r \leq A_p$ and $B_r \leq B_p$ for $r \leq p$.

So, it is enough to show that $A_1 > 0$ and $B_{2k} < \infty$ for all $k \in \mathbb{N}$.

We start with B_{2k}. Let us observe that

$$\begin{aligned} E := \int_0^1 \left| \sum_{n=1}^m a_n r_n(t) \right|^{2k} dt &= \int_0^1 \left(\sum_{n=1}^m a_n r_n(t) \right)^{2k} dt \\ &= \sum_{|\alpha|=2k} \frac{(2k)!}{\alpha_1! \ldots \alpha_m!} a_1^{\alpha_1} \ldots a_m^{\alpha_m} \int_0^1 r_1^{\alpha_1}(t) \ldots r_m^{\alpha_m}(t) dt \\ &= \sum_{|\alpha|=k} \frac{(2k)!}{(2\alpha_1)! \ldots (2\alpha_m)!} a_1^{2\alpha_1} \ldots a_m^{2\alpha_m} \int_0^1 r_1^{2\alpha_1}(t) \ldots r_m^{2\alpha_m}(t) dt \\ &= \sum_{|\alpha|=k} \frac{(2k)!}{(2\alpha_1)! \ldots (2\alpha_m)!} a_1^{2\alpha_1} \ldots a_m^{2\alpha_m}, \end{aligned}$$

where we have used the multinomial theorem (a generalisation of the binomial theorem to a bigger number of summands) and the fact that

$$\int_0^1 r_1^{\alpha_1}(t)\dots r_m^{\alpha_m}(t)dt$$

is equal to zero if some of the α_i's is odd and equal to 1 if all of them are even.

Let $\alpha = (\alpha_1, \dots, \alpha_m) \in \mathbb{N}_0^m$ be integers with $|\alpha| = k$, then

$$2^k\alpha_1!\dots\alpha_m! = (2^{\alpha_1}\alpha_1!)\dots(2^{\alpha_m}\alpha_m!) \le (2\alpha_1)!\dots(2\alpha_m)!.$$

This implies

$$\begin{aligned} E &\le \frac{(2k)!}{2^k k!}\sum_{|\alpha|=k}\frac{k!}{\alpha_1!\dots\alpha_m!}a_1^{2\alpha_1}\dots a_m^{2\alpha_m} \\ &= \frac{(2k)!}{2^k k!}\left(\sum_{n=1}^m |a_n|^2\right)^k = \frac{(2k)!}{2^k k!}\|a\|_2^{2k} \end{aligned}$$

and

$$E^{1/(2k)} \le \left(\frac{(2k)!}{2^k k!}\right)^{1/(2k)} \|a\|_2.$$

Hence, the statement holds with [16]

$$B_{2k} := \left(\frac{(2k)!}{2^k k!}\right)^{1/(2k)}.$$

Finally, we have to show the existence of $A_1 > 0$. We proceed by a nice duality trick using the (already proven) first part of this theorem.

Let $f(t) := \sum_{n=1}^m a_n r_n(t)$. By Hölder's inequality for $p = 3/2$ and $p' = 3$, we have

$$\begin{aligned} \int_0^1 |f(t)|^2 dt &= \int_0^1 |f(t)|^{2/3}\cdot|f(t)|^{4/3}dt \le \left(\int_0^1 |f(t)|dt\right)^{2/3}\cdot\left(\int_0^1 |f(t)|^4\right)^{1/3} \\ &\le \left(\int_0^1 |f(t)|dt\right)^{2/3} B_4^{4/3}\cdot\|a\|_2^{4/3} = \left(\int_0^1 |f(t)|dt\right)^{2/3} B_4^{4/3}\cdot\|f\|_2^{4/3}. \end{aligned}$$

Therefore,

$$\left(\int_0^1 |f(t)|dt\right)^{2/3} \ge B_4^{-4/3}\left(\int_0^1 |f(t)|^2 dt\right)^{1/3},$$

that is

$$\int_0^1 |f(t)|dt \ge B_4^{-2}\left(\int_0^1 |f(t)|^2 dt\right)^{1/2} = B_4^{-2}\|a\|_2.$$

Hence, $A_1 \ge B_4^{-2}$. □

Remark 15.8.2. Stochastic reformulation of Khintchine's inequalities sounds as follows. Let $\varepsilon_i, i = 1, \dots, m$ be independent variables with $\mathbb{P}(\varepsilon_i = 1) = 1/2$ and $\mathbb{P}(\varepsilon_i = -1) = 1/2$. Let $1 \le p < \infty$. Then there are constants A_p, B_p such that for every $a_1, \dots, a_m \in \mathbb{R}$ [17]

$$A_p\|a\|_2 \le \left(\mathbb{E}\left|\sum_{i=1}^m a_i\varepsilon_i\right|^p\right)^{1/p} \le B_p\|a\|_2.$$

[16] By Stirling's formula, one can show quite easily that B_{2k} grows as $\sqrt{2k}$ for $k \to \infty$.

[17] Also $a_1, \dots, a_m \in \mathbb{C}$ can be considered with slightly modified proof.

Choosing p large enough, this estimate gives very quickly the so-called *tail bound estimates on sum of independent Rademacher variables*, i.e. the assymptotic estimates of

$$\mathbb{P}\left(\left|\sum_{i=1}^{m} a_i\varepsilon_i\right| > t\right)$$

for $t \to \infty$.

We use this reformulation of Khintchine's inequalities to give another proof of Theorem 15.8.1.

Proof. (of the upper estimate in Theorem 15.8.1).
We normalize to $\|a\|_2 = 1$. Then

$$\mathbb{E}\exp\left(\sum_{i=1}^{m} a_i\varepsilon_i\right) = \mathbb{E}\prod_{i=1}^{m}\exp(a_i\varepsilon_i) = \prod_{i=1}^{m}\mathbb{E}\exp(a_i\varepsilon_i) = \prod_{i=1}^{m}\cosh(a_i).$$

Using Taylor's expansion, one obtains $\cosh(a_j) \le \exp(a_j^2/2)$. Hence,

$$\mathbb{E}\exp\left(\sum_{i=1}^{m} a_i\varepsilon_i\right) \le \prod_{i=1}^{m}\exp(a_i^2/2) \lesssim 1,$$

and by Markov's inequality

$$\begin{aligned}\mathbb{P}\left(\sum_{i=1}^{m} a_i\varepsilon_i > \lambda\right) &= \mathbb{P}\left(\exp\left(\sum_{i=1}^{m} a_i\varepsilon_i\right) > \exp(\lambda)\right) = \mathbb{P}\left(\exp\left(\sum_{i=1}^{m} a_i\varepsilon_i - \lambda\right) > 1\right)\\ &\le \mathbb{E}\exp\left(\sum_{i=1}^{m} a_i\varepsilon_i - \lambda\right) \lesssim e^{-\lambda}.\end{aligned}$$

By symmetry of ε_i's, we get also $\mathbb{P}\left(\left|\sum_{i=1}^{m} a_i\varepsilon_i\right| > \lambda\right) \lesssim e^{-\lambda}$. The rest then follows by distributional representation of the L_p-norm. ☐

Khintchine's inequalities have an interesting application in operator theory. Let $1 \le p < \infty$ and let $T : L_p(\mathbb{R}^n) \to L_p(\mathbb{R}^n)$ be a bounded linear operator. Then

$$\left\|\left(\sum_{j=0}^{N}|Tf_j|^2\right)^{1/2}\right\|_p \le c_p\left\|\left(\sum_{j=0}^{N}|f_j|^2\right)^{1/2}\right\|_p, \tag{15.1}$$

where the constant c_p depends only on p and $\|T\|$.

The proof follows by considering Rademacher functions $r_1, \ldots, r_N$ and

$$\begin{aligned}\left\|\left(\sum_{j=0}^{N}|Tf_j|^2\right)^{1/2}\right\|_p^p &= \int_{\mathbb{R}^n}\left(\sum_{j=1}^{N}|(Tf_j)(x)|^2\right)^{p/2}dx \le c\int_{\mathbb{R}^n}\left(\int_0^1|\sum_{j=1}^{N}Tf_j(x)r_j(t)|^p dt\right)^{p/p}dx\\ &= c\int_0^1\int_{\mathbb{R}^n}|\sum_{j=1}^{N}Tf_j(x)r_j(t)|^p dxdt = c\int_0^1\int_{\mathbb{R}^n}|T\Big(\sum_{j=1}^{N}f_jr_j(t)\Big)(x)|^p dxdt\\ &\le c\|T\|^p\int_0^1\int_{\mathbb{R}^n}|\sum_{j=1}^{N}f_j(x)r_j(t)|^p dxdt \le c'\int_{\mathbb{R}^n}\left(\sum_{j=1}^{N}|f_j(x)|^2\right)^{p/2}dx\\ &= c_p^p\left\|\left(\sum_{j=0}^{N}|f_j|^2\right)^{1/2}\right\|_p^p.\end{aligned}$$

By letting $N \to \infty$, the same result holds also for infinite sums.

15.9 Littlewood-Paley Theory

Let $\{I_j\}$ be a sequence of intervals on the real line, finite or infinite, and let $\{S_j\}$ be the sequence of operators defined by $(S_j f)^\wedge(\xi) = \chi_{I_j}(\xi)\hat{f}(\xi)$. Later on, we shall concentrate on the *dyadic decomposition* of $\mathbb{R}$ (strictly speaking of $\mathbb{R}\setminus\{0\}$) given by

$$I_j = (-2^{j+1}, -2^j] \cup [2^j, 2^{j+1}), \quad j \in \mathbb{Z}. \tag{15.2}$$

Furthermore, we denote $S_j^* = S_{j-1} + S_j + S_{j+1}$. Let us observe that this implies $S_j^* S_j = S_j S_j^* = S_j$.

Finally, we addopt this concept also to smooth dyadic decompositions. Let $\psi \in \mathscr{S}(\mathbb{R})$ be non-negative, have support in $1/2 \le \|\xi\|_2 \le 4$ and be equal to 1 on $1 \le \|\xi\|_2 \le 2$. Then we define

$$\psi_j(\xi) = \psi(2^{-j}\xi) \quad \text{and} \quad (\tilde{S}_j f)^\wedge(\xi) = \psi_j(\xi)\hat{f}(\xi), \quad \xi \in \mathbb{R}.$$

Theorem 15.9.1. (Littlewood-Paley Theory) *Let* $1 < p < \infty$.

i) Then there exist two constants $C_p > c_p > 0$, *such that*

$$c_p\|f\|_p \le \left\| \left(\sum_j |S_j f|^2 \right)^{1/2} \right\|_p \le C_p \|f\|_p.$$

The same holds for S_j^*.

ii) There exists a constant $C_p > 0$ *such that*

$$\left\| \left(\sum_j |\tilde{S}_j f|^2 \right)^{1/2} \right\|_p \le C_p \|f\|_p.$$

iii) Finally, if $\sum_j |\psi(2^{-j}\xi)|^2 = 1$, *then there is also a constant* $c_p > 0$, *such that*

$$c_p\|f\|_p \le \left\| \left(\sum_j |\tilde{S}_j f|^2 \right)^{1/2} \right\|_p.$$

Proof. Step 1.
We know, that $(S_j f)^\wedge = \chi_{I_j}\hat{f}$, where I_j was defined by (15.2). We define

$$I_j^- := (-2^{j+1}, -2^j], \qquad I_j^+ := [2^j, 2^{j+1}), \quad j \in \mathbb{Z}$$

and split $S_j = S_j^- + S_j^+$, where $(S_j^- f)^\wedge = \chi_{I_j^-}\hat{f}$ and $(S_j^+ f)^\wedge = \chi_{I_j^+}\hat{f}$.

We observe, that

$$\chi_{I_j^+}(x) = \frac{1}{2}\left(\operatorname{sgn}(x - 2^j) - \operatorname{sgn}(x - 2^{j+1})\right) \quad \text{for (almost) all } x \in \mathbb{R},$$

and

$$S_j^+ f = (\chi_{I_j^+}\hat{f})^\vee = \frac{1}{2}\left((\operatorname{sgn}(\cdot - 2^j)\hat{f})^\vee - (\operatorname{sgn}(\cdot - 2^{j+1})\hat{f})^\vee\right).$$

Finally, we write

$$\operatorname{sgn}(\xi - 2^j)\hat{f}(\xi) = \tau_{2^j}[\operatorname{sgn}(\xi)\hat{f}(\xi + 2^j)] = \tau_{2^j}[\operatorname{sgn}(\xi) \cdot \tau_{-2^j}\hat{f}(\xi)],$$

leading to

$$\begin{aligned}
(\operatorname{sgn}(\xi-2^j)\hat f(\xi))^\vee &= M_{2^j}(\operatorname{sgn}\cdot\tau_{-2^j}\hat f)^\vee = (2\pi)^{-1/2}\cdot M_{2^j}(\operatorname{sgn}(\cdot)^\vee * (\tau_{-2^j}\hat f)^\vee)\\
&= (2\pi)^{-1/2}M_{2^j}\left(\frac{\sqrt{2\pi}}{-i}\left(\frac{1}{\pi}p.v.\frac{1}{x}\right) * M_{-2^j}f\right)\\
&= iM_{2^j}\left(\left(\frac{1}{\pi}p.v.\frac{1}{x}\right) * M_{-2^j}f\right)\\
&= iM_{2^j}HM_{-2^j}f.
\end{aligned}$$

Using the boundedness of H on $L_p(\mathbb{R})$ for $1<p<\infty$, we immediately obtain that $\|S_jf\|_p\le c\|f\|_p$, and the same is true also for S_j^+ and S_j^-.

Step 2.

We combine Step 1. with (15.1) to obtain

$$\begin{aligned}
\left\|\left(\sum_{j\in\mathbb{Z}}|S_j^+f_j|^2\right)^{1/2}\right\|_p &\le \frac12\left\{\left\|\left(\sum_{j\in\mathbb{Z}}|(\operatorname{sgn}(\xi-2^j)\hat f_j(\xi))^\vee|^2\right)^{1/2}\right\|_p+\left\|\left(\sum_{j\in\mathbb{Z}}|(\operatorname{sgn}(\xi-2^{j+1})\hat f_j(\xi))^\vee|^2\right)^{1/2}\right\|_p\right\}\\
&\le \frac12\left\{\left\|\left(\sum_{j\in\mathbb{Z}}|M_{2^j}HM_{-2^j}f_j|^2\right)^{1/2}\right\|_p+\dots\right\}\\
&= \frac12\left\{\left\|\left(\sum_{j\in\mathbb{Z}}|HM_{-2^j}f_j|^2\right)^{1/2}\right\|_p+\dots\right\}\\
&\le c_p\left\{\left\|\left(\sum_{j\in\mathbb{Z}}|M_{-2^j}f_j|^2\right)^{1/2}\right\|_p+\dots\right\} = c_p'\left\|\left(\sum_{j\in\mathbb{Z}}|f_j|^2\right)^{1/2}\right\|_p.
\end{aligned}$$

The same holds of course for S_j^- and, therefore, also for S_j.

Step 3.

This, together with the identity $S_j=S_j\tilde S_j$ implies

$$\left\|\left(\sum_{j\in\mathbb{Z}}|S_jf|^2\right)^{1/2}\right\|_p=\left\|\left(\sum_{j\in\mathbb{Z}}|S_j\tilde S_jf|^2\right)^{1/2}\right\|_p\le\left\|\left(\sum_{j\in\mathbb{Z}}|\tilde S_jf|^2\right)^{1/2}\right\|_p.$$

Step 4.

This shows, that the second inequality in part (i) of the theorem follows from (ii). Therefore, we prove (ii) now.

Let $\hat\Psi=\psi$ and $\Psi_j(x)=2^j\Psi(2^jx)$. Then $\hat\Psi_j=\psi_j$ and $\tilde S_jf=\Psi_j*f$. It is enough to prove that the vector-valued mapping

$$f\to(\tilde S_jf)_j$$

is bounded from L_p to $L_p(\ell_2)$. If $p=2$, this follows by Plancherel theorem:

$$\left\|\left(\sum_j|\tilde S_jf|^2\right)^{1/2}\right\|_2^2=\int_{\mathbb{R}}\sum_j|\psi_j(\xi)|^2\cdot|\hat f(\xi)|^2d\xi\le 3\|f\|_2^2.$$

The proof for $p\ne2$ is based on the Hörmander's condition for vector-valued singular integrals, i.e. that

$$\|\Psi_j'(x)\|_{\ell_2}\le C\|x\|_2^{-2}.$$

Using that $\Psi \in \mathscr{S}(\mathbb{R})$, we obtain

$$\left(\sum_j |\Psi_j'(x)|^2\right)^{1/2} \leq \sum_j |\Psi_j'(x)| = \sum_j 2^{2j}|\Psi'(2^j x)| \leq C \sum_j 2^{2j} \max(1, 2^j|x|)^{-3} \leq \frac{C}{|x|^2}.$$

Step 5.

Finally, we prove the first inequality in part (i) and part (iii) of the theorem. Surprisingly enough, they follow quite quickly from previous steps and duality.

The identity

$$\left\| \Big(\sum_{j\in\mathbb{Z}} |S_j f|^2\Big)^{1/2} \right\|_2 = \|f\|_2 \tag{15.3}$$

follows by Plancherel's theorem and, by polarization, also

$$\int_{\mathbb{R}} \sum_j S_j f \cdot \overline{S_j g} = \int_{\mathbb{R}} f\overline{g}$$

follows. Using this, and the first part of the theorem for p' with $1/p + 1/p' = 1$ allows the following estimate.

$$\begin{aligned}
\|f\|_p &= \sup\left\{ \left| \int_{\mathbb{R}} f\overline{g} \right| : \|g\|_{p'} \leq 1 \right\} \\
&= \sup\left\{ \left| \int_{\mathbb{R}} \sum_j S_j f \cdot \overline{S_j g} \right| : \|g\|_{p'} \leq 1 \right\} \\
&\leq \sup\left\{ \left\| \Big(\sum_{j\in\mathbb{Z}} |S_j f|^2\Big)^{1/2} \right\|_p \cdot \left\| \Big(\sum_{j\in\mathbb{Z}} |S_j g|^2\Big)^{1/2} \right\|_p : \|g\|_{p'} \leq 1 \right\} \\
&\leq c_p \left\| \Big(\sum_{j\in\mathbb{Z}} |S_j f|^2\Big)^{1/2} \right\|_p .
\end{aligned}$$

Part (iii) of the theorem follows in exactly the same way - the assumption of the theorem gives exactly the identity (15.3). $\square$

16 Selected topics

16.1 Riesz representation theorem

We prove the Riesz Representation Theorem, which was used to construct the Borel measurable functional calculus. We follow essentially [14], where one can find a lot more to measure and integration theory.

16.1.1 Outer measures

Definition 16.1.1. By an *outer measure* on a set X we understand a set function γ, which assigns to *every* subset $A \subset X$ a non-negative number $\gamma(A) \in [0, \infty]$, such that the following conditions are satisfied:

i) $\gamma(\varnothing) = 0$,

ii) if $A \subset B$, then $\gamma(A) \le \gamma(B)$,

iii) $\gamma(\cup A_n) \le \sum \gamma(A_n)$ for all $A_n \subset X$.

We describe a construction (due to Carathéodory) which constructs a measure from an outer measure.

Definition 16.1.2. A set $M \subset X$ is said to be γ*-measurable (in the sense of Carathéodory)* if

$$\gamma(T) = \gamma(T \cap M) + \gamma(T \setminus M)$$

for each set $T \subset X$ (in other words, M splits additively each set $T \subset X$). The collection of all γ-measurable sets will be denoted by $\mathfrak{M}(\gamma)$.

Let us note the inequality $\gamma(T) \le \gamma(T \cap M) + \gamma(T \setminus M)$ always holds. To prove measurability, it is therefore sufficient to show that $\gamma(T) \ge \gamma(T \cap M) + \gamma(T \setminus M)$ for all $T \subset X$ with $\gamma(T) < \infty$.

Theorem 16.1.3. *$\mathfrak{M}(\gamma)$ is a σ-algebra and γ is a complete measure on $\mathfrak{M}(\gamma)$.*

Proof. The proof will be divided into several steps.
Step 1.: It is easy to check that $X \in \mathfrak{M}(\gamma)$, that $X \setminus M \in \mathfrak{M}(\gamma)$ provided $M \in \mathfrak{M}(\gamma)$ and that $A \in \mathfrak{M}(\gamma)$ whenever $\gamma(A) = 0$.
Step 2.: Let $M, N \in \mathfrak{M}(\gamma)$. We would like to show that $M \cap N \in \mathfrak{M}(\gamma)$. Let $T \subset X$ be arbitrary. We have

$$\gamma(T) = \gamma(T \cap M) + \gamma(T \setminus M)$$

and

$$\gamma(T \cap M) = \gamma(T \cap M \cap N) + \gamma((T \cap M) \setminus N)$$

together with

$$\begin{aligned}\gamma(T \setminus (M \cap N)) &= \gamma((T \setminus (M \cap N)) \cap M) + \gamma((T \setminus (M \cap N)) \setminus M)\\ &= \gamma((T \cap M) \setminus N) + \gamma(T \setminus M).\end{aligned}$$

We put these three inequalities together and obtain

$$\begin{aligned}\gamma(T) &= \gamma(T \cap M \cap N) + \gamma((T \cap M) \setminus N) + \gamma(T \setminus M)\\ &= \gamma(T \cap M \cap N) + \gamma(T \setminus (M \cap N))\\ &= \gamma(T \cap (M \cap N)) + \gamma(T \setminus (M \cap N)).\end{aligned}$$

Since $\mathfrak{M}(\gamma)$ is closed under complements and under finite intersections, it is also closed under finite unions.
Step 3.: To show that γ is σ-additive on $\mathfrak{M}(\gamma)$, choose $M_n \in \mathfrak{M}(\gamma)$ pairwise disjoint. Setting $T = M_1 \cup M_2$ and using γ-measurability of M_1, we obtain

$$\gamma(M_1 \cup M_2) = \gamma(M_1) + \gamma(M_2).$$

Thus γ is finitely additive. Further

$$\sum_{n=1}^{\infty} \gamma(M_n) = \lim_{k\to\infty} \sum_{n=1}^{k} \gamma(M_n) = \lim_{k\to\infty} \gamma\left(\bigcup_{n=1}^{k} M_n\right) \leq \gamma\left(\bigcup_{n=1}^{\infty} M_n\right),$$

and since the reverse inequality always holds, we reach the conclusion.
Step 4.: Let now $M_n \in \mathfrak{M}(\gamma)$ be pairwise disjoint. Our aim is to show that $\bigcup_{n=1}^{\infty} M_n \in \mathfrak{M}(\gamma)$. Choosing a test set $T \subset X$, we have

$$\gamma(T) = \gamma\left(T \setminus \bigcup_{n=1}^{k} M_n\right) + \gamma\left(T \cap \bigcup_{n=1}^{k} M_n\right) \geq \gamma\left(T \setminus \bigcup_{n=1}^{\infty} M_n\right) + \sum_{n=1}^{k} \gamma(T \cap M_n)$$

for each $k \in \mathbb{N}$. Since γ is σ-additive (Step 3), it follows

$$\gamma(T) \geq \gamma\left(T \setminus \bigcup_{n=1}^{\infty} M_n\right) + \gamma(T \cap \bigcup_{n=1}^{\infty} M_n),$$

which is what we wanted as the second inequality is trivial. □

16.1.2 The Theorem

In this section, (K, τ) is a compact topological vector space. Although Riesz Representation Theorem holds even in a more general setting, we shall restrict to this and discuss the extensions separately later on.

Theorem 16.1.4. *Urysohn's lemma (for locally compact spaces).*
If K is a compact set and U an open subset of a locally compact space X, $K \subset U \subset X$, then there exists a continuous function f and a compact set L with

$$K \subset L \subset U, \quad 0 \leq f \leq 1, \quad f = 1 \text{ on } K, \quad f = 0 \text{ on } X \setminus L.$$

Theorem 16.1.5. *Dini's property.*
If (f_n) is a monotone sequence of continuous functions on a compact space X, which converges pointwise to a continuous function, then the convergence of (f_n) is uniform on X.

Theorem 16.1.6. *(Daniell's property) Let A be a positive linear functional on $\mathcal{C}(K)$. Let $f_n \in \mathcal{C}(K)$ with $f_n \searrow 0$. Then $Af_n \to 0$.*

Proof. From positivity of A (i.e. $f \geq 0$ implies $Af \geq 0$) and linearity of A, the monotonicity of A follows (i.e. $f \leq g$ implies $Af \leq Ag$). Therefore, the limit $b := \lim_{n\to\infty} Af_n \geq 0$ always exists. By Dini's theorem, $f_n \rightrightarrows 0$ and, therefore, there exists a sequence (n_k), such that $\|f_{n_k}\|_\infty \leq k^{-2}$ for each $k \in \mathbb{N}$. Thus the series $\sum_k f_{n_k}$ converges uniformly on K and if $f := \sum_k f_{n_k}$, then $f \in \mathcal{C}(K)$. Finally, we obtain for each $k \in \mathbb{N}$

$$kb \leq \sum_{i=1}^{k} Af_{n_i} = A\left(\sum_{i=1}^{k} f_{n_i}\right) \leq Af$$

and $b = 0$ follows. □

Let us stress, that we did not suppose continuity of A in this theorem.

Theorem 16.1.7. *(Riesz Representation Theorem)*
Let A be a positive linear functional on $\mathcal{C}(K)$. Then there exists a unique complete Radon measure μ on K, such that $\mathcal{C}(K) \subset L_1(\mu)$ and $Af = \int_K f d\mu$ for all $f \in \mathcal{C}(K)$.

Proof. The proof is based on the following idea. We first define

$$\begin{aligned} \mu_A^*(G) &:= \sup\{Af : f \in \mathcal{C}(K) : 0 \leq f \leq 1, f = 0 \text{ on } K \setminus G\}, \quad G \text{ is open}, \\ \mu_A^*(E) &:= \inf\{\mu_A^*(G) : G \text{ open}, G \supset E\}, \ E \subset K \text{ arbitrary}. \end{aligned}$$

As μ_A^* is monotone on open sets, the second line of this definition extends the first line and the notation is correct. Such a set function is called *outer Radon measure* (associated to A). We shall show, that this mapping is really an outer measure (in the sense described in Definition 16.1.2), and we shall prove its regularity properties. The construction of Carathéodory will then give a σ-algebra (and we shall show, that it contains all open sets) on which μ_A^* is a measure. Finally, we shall prove, that this measure has the property from the statement of the theorem. Let us work out this program.

Step 1.: We shall show the following

i) $\mu_A^*(C) = \inf\{Ag : g \in \mathcal{C}(K), 0 \leq g \leq 1, g = 1 \text{ on } C\}$ for every compact set $C \subset K$.

ii) $\mu_A^*(G) = \sup\{\mu_A^*(C) : C \text{ compact}, C \subset G\}$ for every open set $G \subset K$.

iii) μ_A^* is an outer measure.

Step 1.(i): Let $C \subset K$ be a compact set and let $g \in \mathcal{C}(K), 0 \leq g \leq 1, g = 1$ on C. First we show that $\mu_A^*(G) \leq Ag$. Fix $\varepsilon \in (0,1)$ and denote $G := \{x \in K : g(x) > 1 - \varepsilon\}$. Obviously, $C \subset G$. If $f \in \mathcal{C}(K), 0 \leq f \leq 1$ and $f = 0$ on $K \setminus G$, then $f \leq \frac{g}{1-\varepsilon}$ and $Af \leq \frac{Ag}{1-\varepsilon}$ holds for every f with such properties. Hence $\mu_A^*(G) \leq \frac{Ag}{1-\varepsilon}$ and

$$Ag \geq (1-\varepsilon)\mu_A^*(G) \geq (1-\varepsilon)\mu_A^*(C).$$

We see that $Ag \geq \mu_A^*(C)$, and thus

$$\mu_A^*(C) \leq \inf\{Ag : g \in \mathcal{C}(K), 0 \leq g \leq 1, g = 1 \text{ on } C\}.$$

To prove the reverse inequality, select an open set $G \supset C$. Urysohn's lemma provides a function $g \in \mathcal{C}(K), 0 \leq g \leq 1, g = 0$ on $K \setminus G$ and $g = 1$ on C. The definition of $\mu_A^*(G)$ for G open gives $Ag \leq \mu_A^*(G)$. Finally, we observe that

$$\begin{aligned} \mu_A^*(C) &= \inf\{\mu_A^*(G) : G \text{ open}, G \supset C\} \\ &\geq \inf\{Ag : g \in \mathcal{C}(K), 0 \leq g \leq 1, g = 1 \text{ on } C\}. \end{aligned}$$

Step 1.(ii): Suppose we are given an open set $G \subset K$ and $\varepsilon > 0$. Let $f \in \mathcal{C}(K)$ be a function such that $0 \leq f \leq 1$, $f = 0$ on $K \setminus G$. Since $f_n := \min(f, 1/n) \searrow 0$, an appeal to Daniell's property gives the existence of an $n \in \mathbb{N}$, such that $Af_n < \varepsilon$. If $C := \{x \in K : f(x) \geq 1/n\}$, then C is a compact subset of K. Due to (i), there exists $g \in \mathcal{C}(K)$ such that $0 \leq g \leq 1, g = 1$ on C and $Ag \leq \mu_A^*(C) + \varepsilon$. Since $f - f_n \leq g$, we get

$$Af \leq Af_n + Ag \leq \mu_A^*(C) + 2\varepsilon.$$

Hence,

$$\begin{aligned}\mu_A^*(G) &= \sup\{Af : f\in\mathcal{C}(K) : 0\le f\le 1, f=0 \text{ on } K\setminus G\}\\ &\le \sup\{\mu_A^*(C) : C \text{ compact}, C\subset G\}+2\varepsilon,\end{aligned}$$

and we finish the proof of one part of (ii). The second one follows directly by monotonicity of μ_A^*.
Step 1.(iii): Finally, we prove that μ_A^* is an outer measure. Clearly, we have $\mu_A^*(\varnothing)=0$ and $\mu_A^*(S)\le\mu_A^*(T)$ whenever $S\subset T$. It remains to show that μ_A^* is σ-subadditive.

To this end, let $\varepsilon>0$ and $C_1,C_2\subset K$ be two given compact sets. By (i), we find $f_j\in\mathcal{C}(K)$ such that $f_j=1$ on C_j and $Af_j\le\mu_A^*(C_j)+\varepsilon; j=1,2$. Then (by (i) again)

$$\mu_A^*(C_1\cup C_2)\le A(\min(f_1+f_2,1))\le A(f_1+f_2)=Af_1+Af_2\le\mu_A^*(C_1)+\mu_A^*(C_2)+2\varepsilon$$

and we see that μ_A^* is subadditive on compact sets.

Next, consider open sets $G_1,G_2\subset K$ and pick a compact set $C\subset G_1\cup G_2$. For any point $x\in C$, there is a neighborhood V_x whose closure is either in G_1 or in G_2. Thanks to the compactness of C, we obtain finite collections of open sets $(V_i^1)_i,(V_i^2)_i$ such that $\overline{V_i^j}\subset G_j$ and $\bigcup_i V_i^1\cup\bigcup_i V_i^2\supset C$. Set $C_j:=C\cap\bigcup_i\overline{V_i^j}; j=1,2$. Then C_j are compact, $C_j\subset G_j$ and $C=C_1\cup C_2$. Hence, $\mu_A^*(C)\le\mu_A^*(C_1)+\mu_A^*(C_2)\le\mu_A^*(G_1)+\mu_A^*(G_2)$ and it readily follows that μ^* is subadditive on open sets.

Now, let (G_n) be a sequence of open sets. Chose a compact set $C\subset\bigcup_{i=1}^\infty G_i$. Then $C\subset\bigcup_{i=1}^n G_i$ for some $n\in\mathbb{N}$, and

$$\mu_A^*(C)\le\mu_A^*\left(\bigcup_{i=1}^n G_i\right)\le\sum_{i=1}^n\mu_A^*(G_i)\le\sum_{i=1}^\infty\mu_A^*(G_i)$$

and from (ii), we conclude that $\mu_A^*(\bigcup_{i=1}^\infty G_i)\le\sum_{i=1}^\infty\mu_A^*(G_i)$.

Finally, let (E_n) be a system of arbitrary subsets of K. We wish to show that $\mu_A^*(\bigcup E_n)\le\sum\mu_A^*(E_n)$. It is clearly sufficient to assume that $\mu_A^*(E_n)<\infty$ for all n. Given $\varepsilon>0$, we find open sets G_n such that $E_n\subset G_n$ and $\mu_A^*(G_n)<\mu_A^*(E_n)+2^{-n}\varepsilon$. Then

$$\mu_A^*\left(\bigcup E_n\right)\le\mu_A^*\left(\bigcup G_n\right)\le\sum\mu_A^*(E_n)+\varepsilon.$$

As $\varepsilon>0$ was arbitrary, μ_A^* is subadditive as needed.
Step 2.: We show, that every Borel subset of K is μ_A^* measurable (in the sense of Carathéodory). It is sufficient to prove measurability of open subsets. Notice, that we have $\mu_A^*(G_1\cup G_2)=\mu_A^*(G_1)+\mu_A^*(G_2)$ for open disjoint sets G_1 and G_2. This follows easily from definition of $\mu_A^*(G)$ for G open. Indeed, there are $f_j\in\mathcal{C}(K), j=1,2$ with $0\le f_j\le 1$, $f_j=0$ on $K\setminus G_j$ and $\mu_A^*(G_j)\le Af_j+\varepsilon$. Then $f_1+f_2\in\mathcal{C}(K)$ and $0\le f_1+f_2\le 1$, $f_1+f_2=0$ on $K\setminus(G_1\cup G_2)$ gives $\mu_A^*(G_1\cup G_2)\ge A(f_1+f_2)=A(f_1)+A(f_2)\ge\mu_A^*(G_1)+\mu_A^*(G_2)-2\varepsilon$. The reverse inequality appeared in Step 1. (iii).

Now given an open set $G\subset K$, and a test set $T\subset K$ and $\varepsilon>0$, we approximate $T\cap G$ and $T\setminus G$ be disjoint open sets.

Hence, there exists open sets V_1 and V_2, such that

$$\begin{aligned}V_1\supset T,&\quad \mu_A^*(V_1)<\mu_A^*(T)+\varepsilon,\\ V_2\supset V_1\setminus G,&\quad \mu_A^*(V_2)<\mu_A^*(V_1\setminus G)+\varepsilon.\end{aligned}$$

We can also find a compact set $C\subset V_1\cap G$ such that $\mu_A^*(C)+\varepsilon>\mu_A^*(V_1\cap G)$ and an open set W with a compact closure $\overline{W}$ such that $C\subset W\subset\overline{W}\subset V_1\cap G$. Set

$W_0 := (V_1 \cap V_2) \setminus \overline{W}$. Then W_0 is an open set, $W \cap W_0 = \varnothing$, $W \cup W_0 \subset V_1$, $V_1 \setminus G \subset W_0$ and $\mu_A^*(W) + \varepsilon > \mu_A^*(V_1 \cap G)$. Thus

$$\begin{aligned}\mu_A^*(T \cap G) + \mu_A^*(T \setminus G) &\le \mu_A^*(V_1 \cap G) + \mu_A^*(V_1 \setminus G) < \mu_A^*(W) + \varepsilon + \mu_A^*(W_0)\\ &= \mu_A^*(W \cup W_0) + \varepsilon \le \mu_A^*(V_1) + \varepsilon < \mu_A^*(T) + 2\varepsilon.\end{aligned}$$

Step 3.: We denote by $\mathfrak{M}_A := \mathfrak{M}(\mu_A^*)$ the σ-algebra of measurable sets of μ_A^*. The restriction of μ_A^* to $\mathfrak{M}_A$ will be denoted by μ_A. We shall show, that this measure satisfies the statement of the theorem.

First, μ_A is a complete Radon measure. As we restricted ourselves to K compact, it is easy to see that $\mathcal{C}(K) \subset L_1(\mu_A)$. Hence, we only have to show that

$$Af = \int_K f d\mu_A, \quad f \in \mathcal{C}(K).$$

Let $f \in \mathcal{C}(K)$ be given. Without restriction, we may assume that $0 \le f \le 1$. For $n \in \mathbb{N}$ and $k = 0, 1, \dots, n$ we denote

$$f_k := \min(f, k/n), \quad G_k := \{x \in K : f(x) > k/n\}.$$

Using the definition of $\mu_A^*(G_{k-1})$ and $n(f_k - f_{k-1}) = 0$ on $K \setminus G_{k-1}$, we get

$$\le A(f_k - f_{k-1}) \le \frac{1}{n}\mu_A(G_{k-1})$$

for each $k = 1, \dots, n$. On the other hand, as $n(f_k - f_{k-1}) = 1$ on G_k, we first obtain $\mu_A(C) \le nA(f_k - f_{k-1})$ for every compact $C \subset G_k$, which implies also $\mu_A(G_k) \le nA(f_k - f_{k-1})$ for all $k = 1, \dots, n$.

As $\chi_{G_k} \le n(f_k - f_{k-1}) \le \chi_{G_{k-1}}$, we get also

$$\frac{1}{n}\mu_A(G_k) \le \int_K (f_k - f_{k-1}) d\mu_A \le \frac{1}{n}\mu_A(G_{k-1})$$

for each $k = 1, \dots, n$.

Thus (using $f_0 = 0$ and $f_n = f$)

$$\begin{aligned}\left|Af - \int_K f d\mu_A\right| &= \left|\sum_{k=1}^n \left(A(f_k - f_{k-1}) - \int_K (f_k - f_{k-1}) d\mu_A\right)\right|\\ &\le \sum_{k=1}^n \frac{1}{n}(\mu_A(G_{k-1}) - \mu_A(G_k)) = \frac{1}{n}\mu_A(G_0) = \frac{1}{n}\mu_A(\{x \in K : f(x) > 0\}).\end{aligned}$$

Since $\mu_A(\{x \in K : f(x) > 0\}) < \infty$ and n may be chosen arbitrarily large, we get $Af = \int_K f d\mu_A$.

Step 4.: Finally, we prove the uniqueness.

If ν is a complete Radon measure on K satisfying the statement of the theorem, we have

$$Af = \int_K f d\nu \le \chi_G d\nu = \nu(G)$$

for any open set $G \subset K$ and $f \in \mathcal{C}(K)$ with $0 \le f \le 1$, $f = 0$ on $K \setminus G$. Hence, $\mu_A(G) = \mu_A^*(G) \le \nu(G)$. On the other hand, by a similar argument and by the point (i) above, we have $\mu_A(C) \ge \nu(C)$ for each compact set $C \subset K$ and from regularity of μ_A and ν it immediately follows that they coincide on Borel sets, which concludes the assertion. $\square$

16.2 Lemma of Johnson and Lindenstrauss

In dieser Vorlesung werden einige Eigenschaften *Zufälliger Matrizen* untersucht. Das Hauptergebnis ist dann der Beweis des Lemmas von Johnson und Lindenstrauss.

Wir benutzen nur normalverteilte Zufallsvariablen $\mathcal{N}(0,1)$, d.h. Zufallsvariablen mit Wahrscheinlichkeitsdichte

$$p(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2}.$$

Lemma 16.2.1. *(i) Sei $\omega \sim \mathcal{N}(0,1)$. Dann gilt $\mathbb{E}\,(e^{\lambda\omega^2}) = 1/\sqrt{1-2\lambda}$ für $-\infty < \lambda < 1/2$.*

(ii) (2-Stabilität der Normalverteilung) Sei k eine natürliche Zahl, sei $\lambda = (\lambda_1, \dots, \lambda_k) \in \mathbb{R}^k$ und seien $\omega_1, \dots, \omega_k$ unabhängige identisch verteilte (u.i.v.) $\mathcal{N}(0,1)$ Zufallsvariablen. Dann gilt $\lambda_1\omega_1 + \dots + \lambda_k\omega_k \sim (\sum_{i=1}^k \lambda_i^2)^{1/2} \cdot \mathcal{N}(0,1)$.

Proof. Der Beweis von (i) folgt aus der Rechnung ($y := \sqrt{1-2\lambda} \cdot x$)

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\lambda x^2} \cdot e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{(\lambda - 1/2)x^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} \cdot \frac{dy}{\sqrt{1-2\lambda}} = \frac{1}{\sqrt{1-2\lambda}}.$$

Um (ii) zu beweisen, reicht es nur den Fall $k = 2$ zu betrachten. Der Rest folgt dann durch Induktion. Sei also $\lambda = (\lambda_1, \lambda_2) \in \mathbb{R}^2, \lambda \neq 0$ fest und seien ω_1 und ω_2 u.i.v. $\mathcal{N}(0,1)$ Zufallsvariablen. Wir setzen $S := \lambda_1\omega_1 + \lambda_2\omega_2$. Sei $t \in \mathbb{R}$ beliebig. Wir berechnen

$$\begin{aligned}
\mathcal{P}(S \le t) &= \frac{1}{2\pi} \int_{(x,y):\lambda_1 x + \lambda_2 y \le t} e^{-(x^2+y^2)/2} dx dy \\
&= \frac{1}{2\pi} \int_{x \le c; y \in \mathbb{R}} e^{-(x^2+y^2)/2} dx dy \\
&= \frac{1}{\sqrt{2\pi}} \int_{x \le c} e^{-x^2/2} dx.
\end{aligned}$$

In dem letzten Schritt haben wir die Rotationsinvarianz der Funktion $(x,y) \to e^{-(x^2+y^2)/2}$ ausgenutzt. Den Wert c berechnen wir aus der Skizze (!Tafel!) als die Länge der Normalen von Null auf die Gerade $\{(x,y) : \lambda_1 x + \lambda_2 y = t\}$. Es ergibt sich

$$c = \left\| \left(\frac{\lambda_1 t}{\lambda_1^2 + \lambda_2^2}, \frac{\lambda_2 t}{\lambda_1^2 + \lambda_2^2} \right) \right\| = \frac{t}{\sqrt{\lambda_1^2 + \lambda_2^2}}.$$

Wir erhalten also

$$\mathcal{P}(S \le t) = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{\lambda_1^2+\lambda_2^2} \cdot x \le t} e^{-x^2/2} dx = \mathcal{P}\left(\sqrt{\lambda_1^2 + \lambda_2^2} \cdot \omega \le t \right).$$

□

Falls $\omega_1, \dots, \omega_k$ (auch abhängige) normalverteilte Zufallsvariablen sind, dann gilt $\mathbb{E}(\omega_1^2 + \dots + \omega_k^2) = k$. Falls $\omega_1, \dots, \omega_k$ auch unabhängig sind, konzentriert sich der Wert von $\omega_1^2 + \dots + \omega_k^2$ auch sehr stark um k. Dieser Effekt (auch *Konzentration des Maßes* genannt) wird genau im nächsten Lemma beschrieben.

Lemma 16.2.2. *Sei k eine natürliche Zahl und seien $\omega_1, \dots, \omega_k$ unabhängige normalverteilte Zufallsvariablen. Sei $0 < \varepsilon < 1$. Dann gilt*

$$\mathcal{P}(\omega_1^2 + \dots + \omega_k^2 > (1+\varepsilon)k) \le e^{-\frac{k}{2}[\varepsilon^2/2 - \varepsilon^3/3]}$$

und

$$\mathcal{P}(\omega_1^2 + \dots + \omega_k^2 < (1-\varepsilon)k) \le e^{-\frac{k}{2}[\varepsilon^2/2 - \varepsilon^3/3]}.$$

Proof. Wir beweisen nur die erste Ungleichung. Der Beweis der zweiten Ungleichung folgt sehr ähnlich. Wir setzen $\beta := 1 + \varepsilon > 1$ und rechnen

$$\begin{aligned}\mathcal{P}(\omega_1^2 + \dots + \omega_k^2 > \beta k) &= \mathcal{P}(\omega_1^2 + \dots + \omega_k^2 - \beta k > 0) = \mathcal{P}(\lambda(\omega_1^2 + \dots + \omega_k^2 - \beta k) > 0)\\ &= \mathcal{P}(\exp(\lambda(\omega_1^2 + \dots + \omega_k^2 - \beta k)) > 1)\\ &\le \mathbb{E}\exp(\lambda(\omega_1^2 + \dots + \omega_k^2 - \beta k)),\end{aligned}$$

wobei $\lambda > 0$ eine beliebige positive reelle Zahl ist und erst später bestimmt wird. Wir haben im letzten Schritt die Markov-Ungleichung benutzt. Weiter benutzen wir die Eigenschaften der Exponentialfunktion und die Unabhängigkeit von $\omega_1, \dots, \omega_k$. Es ergibt sich

$$\mathbb{E}\exp(\lambda(\omega_1^2 + \dots + \omega_k^2 - \beta k)) = e^{-\lambda\beta k} \cdot \mathbb{E}\, e^{\lambda\omega_1^2} \cdots e^{\lambda\omega_k^2} = e^{-\lambda\beta k} \cdot (\mathbb{E}\, e^{\lambda\omega_1^2})^k$$

und mit Hilfe von Lemma 16.2.1 erhalten wir schließlich (für $0 < \lambda < 1/2$)

$$\mathbb{E}\exp(\lambda(\omega_1^2 + \dots + \omega_k^2 - \beta k)) = e^{-\lambda\beta k} \cdot (1-2\lambda)^{-k/2}.$$

Wir suchen jetzt einen Wert von $0 < \lambda < 1/2$, für den der letzte Ausdruck möglichst klein wird. Wir berechnen also die Ableitung von $e^{-\lambda\beta k} \cdot (1-2\lambda)^{-k/2}$ und setzen sie gleich Null. So erhalten wir

$$\begin{aligned}e^{-\lambda\beta k} \cdot (-\beta k) \cdot (1-2\lambda)^{-k/2} + e^{-\lambda\beta k} \cdot (-k/2) \cdot (1-2\lambda)^{-k/2-1} \cdot (-2) &= 0,\\ (-\beta) \cdot (1-2\lambda) + 1 &= 0,\\ 1 - 2\lambda &= 1/\beta,\\ \lambda &= \frac{1-1/\beta}{2}.\end{aligned}$$

Wir sehen auch sofort, dass für diesen Wert von λ wirklich die Abschätzungen $0 < \lambda < 1/2$ erfüllt sind. Wir benutzen diesen Wert von λ und erhalten

$$\begin{aligned}\mathcal{P}(\omega_1^2 + \dots + \omega_k^2 > \beta k) &\le e^{-\frac{1-1/\beta}{2}\cdot\beta k} \cdot (1 - (1-1/\beta))^{-k/2} = e^{-\frac{\beta-1}{2}k} \cdot \beta^{k/2}\\ &= e^{-\frac{\varepsilon k}{2}} \cdot e^{\frac{k}{2}\ln(1+\varepsilon)}.\end{aligned}$$

Das Ergebnis folgt dann aus der Ungleichung

$$\ln(1+t) \le t - \frac{t^2}{2} + \frac{t^3}{3}, \quad -1 < t < 1.$$

□

Remark 16.2.3. Überlegen Sie sich, dass die Unabhängigkeit von $\omega_1, \dots, \omega_k$ in Lemma 16.2.2 nicht weggelassen werden darf.

Lemma 16.2.4 (Johnson und Lindenstrauss, 1984)**.** *Sei $0 < \varepsilon < 1$ und seien k, n und d natürliche Zahlen mit*

$$k \geq 4(\varepsilon^2/2 - \varepsilon^3/3)^{-1} \ln n.$$

Dann existiert für jede Menge $\mathcal{P} = \{x_1, \dots, x_n\}$ eine Abbildung $f : \mathbb{R}^d \to \mathbb{R}^k$, so dass

$$(1-\varepsilon)\|x_i - x_j\|_2^2 \leq \|f(x_i) - f(x_j)\|_2^2 \leq (1+\varepsilon)\|x_i - x_j\|_2^2, \qquad i, j \in \{1, \dots, n\}. \tag{16.1}$$

Proof. Wir setzen

$$f(x) = \frac{1}{\sqrt{k}} \cdot Mx = \frac{1}{\sqrt{k}} \begin{pmatrix} \omega_{1,1} & \dots & \omega_{1,d} \\ \vdots & \ddots & \vdots \\ \omega_{k,1} & \dots & \omega_{k,d} \end{pmatrix} x,$$

wobei $(\omega_{u,v})$ u.i.v. $\mathcal{N}(0,1)$ Zufallsvariablen sind. Wir zeigen, dass diese Wahl von f mit positiver Wahrscheinlichkeit (16.1) erfüllt. Damit wäre die Existenz einer solchen Abbildung bewiesen.

Sei $i, j \in \{1, \dots, n\}$ beliebig mit $x_i \neq x_j$. Dann setzen wir $y = \frac{x_i - x_j}{\|x_i - x_j\|_2}$ und berechnen die Wahrscheinlichkeit, dass die rechte Ungleichung in (16.1) *nicht* erfüllt ist. Wir benutzen die 2-Stabilität der Normalverteilung und erhalten

$$\begin{aligned}
&\mathcal{P}(\|f(x_i) - f(x_j)\|_2^2 > (1+\varepsilon)\|x_i - x_j\|_2^2) = \mathcal{P}(\|f(y)\|^2 > (1+\varepsilon)) = \mathcal{P}(\|My\|_2^2 > (1+\varepsilon)k) \\
&\quad = \mathcal{P}((\omega_{1,1}y_1 + \dots + \omega_{1,d}y_d)^2 + \dots + (\omega_{k,1}y_1 + \dots + \omega_{k,d}y_d)^2 > (1+\varepsilon)k) \\
&\quad = \mathcal{P}(\omega_1^2 + \dots + \omega_k^2 > (1+\varepsilon)k)
\end{aligned}$$

für unabhängige normalverteilte $\omega_1, \dots, \omega_k$. Mit Hilfe von Lemma 16.2.2 läßt sich der letzte Ausdruck von oben durch

$$e^{-\frac{k}{2}[\varepsilon^2/2 - \varepsilon^3/3]}$$

abschätzen. Dieselbe Abschätzung gilt auch für die linke Ungleichung in (16.1) und für alle $\binom{n}{2}$ Paare $\{i, j\} \subset \{1, \dots, n\}$ mit $i \neq j$. Die Wahrscheinlichkeit, dass eine der Ungleichungen in (16.1) *nicht* gilt, ist also höchstens

$$2 \cdot \binom{n}{2} \cdot e^{-\frac{k}{2}[\varepsilon^2/2 - \varepsilon^3/3]} < n^2 \cdot e^{-\frac{k}{2}[\varepsilon^2/2 - \varepsilon^3/3]} = \exp(2 \ln n - \frac{k}{2}[\varepsilon^2/2 - \varepsilon^3/3]) \leq e^0 = 1$$

für $k \geq 4(\varepsilon^2/2 - \varepsilon^3/3)^{-1} \ln n$. □

Remark 16.2.5. Wie ändert sich die Wahrscheinlichkeit, dass $f = \frac{1}{\sqrt{k}} \cdot M$ die Formel (16.1) erfüllt, wenn k größer wird, d.h. wenn

$$k \geq c(\varepsilon^2/2 - \varepsilon^3/3)^{-1} \ln n$$

mit $c > 4$?

Remark 16.1. Der Beweis des Lemmas 16.2.4 benutzt die sogenannte *Probabilistische Methode*: Die Existenz von einem Objekt (in unserem Fall der Abbildung $f : \mathbb{R}^d \to \mathbb{R}^k$ mit (16.1)) wird dadurch bewiesen, dass man zeigt, dass zufällig konstruierte Objekte mit positiver Wahrscheinlichkeit die gewünschten Eigenschaften erfüllen. Als Urvater dieser Methode wird meistens *Paul Erdös* bezeichnet, es wurde aber auch schon vorher benutzt. Bis jetzt gibt es keine konstruktive Beweise von Lemma 16.2.4.

Der ursprüngliche Beweis von Johnson und Lindenstrauss ist viel mehr geometrisch. Man beweist, dass eine (richtig normierte) Projektion auf einen zufälligen k-dim. Unterraum von $\mathbb{R}^d$ mit positiver Wahrscheinlichkeit (16.1) erfüllt.

Seit ca. 10-15 Jahren gibt es viele Varianten von diesem Beweis (meistens durch spezifische Anwendungen motiviert).

i) Die Auswertung von $f(x)$ soll möglichst schnell sein (in unserem Fall nur $k \times d$, sonst $d \ln(k)$ durch *FFT, Fast Fourier Transform*)
ii) Die Anzahl der Zufallsvariablen soll minimiert werden (in unserem Fall $k \times d$)
iii) Die Implementierung soll möglichst einfach sein
iv) u.s.w.

References

[1] H. W. Alt, *Lineare Funktionalanalysis. Eine anwendungsorientierte Einführung*, 6. Aufl., Springer, 2012

[2] J. B. Conway, *A course in functional analysis*, 2. Aufl., Springer, 1990.

[3] H. Heuser, *Funktionalanalysis. Theorie und Anwendung*, 4. Aufl., Teubner, 2006

[4] W. Rudin, *Functional Analysis*, McGraw-Hill, 1991.

[5] D. Werner, *Funktionalanalysis*, 7. Aufl., Springer, 2011.

[6] O. Christensen, *Frames and bases: an introduction course*, Birkhäuser, 2008

[7] E. M. Stein, *Singular integrals and differentiability properties of functions*, Princeton Mathematical Series, No. 30 Princeton University Press, Princeton, N.J., 1970.

[8] E. M. Stein and G. Weiss, *Introduction to Fourier analysis on Euclidean spaces*, Princeton Mathematical Series, No. 32. Princeton University Press, Princeton, N.J., 1971.

[9] H. Triebel, *Higher analysis*, University Books for Mathematics, Johann Ambrosius Barth Verlag GmbH, Leipzig, 1992.

[10] D. D. Haroske and H. Triebel, *Distributions, Sobolev spaces, elliptic equations*, EMS Textbooks in Mathematics. European Mathematical Society (EMS), Zürich, 2008.

[11] K. Gröchenig, *Foundations of time-frequency analysis*, Applied and Numerical Harmonic Analysis. Birkhäuser Boston, Inc., Boston, MA, 2001.

[12] L. Grafakos, *Classical Fourier analysis*, Second edition, Graduate Texts in Mathematics, 249. Springer, New York, 2008.

[13] L. Grafakos, *Modern Fourier analysis*, Second edition, Graduate Texts in Mathematics, 250. Springer, New York, 2009.

[14] J. Lukeš and J. Malý, *Measure and Integral*, Matfyzpress, 2005; available at "http://www.mff.cuni.cz/fakulta/mfp/download/books/lukes-maly_-_measure_and_integral.pdf".