Introduction to Oscillatory Integrals and Pseudodifferential Operators

Lecture Notes

Max Lein

January 20, 2021

WPI-AIMR
Tohoku University

Contents

1 Chapter 1
Introduction

Integration is a fundamental aspect of analysis, initially as the "inverse" of differentiation. With time aspiring mathematicians learn there is not a single notion of integral, but several interrelated ones. The purpose of this course is to compare classical integrals, most notably the Riemann and Lebesgue integrals, with oscillatory integrals. Let us quickly review the first two before we proceed to the third.

1.1 The Riemann integral

The first notion of integral students of mathematics typically encounter is the so-called *Riemann integral*. It defines the integral of a function f as the limit of upper or lower sums, *provided* both limits exist and coincide; we will briefly review Riemann integration and its shortcomings in Chapter 2.

The price to pay for the simplicity of definition of Riemann integration (at least on $\mathbb{R}$ and cubes on $\mathbb{R}^d$) is that only a relatively small class of functions is Riemann integrable. Requiring that limits of upper and lower sums exist and agree puts a strong constraint on the variation of the function.

1.2 The Lebesgue integral

That is the main motivation to introduce the second notion of integration, the *Lebesgue integral*, which approaches integration from the measure theoretic end. Not only does this generalize the notion of integrability beyond integration of suitable subsets on $\mathbb{R}^d$, rougher functions are integrable. Just like with sums, to be able to declare the sum of the integrals as the integral of the sum of the integrands,

$$\int f + \int g = \int (f + g),$$

we need to work with *absolutely* integrable functions, i. e. those from

$$\mathcal{L}^1(\mathbb{R}^d) := \Big\{ f : \mathbb{R}^d \longrightarrow \mathbb{C} \text{ measurable } \;\Big|\; \int_{\mathbb{R}^d} \mathrm{d}x \, |f(x)| < \infty \Big\}.$$

Similarly, for $p \in [1, \infty)$ we can define the notion of p-integrable functions,

$$\mathcal{L}^p(\mathbb{R}^d) := \Big\{ f : \mathbb{R}^d \longrightarrow \mathbb{C} \text{ measurable } \;\Big|\; \int_{\mathbb{R}^d} \mathrm{d}x \, |f(x)|^p < \infty \Big\}.$$

These form vector spaces that can endowed with a seminorm

$$\|f\|_{\mathcal{L}^p(\mathbb{R}^d)} := \left(\int_{\mathbb{R}^d} \mathrm{d}x \, |f(x)|^p \right)^{1/p}.$$

Once we identify functions $f \sim g$ which coincide almost everywhere, we obtain the Banach spaces $L^p(\mathbb{R}^d)$ that commonly used in analysis and functional analysis.

There exist a lot of theory for this type of integrals. For example, Fubini's Theorem (cf. [LL01, Theorem 1.12]) tells us under what circumstances we are allowed to exchange the order of integration. Monotone and Dominated Convergence Theorems give us sufficient conditions when we are allowed to exchange limits and integration. These are the basis for things like parameter-dependent integrals and exchanging differentiation and integration.

1.3 Oscillatory integrals

To motivate why a *third* notion of integral is useful, we use an example that comes from operator theory. Here, one wants to *define* a so-called *pseudodifferential operator* via the integral

$$\big\langle \varphi, \mathrm{Op}(f)\psi \big\rangle := \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathrm{d}x \int_{\mathbb{R}^d} \mathrm{d}y \int_{\mathbb{R}^d} \mathrm{d}\eta \, \mathrm{e}^{-\mathrm{i}\eta \cdot (y-x)} \, \overline{\varphi(x)} \, f\big(\tfrac{1}{2}(x+y), \eta\big) \, \psi(y). \tag{1.3.1}$$

The functions $\varphi, \psi \in \mathcal{C}^\infty_{\mathrm{c}}(\mathbb{R}^d) \subset L^2(\mathbb{R}^d)$ are smooth functions with compact support and $f \in \mathcal{C}^\infty_{\mathrm{u,pol}}(\mathbb{R}^d \times \mathbb{R}^d)$ is a smooth, polynomially bounded function whose derivatives are bounded by a polynomial of fixed order. An example coming from physics would be the semirelativistic Hamilton function

$$f(x,\xi) = \sqrt{m^2 + \xi^2} + V(x), \qquad\qquad m > 0.$$

The associated operator $\mathrm{Op}(f)$ is then called the Weyl quantization of f.

After picking an orthonormal basis $\{\varphi_n\}_{n\in\mathbb{N}}$ of $L^2(\mathbb{R}^d)$, the matrix elements

$$\Big(\big\langle \varphi_j, \mathrm{Op}(f)\varphi_n \big\rangle \Big)_{j,n\in\mathbb{N}}$$

uniquely determine the operator $\mathrm{Op}(f)$, the so-called Weyl quantization of f [Hör79; Fol89; Rob87]. In fact, the diagonal elements $\big\langle \varphi_n, \mathrm{Op}(f)\varphi_n \big\rangle$ suffice to define the operator uniquely.

At closer inspection, the integral (1.3.1) is not well-defined as an absolutely convergent integral. For simplicity, let us consider the constant function $f(x,\xi) = 1$ first. At the end, we would like to show that $\mathrm{Op}(1) = \mathrm{id}_{L^2(\mathbb{R}^d)}$ is the identity operator. Assuming $\varphi, \psi \neq 0$, we start by taking absolute values of the right-hand side of

$$I(\varphi,\psi) := \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathrm{d}x \int_{\mathbb{R}^d} \mathrm{d}y \int_{\mathbb{R}^d} \mathrm{d}\eta \, \mathrm{e}^{-\mathrm{i}\eta\cdot(y-x)} \, \overline{\varphi(x)} \, \psi(y)$$

and pulling it into the two inner integral gives an integrand that is *independent* of the first integration variable η,

$$\begin{aligned} \left\lvert \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathrm{d}y \int_{\mathbb{R}^d} \mathrm{d}\eta \, \mathrm{e}^{-\mathrm{i}\eta\cdot(y-x)} \, \overline{\varphi(x)} \, \psi(y) \right\rvert &\leq \\ &\leq \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathrm{d}y \int_{\mathbb{R}^d} \mathrm{d}\eta \left\lvert \mathrm{e}^{-\mathrm{i}\eta\cdot(y-x)} \, \overline{\varphi(x)} \, \psi(y) \right\rvert \\ &= \frac{\lvert\varphi(x)\rvert}{(2\pi)^d} \left(\int_{\mathbb{R}^d} \mathrm{d}y \, \lvert\psi(y)\rvert \right) \int_{\mathbb{R}^d} \mathrm{d}\eta \\ &= \underbrace{\left(\frac{\lvert\varphi(x)\rvert}{(2\pi)^d} \, \lVert\psi\rVert_{L^1(\mathbb{R}^d)} \right)}_{<\infty} \underbrace{\int_{\mathbb{R}^d} \mathrm{d}\eta}_{=\infty} = \infty. \end{aligned} \tag{1.3.2}$$

Consequently, the integral $I(\varphi,\psi)$ does not exist as an absolutely convergent integral.

We may try to remedy this by integrating with respect to η *last*, but that is not of much help either. While for $\varphi, \psi \in \mathcal{C}_c^\infty(\mathbb{R}^d)$ the estimate

$$\begin{aligned} \left| \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathrm{d}x \int_{\mathbb{R}^d} \mathrm{d}y \, \mathrm{e}^{-\mathrm{i}\eta \cdot (y-x)} \, \overline{\varphi(x)} \, \psi(y) \right| &\leq \\ &\leq \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathrm{d}x \int_{\mathbb{R}^d} \mathrm{d}y \, \left| \mathrm{e}^{-\mathrm{i}\eta \cdot (y-x)} \, \overline{\varphi(x)} \, \psi(y) \right| \\ &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathrm{d}x \int_{\mathbb{R}^d} \mathrm{d}y \, \left| \varphi(x) \right| \left| \psi(y) \right| \\ &= (2\pi)^{-d} \, \| \varphi \|_{L^1(\mathbb{R}^d)} \, \| \psi \|_{L^1(\mathbb{R}^d)} < \infty \end{aligned}$$

is finite as our functions have compact support, the right-hand side is independent of η, and therefore cannot be integrable with respect to the last, remaining variable η.

Another approach that takes a step in the right direction is to keep the oscillating phase factor around. *Formally,* we can reorder the integrals, split the phase factor

$$\mathrm{e}^{-\mathrm{i}\eta \cdot (y-x)} = \overline{\mathrm{e}^{-\mathrm{i}\eta x}} \, \mathrm{e}^{-\mathrm{i}\eta \cdot y}$$

in two and distribute it amongst the factors to obtain

$$\begin{aligned} I(\varphi, \psi) &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathrm{d}x \int_{\mathbb{R}^d} \mathrm{d}y \int_{\mathbb{R}^d} \mathrm{d}\eta \, \overline{\mathrm{e}^{-\mathrm{i}\eta \cdot x} \varphi(x)} \, \mathrm{e}^{-\mathrm{i}\eta \cdot y} \psi(y) \\ &= \bigl\langle \mathcal{F}\varphi, \mathcal{F}\psi \bigr\rangle. \end{aligned}$$

Here, $\mathcal{F} : L^2(\mathbb{R}^d) \longrightarrow L^2(\mathbb{R}^d)$, is the Fourier transform on $L^2(\mathbb{R}^d)$. This is known to be a unitary [Lei20, Theorem 6.2.15], i. e.

$$I(\varphi, \psi) = \bigl\langle \mathcal{F}\varphi, \mathcal{F}\psi \bigr\rangle = \langle \varphi, \psi \rangle < \infty.$$

That seems to work as advertised and we were able to circumvent the problems above, correct? Not quite. The issue is that we have reordered the integral with the help of Fubini's Theorem [LL01, Theorem 1.12], and we were only allowed to do so if we know *a priori* that the original integral, where we integrate with respect to η last, is absolutely integrable. But our first estimate (1.3.2) tells us it is not. The second problem concerns the integral formula for the Fourier transform. We can write $\mathcal{F}$ as the integral

$$(\mathcal{F}\varphi)(\xi) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}x \, \mathrm{e}^{-\mathrm{i}\xi \cdot x} \, \varphi(x)$$

only when the function φ we want to Fourier transform is integrable. Given that there are functions $\varphi \in L^2(\mathbb{R}^d)$ that are square integrable, but *not* integrable, $\varphi \not\in L^1(\mathbb{R}^d)$, the integral above need not exist. Furthermore, the Fourier transform $\mathcal{F} : L^1(\mathbb{R}^d) \longrightarrow \mathcal{C}_\infty(\mathbb{R}^d) \subset L^\infty(\mathbb{R}^d)$ of absolutely integrable functions gives a function that is not necessarily integrable or square integrable. Indeed, defining the Fourier transform as a isometric map on $L^2(\mathbb{R}^d)$ is tricker and requires us to take various limits (cf. e. g. [LL01, Chapter 5.4] or [Lei20, Chapter 6.2.3]).

The last nail in the coffin is that our integral only nicely splits into a product of two integrals for certain "nice" functions such as $f(x,\xi) = \text{const.}$, $f(x,\xi) = K(\xi)$ or $f(x,\xi) = V(x)$. Perhaps we could deal with different functions on a case-by-case basis, but obviously a more general integration theory would be very desirable. These are *oscillatory integrals*.

What is nice is that all these formal manipulations, which are strictly forbidden without proper justification are completely ok if one works with oscillatory integrals. The purpose of this lecture is to contrast and compare the different types of integration and to show their utility with the example of a special class of operators, so-called pseudodifferential operators. 2020.10.07

Chapter 2
Classical integration theory

To better understand how oscillatory integrals differ from the integrals that have been introduced prior, we shall review "classical" integration theory, i. e. the Riemann and the Lebesgue integral. They differ in how we slice up the integral: to define the Riemann integral, we partition the domain and compare upper and lower sums. For a function to be Riemann integrable, its variation needs to be "well-behaved". In contrast, the Lebesgue integral looks at the level sets and measures their "volume". Lebesgue integrability allows the function to be less nicely behaved, but the price we have to pay is that the level sets need to be "nice" (measurable). Fortunately, this price is worth paying, since not only are "more" functions Lebesgue integrable than Riemann integrable, the Lebesgue integral plays nicer with limits.

2.1 The Riemann integral and its shortcomings

The starting point of the Riemann integral is a partition of the domain. While we can introduce Riemann integrals over $\mathbb{R}^d$, we shall stick to the one-dimensional case for simplicity. Suppose we are given a real-valued function $f : [a, b] \longrightarrow \mathbb{R}$ over an interval, and we would like to define

$$\int_a^b f.$$

The idea of the Riemann integral is to divvy up the interval

$$[a, b] = \bigcup_{k=0}^{n-1} [x_k, x_{k+1}]$$

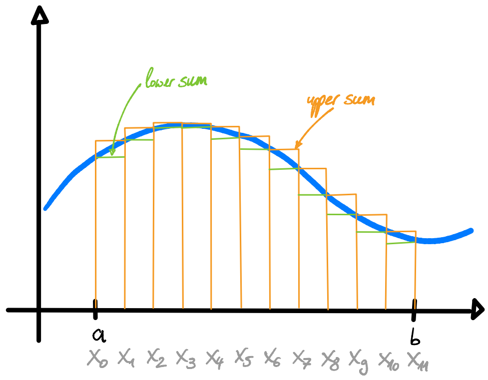

Figure 2.1.1: Upper and lower sums approximate the area under the curve from above and below. Unlike indicated in this picture, upper and lower sums need not converge to the same value, though.

into a *finite* collection of smaller intervals, where $x_k < x_{k+1}$ to avoid that intervals reduce to a single point and the endpoints $x_0 = a$ and $x_n = b$ are the endpoints of the original interval. We call such a finite sequence $P = (x_k)_{k=0}^n = \bigl(a = x_0, x_1, \ldots, x_n = b\bigr)$ of points, thought of as an ordered set, a *partition* of $[a, b]$. To each partition P we define upper and lower sums

$$U(f, P) := \sum_{k=0}^{n-1} \sup_{x \in [x_k, x_{k+1}]} f(x)\,(x_{k+1} - x_k), \tag{2.1.1a}$$

$$L(f, P) := \sum_{k=0}^{n-1} \inf_{x \in [x_k, x_{k+1}]} f(x)\,(x_{k+1} - x_k). \tag{2.1.1b}$$

We say the partition P is *finer* than a second partition P', or $P \leq P'$ for short, if and only if all x_k from P are also contained in P', i. e. $x_k = x'_j$ for all $k = 0, \ldots, n$ and some index j. Pictorially, the finer partition P is a subdivision of the coarser subdivision P'.

Note that we need not be able to compare two partitions: $P \leq P'$ implies all that all beginning and endpoints of the sub intervals of P' are beginning and endpoints of intervals of P. That is why the set of partitions forms a *partially ordered set*, i. e. a relation $P \leq P'$ that satisfies the following axioms:

(a) $P \leq P$ (*reflexivity*)

(b) $P \leq P'$ and $P' \leq P$ implies $P = P'$ (*antisymmetry*)

(c) $P \leq P'$ and $P' \leq P''$ implies $P \leq P''$ (*transitivity*)

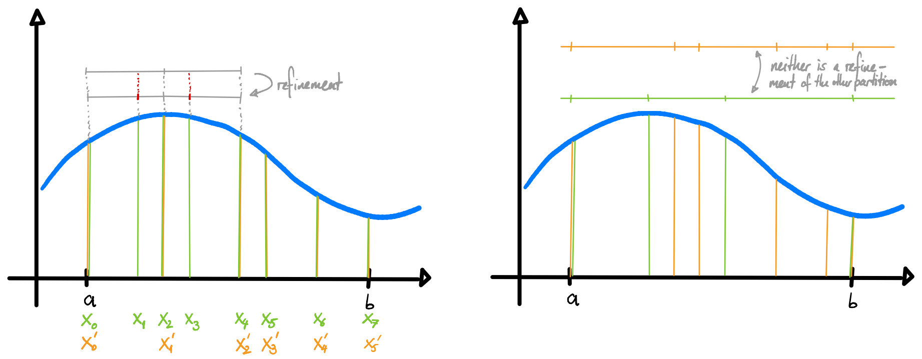

Figure 2.1.2: Partitions are ways to subdivide the interval $[a, b]$ we integrate over into smaller subintervals. Partitions can be refined by subdividing some of the subintervals into even smaller subintervals. However, two arbitrary partitions are usually not refinements of one another.

Even though we may not be able to compare P and P', there exists a partition Q such that $P \leq Q$ and $P' \leq Q$. The smallest such partition is $P \vee P'$ where we take all the x_k and x'_j, order them by size and discard duplicates to obtain another partition. This last property, makes the set of partitions into a *directed set*.

(d) For any two partitions P and P' there exists another partition Q with $P \leq Q$ and $P' \leq Q$ (*existence of an upper bound*).

Another way to quantify the size of a partition is to look at the length of the largest interval it contains, i. e. the *mesh* or norm,

$$\|P\| := \max_{k=0,\ldots,n-1} (x_{k+1} - x_k).$$

The value of the upper sum

$$U(f, P) \leq U(f, P') \qquad \qquad \forall P \leq P'$$

can only go down if we refine the partition, i. e. in this sense $U(f, P)$ is *non-increasing*. Similarly, the lower sum is *non-decreasing*,

$$L(f, P) \geq L(f, P') \qquad \qquad \forall P \leq P'.$$

Putting all inequalities together, we can summarize the situation as

$$L(f, P') \leq L(f, P) \leq U(f, P) \leq U(f, P') \qquad \qquad \forall P \leq P'. \tag{2.1.2}$$

Evidently, if f is bounded from above and below on this interval, then this would mean that upper and lower sums are bounded from below and above,

$$(b-a) \inf_{x\in[a,b]} f(x) \leq L(f,P) \leq U(f,P) \leq (b-a) \sup_{x\in[a,b]} f(x).$$

Now *if* we could formalize the limit

$$\lim_{\|P\|\to 0} U(f,P)$$

for the upper sum, we know this limit had to exist: non-increasing sequences that are bounded from below must have a limit. However, the notion of sequence is not general enough to make our ideas mathematically rigorous. Instead, we need to notion of

Definition 2.1.1 (Net) *(1) Let $\mathfrak{P}$ be a directed set, i. e. a set whose elements satisfy (a)–(d) above, and $\mathcal{X}$ a topological space*[1]*. Any function $f : \mathfrak{P} \longrightarrow \mathcal{X}$ is a net.*

(2) We say a net f converges to $g \in \mathcal{X}$,

$$\lim_P f(P) = g,$$

if and only if for every neighborhood U of $g \in \mathcal{X}$ the net $f(P)$ is eventually in U, i. e.

$$f(P) \in U \qquad\qquad \forall P \leq P_0.$$

Example (Sequences are a special case of nets) (1) To see that nets are a generalization of the notion of sequence, we need to check that sequences are nets. A sequence $(f_n)_{n\in\mathbb{N}}$ in a space $\mathcal{X}$ can be thought of as a function $f : \mathbb{N} \longrightarrow \mathcal{X}$. The natural number come equipped with a partial order through the usual $\leq$, although here the order is total (since for two positive integers $a \neq b$ we either have $a \leq b$ or $b \leq a$).

(2) A sequence $f_n \to g$ converges if and only if for any neighborhood U of g there exists N so that $f_n \in U$ holds for all $n \geq N$. (Note that perhaps confusingly, $\geq$ plays the role of $\leq$ in the above definition. The reason is that we have adopted notation for nets that imitates the limit $\varepsilon \to 0$ rather than the limit $n \to \infty$.)

With this definition in hand, we can define Riemann integrability.

[1] A topology defines the notion of open or, equivalently, the notion of closed sets. In metric spacers, the topology that is usually considered is generated by open balls. We will make this precise later in Definition 2.2.5.

Definition 2.1.2 (Riemann-integrability) *(1) We call a non-negative function $f : [a,b] \longrightarrow [0,\infty]$ Riemann-integrable if and only if the limits of the nets*

$$U(f) := \lim_P U(f,P),$$
$$L(f) := \lim_P L(f,P),$$

exist and agree, $U(f) = L(f)$. In that case we write

$$\int_a^b f := U(f) = L(f)$$

for the (Riemann) integral of f.

(2) A real-valued function $f : [a,b] \longrightarrow \mathbb{R}$ is called Riemann-integrable if and only if its positive part $f_+ := \max\{f, 0\}$ and its negative part $f_- := \max\{-f, 0\}$ are Riemann-integrable.

(3) A complex-valued function $f : [a,b] \longrightarrow \mathbb{C}$ is called Riemann-integrable if and only if its real and imaginary part are Riemann-integrable in the sense of (2).

Given that L and U are monotonous, equation (2.1.2), we can replace the limits of the nets with

$$L(f) = \sup_{P \in \mathfrak{P}} L(f,P),$$
$$U(f) = \inf_{P \in \mathfrak{P}} U(f,P).$$

So when f is a bounded function on $[a,b]$ the limits of upper and lower sum always exist; and when f is not, then $U(f) = +\infty$ and $L(f)$ may or may not be finite. That frees us from having to establish the existence of the limits that define $L(f)$ and $U(f)$.

Example (Integrating $f(x) = x^3$ on $[0,1]$ by hand) From calculus we already know how to integrate $\int_0^1 x^n = {}^1\!/_n$. But we would like to confirm $\int_0^1 x^3 = {}^1\!/_4$ by hand, using upper and lower sums. Given that $f(x)$ is strictly monotonous on $\mathbb{R}$, maxima are attained on the upper end of each interval and minima on the lower end,

$$\sup_{x \in [x_k, x_{k+1}]} f(x) = f(x_{k+1}) = x_{k+1}^3,$$
$$\inf_{x \in [x_k, x_{k+1}]} f(x) = f(x_k) = x_k^3.$$

Consequently, upper and lower sums are

$$U(f,P) = \sum_{k=0}^{n-1} x_{k+1}^3\,(x_{k+1}-x_k),$$
$$L(f,P) = \sum_{k=0}^{n-1} x_k^3\,(x_{k+1}-x_k).$$

Without loss of generality, we may assume that we are using equipartitions P_n, i. e. $x_k = {}^k\!/\!_n$. Then the interval width is a constant $x_{k+1} - x_k = {}^1\!/\!_n$, and upper and lower sums then simplify to

$$U(f,P_n) = \sum_{k=0}^{n-1} \frac{(k+1)^3}{n^4},$$
$$L(f,P_n) = \sum_{k=0}^{n-1} \frac{k^3}{n^4}.$$

These sums can be evaluated explicitly,

$$\sum_{k=0}^{n-1}(k+1)^3 = \sum_{k=1}^{n} k^3 = \frac{(n^2+n)^2}{4},$$

and in the limit as $n \to \infty$ the upper sum gives

$$\lim_{n\to\infty} U(f,P_n) = \lim_{n\to\infty} \frac{n^4+2n^3+n^2}{4n^4} = \frac{1}{4}.$$

Since upper and lower sum only differ by $U(f,P_n) - L(f,P_n) = {}^n\!/\!_{n^4} = {}^1\!/\!_{n^3} \to 0$, the lower sum attains the same limit.

While choosing an equipartition P_n was arbitrary, since a subnet of a convergent net must approach the same limit, we deduce $\int_0^1 x^3 = {}^1\!/\!_4$ as expected.[2]

The Riemann integral has a few fundamental properties that make it useful:

Proposition 2.1.3 *(1) The Riemann integral is linear, i. e. if f and g are integrable functions and $\mu \in \mathbb{R}$, then $f + \mu\, g$ is a Riemann integrable function, and the Riemann integral is given by*

$$\int_a^b (f+\mu\, g) = \int_a^b f + \mu \int_a^b g.$$

[2]The mathematical definition of a subnet is a bit more delicate since properties (a)–(d) must still hold for the subset of the directed set and must be big enough to accommodate the limit. But for our purposes it suffices to say that subnets generalize the notion of subsequence.

(2) For any Riemann-integrable function f and $a < c < b$ the Riemann integral splits

$$\int_a^b f = \int_a^c f + \int_c^b f.$$

(3) Continuous functions $f \in \mathcal{C}([a,b],\mathbb{R})$ are Riemann-integrable.

(4) Functions that are continuous everywhere on $[a,b]$ except for a finite *number of points are Riemann-integrable.*

(5) Non-decreasing and non-increasing functions are Riemann-integrable.

Proof We leave the proofs of (1)–(2) as an exercise to the readers.

To see that continuous functions are Riemann-integrable, item (3), we note that continuous functions on the compact interval $[a,b]$ are *uniformly* continuous. So by definition of uniform continuity, given any $\varepsilon > 0$ we can find a $\delta > 0$ so that $|f(x) - f(y)| < \varepsilon$ as long as $|x - y| < \delta$.

So let P be a partition whose mesh $\|P\| < \delta$ is fine enough. Then on each subinterval, the difference between

$$\begin{aligned} \sup_{x\in[x_k,x_{k+1}]} f(x) - \inf_{x\in[x_k,x_{k+1}]} f(x) &= \max_{x\in[x_k,x_{k+1}]} f(x) - \min_{x\in[x_k,x_{k+1}]} f(x) \\ &= \max_{x,y\in[a,b]} |f(x) - f(y)| < \varepsilon \end{aligned}$$

is at most ε. Consequently, the difference between upper and lower sums is also bounded by ε,

$$\begin{aligned} U(f,P) - L(f,P) &= \sum_{k=0}^{n-1} \Big(\sup_{x\in[x_k,x_{k+1}]} f(x) - \inf_{x\in[x_k,x_{k+1}]} f(x) \Big) \, (x_{k+1} - x_k) \\ &< \sum_{k=0}^{n-1} \varepsilon \, (x_{k+1} - x_k) = \varepsilon \, (b - a), \end{aligned}$$

since the sum is telescoping and only the first and last term survive. Seeing as $\varepsilon > 0$ can be chosen arbitrarily small, upper and lower sums converge to the same limit.

We leave the modifications to allow for finitely many discontinuities, item (4), to the readers.

Lastly, for item (5) we pick any partition P and estimate the difference between upper and lower sums. It suffices to treat only the non-decreasing case, the non-increasing case is analogous. Owing to the non-decreasing nature of the function, minima are attained on the left of any interval while maxima are necessarily on the right. Consequently, the difference reduces to a telescoping sum,

$$\begin{aligned} U(f,P) - L(f,P) &= \sum_{k=0}^{n-1} \big(f(x_{k+1}) - f(x_k)\big) \underbrace{(x_{k+1} - x_k)}_{\leq \|P\|} \\ &\leq \|P\| \sum_{k=0}^{n-1} \big(f(x_{k+1}) - f(x_k)\big) = \|P\| \, \big(f(b) - f(a)\big) \xrightarrow{\|P\| \to 0} 0. \end{aligned}$$

The right-hand side goes to 0 as we make the mesh finer and finer, and hence, the difference of upper and lower sums converges to 0. Given that lower (upper) sum are non-decreasing (non-increasing) as $\|P\| \to 0$ and bounded from above (below), upper and lower sums themselves converge. Therefore, f is Riemann-integrable. □

Example ($1_{\mathbb{Q}}$ is not Riemann-integrable) The assumption that f has only *finitely* many discontinuities in item (4) is essential. A classical counterexample is the indicator function on the rationals

$$f(x) := 1_{\mathbb{Q}}(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}.$$

Given that all subintervals contain both, countably infinite rational and uncountably infinite irrational numbers, the lower sum is always 0 while the upper sum is always 1,

$$L(1_{\mathbb{Q}}, P) = 0 \neq U(1_{\mathbb{Q}}, P) = 1.$$

So this function is not Riemann-integrable. It is, however, Lebesgue-integrable.

The Riemann integral can be extended in a variety of ways. For example, rather than looking at intervals $[a,b] \subset \mathbb{R}$ of the real line, we can look at compact subsets of $\mathbb{R}^d$. In that case, cubes $[a,b] := \prod_{j=1}^{d} [a_j, b_j]$ take the place of intervals. We can also define integrals of $\mathbb{C}$-valued functions by viewing any complex number $z = x + \mathrm{i}y \in \mathbb{C} \simeq \mathbb{R}^2$ as the sum of real and imaginary part. Similarly, extensions to $\mathbb{R}^n$ and $\mathbb{C}^n$ are straightforward. The definition of improper integrals, e. g. integrals over over $[0, \infty)$
2020.10.14 or the entire real line $\mathbb{R} = (-\infty, +\infty)$, is more subtle, but possible.

2.2 The Lebesgue integral

Riemann integration starts by slicing up the domain of integration "vertically" into subintervals ($d = 1$) or cubes ($d > 1$). The definition of the Riemann-integrability asks

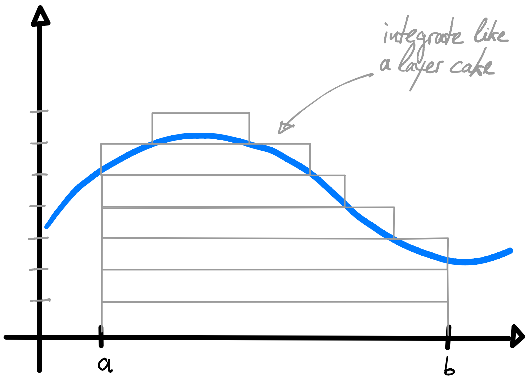

Figure 2.2.1: The Riemann integral corresponds to carving up the area under the curve vertically. Lebesgue proposed to slice horizontally and approximate the area under the curve by layers. Because of the resemblance, this is usually called the layer cake representation.

us to control the variation of the function. At the end of the day that places very strong restrictions on the class of functions that are Riemann-integrable.

What we would instead like to do now is slice our region of integrability "horizontally" and approximate the integral by a "layer cake". For simplicity, let us consider a $f : \mathbb{R} \longrightarrow [0, \infty)$ be a non-negative function. Then we would like to approximate the integral by finitely many layers via e. g. an upper sum of the form

$$\sum_{k=0}^{n-1} \mathrm{Vol}\Big(\big\{x \in \mathbb{R}^d \;\big|\; y_k \leq f(x) \leq y_{k+1}\big\}\Big)\, y_{k+1} \approx \int_{\mathbb{R}} \mathrm{d}x\, f(x).$$

Here, $\mathrm{Vol}(\Omega)$ measures the volume of the set $\Omega \subseteq \mathbb{R}^d$; for $d = 1$ and $\Omega = [a, b]$ the natural choice is to use the interval length $\mathrm{Vol}([a, b]) = b - a$. That almost looks like a Riemann integral for good reason, but at the same time clearly illustrates the trade-off: (1) we will need to find a "volume" function. And (2) the preimages of f need to produce "nice" sets that have a volume associated with them.

Mathematically, the idea of a volume map is implemented by a so-called *measure* and since the reader will likely encounter more general forms of measures in their career, we will give a short primer on measure theory.

2.2.1 A primer on measure theory

Assigning a "volume" to sets is not as easy as it looks at first glance, for example, it will turn out that not all sets can be measured. That is encapsulated in the so-called Banach-Tarski Paradox, which shows how to decompose a ball $B_R := \{x \in \mathbb{R}^3 \;\; | \;\; |x| \leq R\}$ into a *finite* number of (disjoint) pieces and reassemble them only to get *two* spheres. Put another way, this procedure doubles the volume. The resolution of this paradox is simple: our geometric intuition is betraying us, since the pieces we have cut the sphere into do not have a volume associated with them. So the cutting and reassembly procedure is "discontinuous" with respect to the volume measure.

2.2.1.1 Essential definitions

The starting point is the definition of σ-algebra and measurable space. In the following, we shall always assume that X is a set and $\mathfrak{P}(X)$ is the power set, i. e. the set of all subsets of X.

Definition 2.2.1 (σ-algebra and measure space) *A* σ-algebra $\Sigma \subseteq \mathfrak{P}(X)$ *is a set of subsets of* X *with the following three properties:*

(a) $\emptyset \in \Sigma$

(b) Σ *is closed under* countable *intersections, i. e. if* $\mathcal{I}$ *is a countable index set and* $\{A_j\}_{j \in \mathcal{I}} \subseteq \Sigma$*, then also* $\bigcap_{j \in \mathcal{I}} A_j \in \Sigma$.

(c) Σ *is closed under complements, i. e.* $A \in \Sigma$ *implies* $A^{\mathrm{c}} := X \setminus A \in \Sigma$.

Elements of Σ *are called* measurable sets *and the pair* (X, Σ) *is a* measurable space.

Remark 2.2.2 We could have equivalently replaced (b) with the assumption that

(b') Σ is closed under countable *unions*.

Example (The power set is a σ-algebra) It is an easy exercise to show that the power set is a σ-algebra. In fact, it is always the *largest* σ-algebra for a given set X.

(a) $\emptyset \in \mathfrak{P}(X)$ is contained in the power set.

(b) Countable intersects of subsets give another subset, which by definition is contained in $\mathfrak{P}(X)$.

(c) The complement of a set A is just another subset of X, and as such, an element of $\mathfrak{P}(X)$.

At this point we could ask: why do we need to introduce σ-algebras in the first place, why not work on the power set instead? Even for the Lebesgue measure, we cannot work on the power set, though. Even though I have not defined what a measure is at this point, the following theorem holds true:

Theorem 2.2.3 *The* only *measure $\mu : \mathfrak{P}(X) \longrightarrow [0,\infty]$ on the power set $\mathfrak{P}(\mathbb{R})$ that satisfies*

(a) $\mu\big([0,1)\big) < \infty$ *and*

(b) $\mu(x + A) = \mu(A)$ *(invariance under translations)*

is the trivial measure $\mu = 0$.

The proof is part and parcel of any course on measure theory worth its salt. For the purpose of this course, more important than the (rather instructive) proof is its content, namely that in general we cannot assign a volume to *all* subsets, only sufficiently nice ones. That is why we need to deal with σ-algebras. One nice fact about σ-algebras is as follows:

Theorem 2.2.4 *The intersection of* any *family of σ-algebras is again a Σ-algebra.*

That little factoid allows us to characterize σ-algebras by generating sets $S \subseteq \mathfrak{P}(X)$ as the σ-algebra

$$\Sigma(S) := \bigcap_{\substack{\Sigma \subseteq \mathfrak{P}(X)\ \sigma-\text{algebra} \\ S \subseteq \Sigma}} \Sigma$$

as the intersection of all σ-algebras containing S.

2.2.1.2 Comparison between σ-algebras and topology

It is useful and necessary to contrast and compare the notions of σ-algebra with the notion of topology.

Definition 2.2.5 (Topology) *A topology $\mathcal{O} \subseteq \mathfrak{P}(X)$ is the collection of sets which are by definition* open*, i. e. it is a set of subsets with the following properties:*

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

(b) *$\mathcal{O}$ is closed under* finite *intersections, whenever $O_1, \ldots, O_n \in \mathcal{O}$, then also $\bigcap_{j=1}^{n} O_j \in \mathcal{O}$.*

(c) *$\mathcal{O}$ is closed under* arbitrary *unions, i. e. for* any *$\mathcal{O}' \subseteq \mathcal{O}$ we have $\bigcup_{O \in \mathcal{O}'} O \in \mathcal{O}$.*

The pair $(X, \mathcal{O})$ *is called a* topological space.

Remark 2.2.6 (1) We have characterized a topology in terms of open sets. Equivalently, we could have singled out closed sets, neighborhoods or e. g. used Kuratowski's closure axioms.

(2) Certain sets like $\emptyset$ and X are closed and open (sometimes abbreviated with *clopen*) at the same time.

There is a huge zoology of topological spaces, e. g. Hausdorff spaces and Polish spaces, but we will not get into this here. One important fact is that a topology gives rise to a notion of convergence. A neighborhood of a point is a set that characterizes what is happening in the vicinity of the point.

Definition 2.2.7 (Neighborhoods) *Let* $(X, \mathcal{O})$ *be a topological space. A* neighborhood *of a point* $x \in X$ *is any set* U *so that*

(a) $x \in U$ *and*

(b) there exists an open set $O \in \mathcal{O}$ *with* $O \subseteq U$*.*

With that settled, we define convergent sequences as follows:

Definition 2.2.8 (Convergent sequences on topological spaces) *Let* $(X, \mathcal{O})$ *be a topological space. Then a sequence* $(x_n)_{n \in \mathbb{N}}$ *is said to converge to* $x \in X$ *if and only if for every neighborhood* U *of* x *there exists* $N \in \mathbb{N}$ *so that* $x_n \in O$ *for all* $n \in \mathbb{N}$*.*

Convergent nets are defined in an analogous fashion.

σ-algebras and topologies create very different structures. For example, functions which are compatible with respect to σ-algebras are called *measurable.*

Definition 2.2.9 (Measurable function) *Suppose* $f : X \longrightarrow Y$ *is a function between two measurable spaces* (X, Σ_X) *and* (Y, Σ_Y)*. We call* f *measurable if and only if preimages of* Σ_Y*-measurable sets are* Σ_X*-measurable,*

$$A_Y \in \Sigma_Y \implies f^{-1}(A_Y) \in \Sigma_X.$$

Functions which are compatible with respect to topologies are *continuous:*

Definition 2.2.10 (Continuous function) *Suppose* $f : X \longrightarrow Y$ *is a function between two topological spaces* $(X, \mathcal{O}_X)$ *and* $(Y, \mathcal{O}_Y)$*. We call* f *continuous if and only if preimages of neighborhoods in* Y *are neighborhoods in* X*,*

$$\text{neighborhood } N_Y \text{ of } f(x) \implies f^{-1}(N_Y) \text{ neighborhood of } x.$$

For e. g. metric spaces, this definition is equivalent to sequential continuity defined as follows: if $(x_n)_{n\in\mathbb{N}}$ is a sequence in X converging to $x \in X$, then $\big(f(x_n)\big)_{n\in\mathbb{N}}$ is a convergent sequence in Y that converges to $f(x)$,

$$\lim_{n\to\infty} f(x_n) = f\big(\lim_{n\to\infty} x_n\big).$$

For general topological spaces, sequential continuity need not be the same as continuity, though.

Very often, though, $(X, \mathcal{O})$ is a topological space on which we would like to define a σ-algebra that is compatible with the topology. In that case, we pick the so-called Borel σ-algebra.

Definition 2.2.11 *(1) Let $(X, \mathcal{O})$ be a topological space. Then the Borel σ-algebra $\mathfrak{B}(X) := \Sigma(\mathcal{O})$ is the σ-algebra generated from the open (or equivalently, the closed) sets.*

(2) The measurable space $\big(X, \mathcal{O}, \mathfrak{B}(X)\big)$ is called a Borel space.

(3) Measurable functions between Borel spaces are Borel functions.

Remark 2.2.12 We will frequently work with $\mathbb{R}$ and $\mathbb{C}$. Unless explicitly stated otherwise, we will always use the corresponding σ-algebras $\mathfrak{B}(\mathbb{R})$ and $\mathfrak{B}(\mathbb{C})$.

2.2.1.3 Measures

A measure is a non-negative function on the Σ-algebra that is compatible with the structure of a σ-algebra.

Definition 2.2.13 (Measure and measure space) *Let (X, Σ) be a measurable space. A measure is a function*

$$\mu : \Sigma \longrightarrow [0, \infty]$$

with the following properties:

(a) $\mu(\emptyset) = 0$

(b) μ is σ-additive, i. e. for a countable collection $\{A_j\}_{j\in\mathcal{I}} \subseteq \Sigma$ of mutually disjoint *sets, we have*

$$\mu\Big(\bigcup_{j\in\mathcal{I}} A_j\Big) = \sum_{j\in\mathcal{I}} \mu(A_j)$$

The triple (X, Σ, μ) is then called a measure space.

Furthermore, we distinguish between the following types of measures:

(1) We call μ trivial if and only if $\mu = 0$.

(2) We call μ finite *if and only if $\mu(X) < \infty$.*

(3) We call μ σ-finite if and only if there is a countable cover $\{X_j\}_{j\in\mathcal{I}}$, $\mathcal{I} \subseteq \mathbb{N}$, such that $X_j \in \Sigma$ and $\mu(X_j) < \infty$ for all $j \in \mathcal{I}$.

(4) We call μ a probability measure *if and only if $\mu(X) = 1$.*

Example (The Dirac measure on $\mathbb{R}$) Let $\Sigma \subseteq \mathfrak{P}(\mathbb{R})$ be any σ-algebra and define the Dirac measure at 0 as

$$\mu_{\text{Dirac}}(A) := \begin{cases} 1 & 0 \in A \\ 0 & 0 \not\in A \end{cases}$$

for any $A \in \Sigma$ that is measurable. Let us check whether μ_{Dirac} satisfies the axioms of a measure.

(a) $\mu_{\text{Dirac}}(\emptyset) = 0$ as $0 \not\in \emptyset$

(b) Let $\{A_j\}_{j\in\mathcal{I}}$ be a countable collection of *mutually disjoint* sets, where the index set is either finite, $\mathcal{I} \simeq \{1, \ldots, n\}$ or countably infinite, $\mathcal{I} \simeq \mathbb{N}$. Since the sets are mutually disjoint, then 0 lies in *at most* one of the sets.

When $0 \not\in \bigcup_{j\in\mathcal{I}} A_j$, then the σ-additivity condition amounts to $0 = 0$.

In the other case, when $0 \in \bigcup_{j\in\mathcal{I}} A_j$, then $0 \in A_n$ lies in exactly one A_n. Now σ-additivity reduces to $1 = 1$.

2020.10.21 Σ could be any σ-algebra, including the entire power set.

σ-additivity of measures has some powerful implications.

Theorem 2.2.14 *Let (X, Σ, μ) be a measure space. Then the following holds true:*

(1) $A \subseteq B$ implies $\mu(A) \leq \mu(B)$ for any $A, B \in \Sigma$ (monotonicity).

(2) $\mu(A_n) \to \mu(A)$ if $A_n \nearrow A$ (continuity from below)*, where $A_n \nearrow A$ means we are given a* non-decreasing *sequence of nested sets $A_n \subseteq A_{n+1}$ with $A = \bigcup_{n\in\mathbb{N}} A_n$.*

(3) $\mu(A_n) \to \mu(A)$ if $A_n \searrow A$ (continuity from above)*, where $A_n \searrow A$ means we are given a* non-increasing *sequence of nested sets $A_n \supseteq A_{n+1}$ with $A = \bigcap_{n\in\mathbb{N}} A_n$.*

What is often used to construct measures is that they are uniquely determined on a “sufficiently large” set, which generates the σ-algebra.

Theorem 2.2.15 (Uniqueness of measures) *Let $S \subseteq \Sigma$ be a collection of sets with the following properties:*

(a) S *generates* $\Sigma = \Sigma(S)$.

(b) S *is closed under* finite *intersections.*

(c) S *contains a sequence of increasing sets* $X_n \nearrow X$ *of finite measure,* $\mu(X_n) < \infty$.

Then μ *is uniquely determined by its values on* S.

The proof is a not difficult, but requires the notion of Dynkin system, which is defined just like a σ-algebra, but σ-additivity is only assumed to hold for countable collections of *mutually disjoint* sets. *Pre*measures are "measures on Dynkin systems", meaning they satisfy the exact same axioms as measures, just on Dynkin systems rather than σ-algebras. Other collections of sets enter in measure theory, e. g. semialgebras and algebras.

But within this context the purpose is always the same: suppose we are given a map $\widetilde{\mu} : \widetilde{\Sigma} \longrightarrow [0, \infty]$ on some suitable set of sets $\widetilde{\Sigma} \subseteq \mathfrak{P}(X)$. Under what circumstances does there exist a *unique* extension of $\widetilde{\mu}$ to a *bona fide* measure $\mu : \Sigma \longrightarrow [0, \infty]$ on a σ-algebra $\Sigma \supseteq \widetilde{\Sigma}$? This way it often suffices to specify measures on smaller generating sets. For example, the Lebesgue measure can be constructed from such an "almost-measure" on e. g. semiopen cubes of the form $[a, b) := \prod_{j=1}^{d} [a_j, b_j)$, where $a = (a_1, \ldots, a_d), b = (b_1, \ldots, b_d) \in \mathbb{R}^d$ and $a_j \leq b_j$ is assumed for all $j = 1, \ldots, d$. That is, specifying

$$\lambda\big([a, b)\big) := \prod_{j=1}^{d} (b_j - a_j)$$

completely suffices.

The constructions are not difficult, but require us to keep track of the subtle differences between premeasures, outer measures and measures. Given the time constraint, I refer the interersted reader to [Tes19, Chapter 8].

A particularly important role are sets of measure 0.

Definition 2.2.16 *Let* (X, Σ, μ) *be a measure space.*

(1) A null set *or* set of measure 0 *is a set* $N \in \Sigma$ *with* $\mu(N) = 0$. *The set of null sets is denoted with* $\mathcal{N}$.

(2) We say a condition holds μ*-almost everywhere if the set of exceptions is a null set. In case the measure can be implied from the context, we will shorten it to "almost everywhere". If* μ *is a probability measure, we will also say "almost surely" or "with probability* 1*".*

(3) The measure μ *is* complete *if* any *subset* $N' \subset N$ *of a null set* N *is again measurable and hence, a null set.*

These sets are important, because they allow us to quantify how often exceptions occur. Probably one of the better known ones is equality of functions, which will play a crucial role when we introduce L^p-spaces.

Example Later on we will identify functions $f \sim g$ that agree almost everywhere if there exists a null set N so that

$$f(x) = g(x)$$

holds for all $x \in X \setminus N$.

2.2.2 Integration with respect to measures

Measures determine a notion of volume, and we are almost in a position to make the layer cake representation rigorous. So in what follows, (X, Σ, μ) is a measure space, that we will frequently abbreviate with X when there is no risk of confusion as to what the σ-algebra and measure μ are. The starting point of introducing the integral is to define it for simple function. Moreover, $\mathbb{R}$ and $\mathbb{C}$ shall always be endowed with their standard Borel σ-algebras.

We will start with the smallest building block, the characteristic function

$$1_A(x) := \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

associated to a measurable set $A \in \Sigma$. This defines a *measurable* function (the proof is left as homework to the readers); to be precise, the target space of $1_A : X \longrightarrow \mathbb{R}, \mathbb{C}$ is either $\mathbb{R}$ or $\mathbb{C}$ endowed with the Borel σ-algebra.

The integral of characteristic functions is just the volume of the measurable set A,

$$\int_X 1_A := \int_X \mathrm{d}\mu(x)\, 1_A(x) := \mu(A).$$

Given that we want our integral to be linear, we can extend this definition to *simple functions*

$$f(x) = \sum_{k=1}^{n} c_k\, 1_{A_k}(x), \qquad\qquad c_k \in \mathbb{C},\ A_k \in \Sigma\ \forall k = 1, \ldots, n,$$

that are *finite* linear combinations of characteristic functions and set

$$\int_X \mathrm{d}\mu(x)\, f(x) := \sum_{k=1}^{n} c_k\, \mu(A_k).$$

However, as innocent as this may look, this definition has a serious flaw: what happens if some of the $\mu(A_k) = \infty$ have infinite measure? In the simplest case we have to decide what $4 \cdot \infty - 3 \cdot \infty$ is without running afoul of σ-additivity. Moreover, looming on the horizon is the topic of limits of simple functions. Essentially, we are presented with two options:

(1) We can restrict ourselves to simple functions where all $\mu(A_k) < \infty$.

(2) We can consider only simple functions where the coefficients $c_k \geq 0$ are non-negative.

Initially, either way may work, but once we want a lasting definition that is compatible with limits, we quickly see that the second option is much more tractable. In essence, we just have to worry about the summability of *non-negative* sequences, i. e. it is straightforward to declare what

$$\sum_{k=1}^{\infty} c_k \, \mu(A_k)$$

is even when the sum diverges to ∞ for one reason or another.

That definition also allows us to extend the definition of the Lebesgue integral to $\mathbb{R}$- and $\mathbb{C}$-valued functions: we merely require that $|f|$ is Lebesgue-integrable in the sense of non-negative functions. Behind the scenes, we are using the fact that even in the limit absolutely convergent sums retain linearity.

So for the moment, let us assume that $s : X \longrightarrow [0, \infty]$ is a *non-negative simple function* and $A \in \Sigma$ be some measurable set. Then we can define the integral as

$$\int_A \mathrm{d}\mu \, s := \sum_{j=1}^{n} c_j \, \mu\big(A_j \cap A\big). \tag{2.2.1}$$

We use the previously used conventions $n + \infty = \infty$, $n \cdot \infty = \infty$ and set $0 \cdot \infty := 0$.

The extension of this definition to a broader class of functions will give us the Lebesgue integral. So let us collect some properties first:

Lemma 2.2.17 *Let $c \geq 0$ and $s, s_1, s_2 \geq 0$ be non-negative, simple functions. Then the Lebesgue integral* (2.2.1) *has the following properties:*

(1) $\displaystyle\int_A \mathrm{d}\mu \, s = \int_X \mathrm{d}\mu \, (1_A \, s)$

(2) If $\{A_j\}_{j \in \mathcal{I}}$ be a countable collection of disjoint sets, the integral is σ-additive,

$$\int_{\bigcup_{j \in \mathcal{I}} A_j} \mathrm{d}\mu \, s = \sum_{j \in \mathcal{I}} \int_{A_j} \mathrm{d}\mu \, s.$$

(3) $\int_A \mathrm{d}\mu \, c \, s = c \int_A \mathrm{d}\mu \, s$

(4) $\int_A \mathrm{d}\mu \, (s_1 + s_2) = \int_A \mathrm{d}\mu \, s_1 + \int_A \mathrm{d}\mu \, s_2$

(5) $A \subseteq B$ *implies* $\int_A \mathrm{d}\mu \, s \leq \int_B \mathrm{d}\mu \, s$.

(6) $s_1 \leq s_2$ *implies* $\int_A \mathrm{d}\mu \, s_1 \leq \int_A \mathrm{d}\mu \, s_2$.

All these properties follow directly from the definition properties of measures, σ-algebras and simple functions.

Definition 2.2.18 (Lebesgue integral of non-negative functions) *Let $f : X \longrightarrow [0, \infty]$ be a non-negative, measurable function on the measure space (X, Σ, μ). The Lebesgue integral is defined as*

$$\int_X \mathrm{d}\mu \, f := \sup_{\substack{s \text{ simple function} \\ s \leq f}} \int_X \mathrm{d}\mu \, s. \tag{2.2.2}$$

When the supremum on the right is finite, we say the function is (Lebesgue-)integrable.

So of course, we can understand this limit in terms of nets, but we shall not spell out the details. It pays off that we are dealing with non-negative sequences, since then only two things can occur: either the supremum in (2.2.2) is ∞ or it is finite. To prove that we can really use the picture of level sets being added up, and approximating f by simple functions *point-wise*.

Moreover, all parts of Lemma 2.2.17 whose proofs do not make use of infinite sums under the hood, i. e. everything *except* for items (2) and (4), extend immediately to non-negative, measurable functions.

Corollary 2.2.19 *Lemma 2.2.17 (1), (3) and (5)–(6) extend to non-negative, measurable functions.*

To be sure, (2) and (4) are still correct, but we need the Monotone Convergence Theorem 2.2.20 in the proofs.

Theorem 2.2.20 (Monotone convergence/Beppo Levi) *Let f_n be a monotone non-decreasing sequence of non-negative measurable functions and $f_n \nearrow f$. Then f is measurable and we can exchange limit and integration,*

$$\lim_{n \to \infty} \int_A \mathrm{d}\mu \, f_n = \int_A \mathrm{d}\mu \, \lim_{n \to \infty} f_n.$$

Proof The measurability of the limit function $f = \lim_{n\to\infty} f_n = \sup_{n\in\mathbb{N}} f_n$ follows from the fact that lim inf, lim sup, inf and sup of sequences of measurable functions give measurable functions [Tes19, Lemma 8.19]. The idea is that

$$\begin{aligned} f^{-1}\big((a,\infty]\big) &= \Big(\sup_{n\in\mathbb{N}} f_n\Big)^{-1}\big((a,\infty]\big) \\ &= \bigcup_{n\in\mathbb{N}} \underbrace{f_n^{-1}\big((a,\infty]\big)}_{\in\Sigma} \in \Sigma \end{aligned}$$

is again measurable as a countable union of measurable sets. Given that we can recover all other Borel sets as countable unions or intersections of such half-open intervals, this extends to all Borel sets.

The proof uses a combination of Lemma 2.2.2 and a sequential squeeze of the form

$$\tau \int_A \mathrm{d}\mu \, \lim_{n\to\infty} f_n \leq \lim_{n\to\infty} \int_A \mathrm{d}\mu \, f_n \leq \int_A \mathrm{d}\mu \, \lim_{n\to\infty} f_n$$

where $\tau \in (0,1)$ can be chosen arbitrarily close to 1.

Lemma 2.2.2 (6) extends to non-negative, measurable functions, because $s \leq f_n \leq f$ implies $s \leq f$ as well. Consequently, $I_n := \int_A \mathrm{d}\mu \, f_n$ is a monotone, non-decreasing sequence,

$$I_n \leq I_{n+1} \leq \int_A \mathrm{d}\mu \, \lim_{n\to\infty} f_n = \int_A \mathrm{d}\mu \, f,$$

which therefore converges to some number. When $I_n \to \infty$ diverges, then the integral over f must also be ∞, and equality holds in this case.

Hence, we may assume $I_n \to I_\infty < \infty$ converges to some finite, non-negative value, and we only need to show the second estimate in our sequential squeeze argument. So pick a simple function $s \leq f$ and a constant $\tau \in (0,1)$. We define the sets

$$A_n := \big\{ x \in A \;\mid\; \tau\, s \leq f_n(x) \big\}.$$

Since the limiting function $f_n(x) \to f(x)$ is defined pointwise, these sets for a non-decreasing sequence $A_n \nearrow A$ that exhausts A from the inside. By definition of the set A_n and Lemma 2.2.2 (3) and (6), we obtain the estimates

$$\tau \int_{A_n} \mathrm{d}\mu \, s \leq \int_{A_n} \mathrm{d}\mu \, f_n \leq \int_A \mathrm{d}\mu \, f_n.$$

Seeing as the left-hand side only involves limits of simple functions, Lemma 2.2.2 (2) applies directly, which allows us to take $\lim_{n\to\infty}$ on both sides.

$$\tau \int_A \mathrm{d}\mu \, s \leq \lim_{n\to\infty} \int_A \mathrm{d}\mu \, f_n$$

and after taking the supremum over simple functions with $s \leq f$ gives us the second equality in our sequential squeeze. □

Corollary 2.2.21 *All of Lemma 2.2.17extends to non-negative, measurable functions.*

Example We can now *construct* a monotone sequence of simple functions $(s_n)_{n \in \mathbb{N}}$, which converges pointwise to f. The convergence is even uniform if f is bounded. For any $n \in \mathbb{N}$ we set

$$s_n(x) := \sum_{k=0}^{n\,2^n} \frac{k}{2^n} 1_{f^{-1}(A_k)}(x),$$

where up until $y = n$ we split the y-axis into 2^n equally sized intervals,

$$A_k := \left[\tfrac{k}{2^n}, \tfrac{k+1}{2^n}\right],$$

and subsume anything larger than that into the preimage of

$$A_{n\,2^n} := [n, \infty).$$

Example Extending Lemma 2.2.2 (2) and (6) to non-negative, measurable functions Since these involve infinite sums, we need monotone convergence. We leave these as exercises to the readers.

Now we are in a position to extend Lebesgue integration to complex-valued functions

Definition 2.2.22 (Integrable functions and $\mathcal{L}^1(X, \mathrm{d}\mu)$) *Suppose $f = \big(f_{\mathrm{Re}\,,+} - f_{\mathrm{Re}\,,-}\big) + \mathrm{i}\,\big(f_{\mathrm{Im}\,,+} - f_{\mathrm{Im}\,,-}\big)$ is the decomposition of a complex-valued function into four non-negative functions, and that $f_{\mathrm{Re}\,,\pm}$ and $f_{\mathrm{Im}\,,\pm}$ have finite Lebesgue integral. Then we define the Lebesgue integral of f to be*

$$\int_X \mathrm{d}\mu\, f := \int_X \mathrm{d}\mu\, f_{\mathrm{Re}\,,+} - \int_X \mathrm{d}\mu\, f_{\mathrm{Re}\,,-} + \mathrm{i} \int_X \mathrm{d}\mu\, f_{\mathrm{Im}\,,+} - \mathrm{i} \int_X \mathrm{d}\mu\, f_{\mathrm{Im}\,,-}.$$

The set of all complex-valued integrable functions is denoted with

$$\mathcal{L}^1(X, \mathrm{d}\mu) := \Big\{ f : X \longrightarrow \mathbb{C} \text{ measurable} \;\Big|\; \int_X \mathrm{d}\mu(x)\, |f(x)| < \infty \Big\}.$$

When the implied measure is clear, we may also write $\mathcal{L}^1(X)$ instead. Further, we write introduce the seminorm[3]

$$\|f\|_1 := \|f\|_{\mathcal{L}^1(X, \mathrm{d}\mu)} := \int_X \mathrm{d}\mu(x)\, |f(x)|.$$

[3] We have yet to define what a seminorm is and to prove that this is one. We will postpone this.

In a similar fashion, we can introduce the notion of p-integrable functions.

Definition 2.2.23 (p-integrable functions and $\mathcal{L}^p(X, \mathrm{d}\mu)$) *Let $1 \leq p < \infty$. Then $\mathcal{L}^p(X) \equiv \mathcal{L}^p(X, \mathrm{d}\mu)$ is the set of p-integrable functions, i. e.*

$$\mathcal{L}^p(X, \mathrm{d}\mu) := \Big\{ f : X \longrightarrow \mathbb{C} \text{ measurable} \;\Big|\; \int_X \mathrm{d}\mu(x) \, |f(x)|^p < \infty \Big\}.$$

The corresponding seminorm is

$$\|f\|_1 := \|f\|_{\mathcal{L}^1(X,\mathrm{d}\mu)} := \left(\int_X \mathrm{d}\mu(x) \, |f(x)|^p \right)^{1/p}$$

Let us collect a few important facts about the integral.

Lemma 2.2.24 *(1) The Lebesgue integral on $\mathcal{L}^1(X, \mathrm{d}\mu)$ is linear.*

(2) $\displaystyle \left| \int_X \mathrm{d}\mu(x) \, f(x) \right| \leq \int_X \mathrm{d}\mu(x) \, |f(x)|$

(3) $\|f + g\|_1 \leq \|f\|_1 + \|g\|_1$ *(triangle inequality)*

(4) $\|f\|_1 = 0$ *implies* $f(x) = 0$ *μ-almost everywhere.*

(5) Suppose $f, g \in \mathcal{L}^1(X, \mathrm{d}\mu)$ differ only on a set of measure 0. Then their integrals coincide, $\int_X \mathrm{d}\mu \, f = \int_X \mathrm{d}\mu \, g$.

Proof We will leave some of the proofs as an exercise for the readers. □

2.2.3 Fundamental facts on Lebesgue integrals

To study properties of $\mathcal{L}^p(X, \mathrm{d}\mu)$ spaces and the $L^p(X, \mathrm{d}\mu)$ spaces we will introduce below, one fundamental question is under what circumstances limits and integration commute. We have already gotten to know one of these theorems that played a crucial role when extending the definition of the integral to arbitrary, non-negative, measurable functions, namely the Monotone Convergence Theorem 2.2.20. Another one is

Theorem 2.2.25 (Fatou's Lemma) *If $A \in \Sigma$ is a measurable set and f_n is any sequence of non-negative, measurable functions, then for the* $\liminf$ *we have the estimate*

$$\int_A \mathrm{d}\mu \, \liminf_{n \to \infty} f_n \leq \liminf_{n \to \infty} \int_A \mathrm{d}\mu \, f_n .$$

Proof We define the sequence $g_n := \inf_{k \geq n} f_k$; by definition of $\liminf$, the sequence $(g_n)_{n \in \mathbb{N}}$ converges to

$$\lim_{n \to \infty} g_n = \lim_{n \to \infty} \inf_{k \geq n} f_k = \liminf_{n \to \infty} f_n$$

and is monotone non-decreasing.

Clearly, $g_n \leq f_n$ is dominated by f_n, which implies

$$\int_A \mathrm{d}\mu\, g_n \leq \int_A \mathrm{d}\mu\, f_n, \tag{2.2.3}$$

and that the integral on the left is non-degreasing as well.

Taking $\liminf_{n\to\infty}$ on the left yields with the help of the Monotone Convergence Theorem 2.2.20

$$\begin{aligned}
\liminf_{n\to\infty} \int_A \mathrm{d}\mu\, g_n &= \lim_{n\to\infty} \int_A \mathrm{d}\mu\, g_n \\
&= \int_A \mathrm{d}\mu\, \lim_{n\to\infty} g_n \\
&= \int_A \mathrm{d}\mu\, \liminf_{n\to\infty} f_n.
\end{aligned}$$

□

2020.10.28 So taking $\liminf_{n\to\infty}$ on both sides of (2.2.3) yields the claim.

Fatou's Lemma has a generalization that will be useful when proving one of the staples of classical analysis, the Dominated Convergence Theorem 2.2.27.

Corollary 2.2.26 (Generalized Fatou's Lemma) *Suppose $A \in \Sigma$ is a measurable set and $(f_n)_{n\in\mathbb{N}}$ is a sequence of $\mathbb{R}$-valued measurable functions and $g \in \mathcal{L}^1(X, \mathrm{d}\mu)$. Then the following holds true:*

(1) If $f_n \geq g$ is a uniform lower *bound, then we have*

$$\int_A \mathrm{d}\mu\, \liminf_{n\to\infty} f_n \leq \liminf_{n\to\infty} \int_A \mathrm{d}\mu\, f_n.$$

(2) If $f_n \leq g$ is a uniform upper *bound, then we have*

$$\int_A \mathrm{d}\mu\, \limsup_{n\to\infty} f_n \geq \limsup_{n\to\infty} \int_A \mathrm{d}\mu\, f_n.$$

(3) If $|f_n| \leq g$ is a uniform lower bound, then we have

$$\int_A \mathrm{d}\mu\, \liminf_{n\to\infty} f_n \leq \liminf_{n\to\infty} \int_A \mathrm{d}\mu\, f_n \leq \limsup_{n\to\infty} \int_A \mathrm{d}\mu\, f_n \leq \int_A \mathrm{d}\mu\, \limsup_{n\to\infty} f_n. \tag{2.2.4}$$

Proof We leave the proof to the readers as an exercise. □

Item (3) allows for a very compact proof of

Theorem 2.2.27 (Dominated Convergence) *Let $(f_n)_{n \in \mathbb{N}}$ be a convergent sequence in $\mathcal{L}^1(X)$ of measurable functions with limit $f := \lim_{n \to \infty}$. Suppose there exists a measurable function so that $|f_n(x)| \leq g(x)$ holds everywhere but possibly on a set of measure zero. Then $f \in \mathcal{L}^1(X)$ is integrable and we can exchange limit and integration,*

$$\lim_{n \to \infty} \int_A \mathrm{d}\mu \, f_n = \int_A \mathrm{d}\mu \, \lim_{n \to \infty} f_n. \tag{2.2.5}$$

Proof First of all, the assumptions of Corollary 2.2.26 (3) are satisfied. Moreover, since we assume that $f_n \to f$ converges pointwise, $\liminf$ and $\limsup$ coincide. Therefore, equation (2.2.4) is a sequential squeeze that directly implies (2.2.5). □

A classical example that shows the power of the Dominated Convergence Theorem is in the proof of the Riemann-Lebesgue Lemma, which states that the Fourier transform

$$(\mathcal{F}f)(\xi) := \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}x \, \mathrm{e}^{-\mathrm{i}x \cdot \xi} \, f(x)$$

maps $\mathcal{L}^1(\mathbb{R}^d)$ into $\mathcal{C}_\infty(\mathbb{R}^d)$, the continuous functions on $\mathbb{R}^d$ that vanish at ∞.

This is just one case of showing the continuity of parameter-dependent integrals in said parameter. Similarly, it is used to establish $\mathcal{C}^n$-regularity of parameter-dependent integrals since derivatives are also defined as limits.

2.3 L^p spaces

There is one deficiency in the integral, namely that a lot of properties are determined only up to a set of measure 0. For example, if we use the integral to define the "distance"

$$d(f,g) := \|f - g\|_1 = \int_X \mathrm{d}\mu(x) \, |f(x) - g(x)|$$

between two functions. Moreover, this notion of distance is compatible with the linear structure of $\mathcal{L}^p(X, \mathrm{d}\mu)$, and these are examples of a

Definition 2.3.1 (Normed space) *Let $\mathcal{X}$ be a vector space. A mapping $\|\cdot\| : \mathcal{X} \longrightarrow [0, +\infty)$ with properties*

(a) $\|x\| = 0$ *if and only if* $x = 0$*,*

(b) $\|\alpha x\| = |\alpha| \, \|x\|$*, and*

(c) $\|x + y\| \leq \|x\| + \|y\|$ *(triangle inequality),*

for all vectors $x, y \in \mathcal{X}$ and scalars $\alpha \in \mathbb{C}$, is called norm. The pair $(\mathcal{X}, \|\cdot\|)$ is then referred to as normed space.

Strictly speaking, we have only shown the triangle inequality for $p = 1$; for all other $1 < p < \infty$ this is known as Minkowski's inequality (cf. e. g. [Tes19, Corollary 10.8]).

However, in view of Lemma 2.2.24 (5) we can only ensure that

$$f(x) = g(x) \qquad\qquad \mu\text{-almost everywhere}$$

can have exceptions of a set of measure 0. It turns out, this is an incurable inaccuracy inherent in the integral norm. What we do now is simply forget about that difference "by force": we first define the set of functions

$$\begin{aligned}
\mathcal{N}(X, \mathrm{d}\mu) &:= \Bigl\{ f : X \longrightarrow \mathbb{C} \text{ measurable} \; \bigm| \; \int_X \mathrm{d}\mu(x) \, |f(x)| = 0 \Bigr\} \\
&= \bigl\{ f \in \mathcal{L}^1(X, \mathrm{d}\mu) \; \bigm| \; \|f\|_1 = 0 \bigr\} \\
&= \Bigl\{ f : X \longrightarrow \mathbb{C} \text{ measurable} \; \bigm| \; \int_X \mathrm{d}\mu(x) \, |f(x)|^p = 0 \Bigr\} \\
&= \bigl\{ f \in \mathcal{L}^p(X, \mathrm{d}\mu) \; \bigm| \; \|f\|_p = 0 \bigr\}
\end{aligned}$$

that have integral 0. Clearly, f and g agree almost everywhere exactly when the difference $f - g \in \mathcal{N}$ has integral 0. Importantly, we have shown in the homework assignments that this condition defines an equivalence relation, and the well-known

$$L^p(X, \mathrm{d}\mu) := \mathcal{L}^p(X, \mathrm{d}\mu) / \mathcal{N}(X, \mathrm{d}\mu)$$

are obtained as the quotient of the $\mathcal{L}^p(X, \mathrm{d}\mu)$ spaces. Its elements are equivalence classes of functions that agree almost everywhere. Strictly speaking, it makes no sense to write $\mathrm{e}^{-x^2} \in L^1(\mathbb{R}^d)$ since $x \mapsto \mathrm{e}^{-x^2}$ is a function or to claim that a particular element $f \in L^1(\mathbb{R}^d)$ is continuous. However, in practice the distinction between functions and equivalence classes of functions is almost never made. For example, the claim "$f \in L^1(\mathbb{R}^d)$ is continuous" is automatically translated to "$f \in L^1(\mathbb{R}^d)$ has a continuous representative".

The $L^p(X, \mathrm{d}\mu)$ spaces can be shown to be complete, the Riesz-Fischer Theorem (see e. g. [Tes19, Theorem 10.12]), and therefore are one of the prototypical Banach spaces.

Definition 2.3.2 (Banach space) *A complete normed space is a Banach space.*

Example The space $\mathcal{X} = \mathcal{C}\bigl([a, b], \mathbb{C}\bigr)$ of complex-valued, continuous functions on the interval $[a, b]$ has a norm, the sup norm

$$\|f\|_\infty = \sup_{x \in [a,b]} |f(x)| \, .$$

Since $\mathcal{C}\bigl([a, b], \mathbb{C}\bigr)$ is complete, it is a Banach space.

For the special case $p = 2$ we can introduce a notion of geometry through a scalar product

$$\langle f, g \rangle := \int_X \mathrm{d}\mu(x) \, \overline{f(x)} \, g(x),$$

which is well-defined for $f, g \in \mathcal{L}^2(X, \mathrm{d}\mu)$ and $f, g \in L^2(X, \mathrm{d}\mu)$ (with the aforementioned abuse of notation). These are then examples of a

Definition 2.3.3 (Pre-Hilbert space and Hilbert space) *A* pre-Hilbert space *is a complex vector space $\mathcal{H}$ with scalar product*

$$\langle \cdot \, , \cdot \rangle : \mathcal{H} \times \mathcal{H} \longrightarrow \mathbb{C},$$

i. e. a mapping that satisfies

(a) $\langle \varphi, \varphi \rangle \geq 0$ *and* $\langle \varphi, \varphi \rangle = 0$ *implies* $\varphi = 0$ *(positive definiteness),*

(b) $\langle \varphi, \alpha\psi + \chi \rangle = \alpha \, \langle \varphi, \psi \rangle + \langle \varphi, \chi \rangle$ *(linearity in the second argument), and*

(c) $\overline{\langle \varphi, \psi \rangle} = \langle \psi, \varphi \rangle$ *(antilinearity in the first argument)*

for all $\varphi, \psi, \chi \in \mathcal{H}$ and $\alpha \in \mathbb{C}$. This induces a norm $\lVert \varphi \rVert := \sqrt{\langle \varphi, \varphi \rangle}$ and metric $d(\varphi, \psi) := \lVert \varphi - \psi \rVert$ that measures the distance between $\varphi, \psi \in \mathcal{H}$. If $\mathcal{H}$ is complete with respect to the induced metric, it is a Hilbert space.

Example (1) $\mathbb{C}^d$ with the Euclidean scalar product

$$\langle z, w \rangle := \sum_{n=1}^{d} \overline{z_n} \, w_n$$

is a d-dimensional complex Hilbert space. In fact, every d-dimensional complex Hilbert space is isomorphic to it.

(2) $\mathcal{C}\big([-1, +1] \, , \mathbb{C}\big)$ with scalar product

$$\langle f, g \rangle := \int_{-1}^{+1} \mathrm{d}x \, \overline{f(x)} \, g(x)$$

is just a *pre*-Hilbert space, since it is not complete.

Consider for example the sequence of continuous functions

$$f_n(x) := \begin{cases} 0 & x \in [-1, 0] \\ nx & x \in (0, 1/n) \\ 1 & x \in [1/n, 1] \end{cases} ,$$

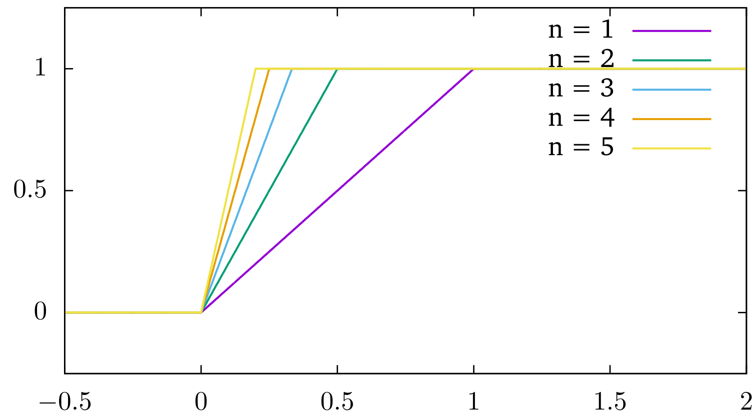

Figure 2.3.1: A sequence of continuous functions that converge pointwise to the step function.

whose graph looks like a ramp of ever increasing steepness (cf. Figure 2.3.1), connecting $y = 0$ with $y = 1$. We can check by hand that indeed $(f_n)_{n \in \mathbb{N}}$ forms a Cauchy sequence in $\mathcal{C}\bigl([-1,+1], \mathbb{C}\bigr)$. Moreover, away from $x = 0$ the *pointwise* limit exists,

$$f(x) := \lim_{n \to \infty} f_n(x) = \begin{cases} 0 & x < 0 \\ 1 & x > 0 \end{cases}.$$

But no matter how we define f at 0, though, it can never be continuous — and therefore, the limit of this particular sequence lies outside of the pre-Hilbert space $\mathcal{C}\bigl([-1,+1], \mathbb{C}\bigr)$.

We will give a few fundamental inequalities without proof:

Theorem 2.3.4 *(1) Let $p, q, r \in [1, \infty)$ such that $1/p + 1/q = 1/p$. Then there exists a constant $C_{p,q,r}$ depending only on these three real numbers such that for all $f \in L^p(X, \mathrm{d}\mu)$ and $g \in L^q(X, \mathrm{d}\mu)$ Hölder's Inequality holds true,*

$$\lVert f \, g \rVert_r \leq C_{p,q,r} \, \lVert f \rVert_p \, \lVert g \rVert_q .$$

(2) On the Hilbert space $L^2(X, \mathrm{d}\mu)$ the Cauchy-Schwartz inequality holds,

$$\lvert \langle \varphi, \psi \rangle \rvert \leq \lVert \varphi \rVert_2 \, \lVert \psi \rVert_2 .$$

There are plenty of other inequalities (see e. g. [LL01]), which are relevant in analysis. To name but one, Young's Inequality is roughly analogous to Hölder's Inequality, but involves the convolution.

Chapter 3

3 Bounded Linear Operators and Linear Functionals

Operators are generalizations of the concept of matrices to infinite-dimensional spaces, and it is therefore natural to try and extend ideas from linear algebra to functional analysis. A lot of physical theories, e. g. quantum mechanics and many classical wave equations, are formulated in terms of operators. Well-known examples are the quantum mechanical Schrödinger equation

$$\mathrm{i}\,\partial_t \psi(t) = \bigl(-\Delta + V\bigr)\,\psi(t)\,, \qquad \psi(0) = \psi_0 \in L^2(\mathbb{R}^d)\,, \tag{3.0.1}$$

and the heat equation

$$\partial_t \psi(t) = -\bigl(-\Delta + V\bigr)\,\psi(t)\,, \qquad \psi(0) = \psi_0 \in L^1(\mathbb{R}^d)\,, \tag{3.0.2}$$

which are both defined in terms of the Laplace operator $-\Delta := -\sum_{j=1}^d \partial_{x_j}^2$ and a potential $V : \mathbb{R}^d \longrightarrow \mathbb{R}$.

Mathematically speaking, operators $T : \mathcal{X} \longrightarrow \mathcal{Y}$ are continuous functions between normed vector spaces that are compatible with the linear structure. Compatible with the linear structure means that for any $x_1, x_2 \in \mathcal{X}$ and $\alpha \in \mathbb{C}$ we have

$$T(x_1 + \alpha x_2) = T x_1 + \alpha\, T x_2. \tag{3.0.3}$$

3.1 Bounded operators

The simplest class of operators are bounded.

Definition 3.1.1 (Bounded operator) *Let $\mathcal{X}$ and $\mathcal{Y}$ be normed spaces. A linear operator $T : \mathcal{X} \longrightarrow \mathcal{Y}$ is called bounded if there exists $M \geq 0$ so that*

$$\|Tx\|_{\mathcal{Y}} \leq M \, \|x\|_{\mathcal{X}} \, .$$

The set of bounded linear operators is denoted with $\mathcal{B}(\mathcal{X}, \mathcal{Y})$. When initial and target space are the same, $\mathcal{X} = \mathcal{Y}$, we abbreviate $\mathcal{B}(\mathcal{X}, \mathcal{X}) = \mathcal{B}(\mathcal{X})$.

A special case of bounded operators are linear functionals where

Definition 3.1.2 (Bounded linear functional) *A bounded linear functional $L : \mathcal{X} \longrightarrow \mathbb{C}$ is a linear operator where the target space is $\mathbb{C}$. The set of bounded linear functional is denoted with $\mathcal{X}' := \mathcal{B}(\mathcal{X}, \mathbb{C})$.*

For normed spaces, continuous and bounded are synonymous.

Theorem 3.1.3 *Let $T : \mathcal{X} \longrightarrow \mathcal{Y}$ be a linear operator between two normed spaces $\mathcal{X}$ and $\mathcal{Y}$. Then the following statements are equivalent:*

(1) T is continuous at $x_0 \in \mathcal{X}$.

(2) T is continuous.

(3) T is bounded.

Proof (1) $\Leftrightarrow$ (2): This follows immediately from the linearity.

(2) $\Rightarrow$ (3): Assume T to be continuous. Then it is continuous at 0 and for $\varepsilon = 1$, we can pick $\delta > 0$ such that

$$\|Tx\|_{\mathcal{Y}} < \varepsilon = 1$$

for all $x \in \mathcal{X}$ with $\|x\|_{\mathcal{X}} < \delta$. By linearity, this implies for any $x' \in \mathcal{X} \setminus \{0\}$ that

$$\left\|T\big(\tfrac{\delta}{2\|x'\|_{\mathcal{X}}} x'\big)\right\|_{\mathcal{Y}} = \tfrac{\delta}{2\|x'\|_{\mathcal{X}}} \left\|Tx'\right\|_{\mathcal{Y}} < 1.$$

Hence, T is bounded with bound $1/\delta$,

$$\left\|Tx'\right\|_{\mathcal{Y}} < \tfrac{2}{\delta} \, \|x'\|_{\mathcal{X}} \, .$$

(3) $\Rightarrow$ (2): Conversely, if T is bounded by $M > 0$,

$$\left\|Tx_1 - Tx_2\right\|_{\mathcal{Y}} \leq M \, \|x_1 - x_2\|_{\mathcal{X}} \qquad \forall x_1, x_2 \in \mathcal{X}.$$

This means, T is continuous: for $\varepsilon > 0$ we pick $\delta = \varepsilon/2M$ and verify that

$$\left\|Tx_1 - Tx_2\right\|_{\mathcal{Y}} \leq M \, \|x_1 - x_2\|_{\mathcal{X}} \leq M \tfrac{\varepsilon}{2M} < \varepsilon \qquad \square$$

2020.11.04 holds for all $x_1, x_2 \in \mathcal{X}$ that satisfy $\|x_1 - x_2\|_{\mathcal{X}} < \varepsilon/2M$.

Next, we introduce the operator norm, which leads to a notion of convergence for sequences of operators:

Definition 3.1.4 (Operator norm) *Let $T : \mathcal{X} \longrightarrow \mathcal{Y}$ be a bounded linear operator between normed spaces. We define the operator norm of T as*

$$\|T\| := \|T\|_{\mathcal{B}(\mathcal{X},\mathcal{Y})} := \sup_{\substack{x\in\mathcal{X} \\ \|x\|=1}} \|Tx\|_{\mathcal{Y}} . \tag{3.1.1}$$

The space of all bounded linear operators between $\mathcal{X}$ and $\mathcal{Y}$ is denoted by $\mathcal{B}(\mathcal{X},\mathcal{Y})$. When the initial and target spaces $\mathcal{X} = \mathcal{Y}$ coincide, we will use the abbreviation $\mathcal{B}(\mathcal{X}) := \mathcal{B}(\mathcal{X},\mathcal{X})$.

One can show that the operator norm $\|T\|$ coincides with

$$\inf \bigl\{ M \geq 0 \;\; \big| \;\; \|Tx\|_{\mathcal{Y}} \leq M \, \|x\|_{\mathcal{X}} \;\; \forall x \in \mathcal{X} \bigr\} = \|T\| .$$

The product of two bounded operators $T \in \mathcal{B}(\mathcal{Y},\mathcal{Z})$ and $S \in \mathcal{B}(\mathcal{X},\mathcal{Y})$ is again a bounded operator whose norm can be estimated from above by

$$\|TS\| \leq \|T\| \, \|S\| .$$

When $\mathcal{Y} = \mathcal{X} = \mathcal{Z}$, this implies that the product is jointly continuous with respect to the norm topology on $\mathcal{X}$.

Let T, S be bounded linear operators between the normed spaces $\mathcal{X}$ and $\mathcal{Y}$. Once we define addition,

$$(T + S)x := Tx + Sx,$$

and scalar multiplication,

$$(\lambda T)x := \lambda T x,$$

the set of bounded linear operators forms a vector space. Once endowed with the norm (3.1.1), we obtain a normed vector space.

Proposition 3.1.5 *The vector space $\mathcal{B}(\mathcal{X},\mathcal{Y})$ of bounded linear operators between normed spaces $\mathcal{X}$ and $\mathcal{Y}$ with operator norm* (3.1.1) *forms a normed space. If $\mathcal{Y}$ is complete, $\mathcal{B}(\mathcal{X},\mathcal{Y})$ is a Banach space.*

Proof Let $(T_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathcal{B}(\mathcal{X},\mathcal{Y})$, i. e. a sequence for which

$$\|T_n - T_k\|_{\mathcal{B}(\mathcal{X},\mathcal{Y})} \xrightarrow{n,k\to\infty} 0.$$

We have to show that $(T_n)_{n\in\mathbb{N}}$ converges to some $T \in \mathcal{B}(\mathcal{X},\mathcal{Y})$. For any $\varepsilon > 0$, there exists $N(\varepsilon) \in \mathbb{N}$ such that we have

$$\|T_n - T_k\|_{\mathcal{B}(\mathcal{X},\mathcal{Y})} < \varepsilon$$

for all $n, k \geq N(\varepsilon)$. This also implies that for any $x \in \mathcal{X}$, the sequence $\bigl(T_n(x)\bigr)_{n\in\mathbb{N}}$ converges as well,

$$\bigl\|T_n x - T_k x\bigr\|_{\mathcal{Y}} \leq \|T_n - T_k\|_{\mathcal{B}(\mathcal{X},\mathcal{Y})} \, \|x\|_{\mathcal{X}} < \varepsilon \, \|x\| \, .$$

The field of complex numbers is complete and $\bigl(T_n(x)\bigr)_{n\in\mathbb{N}}$ converges to some $Tx \in \mathcal{Y}$. We now *define*

$$Tx := \lim_{n\to\infty} T_n x \in \mathcal{Y}$$

for any $x \in \mathcal{X}$. Clearly, T inherits the linearity of the $(T_n)_{n\in\mathbb{N}}$. The map T is also bounded: for any $\varepsilon > 0$, there exists $N(\varepsilon) \in \mathbb{N}$ such that $\|T_k - T_n\|_{\mathcal{B}(\mathcal{X},\mathcal{Y})} < \varepsilon$ is small when $k, n \geq N(\varepsilon)$. Then for $n \geq N(\varepsilon)$ the estimate

$$\begin{aligned} \bigl\|(T - T_n)x\bigr\|_{\mathcal{B}(\mathcal{X},\mathcal{Y})} &= \lim_{k\to\infty} \bigl\|(T_k - T_n)x\bigr\|_{\mathcal{Y}} \leq \lim_{k\to\infty} \|T_k - T_n\|_{\mathcal{B}(\mathcal{X},\mathcal{Y})} \, \|x\|_{\mathcal{X}} \\ &< \varepsilon \, \|x\| \end{aligned}$$

holds true. Since we can write T as $T = T_n + (T - T_n)$, we can estimate the norm of the linear map T by $\|T\|_{\mathcal{B}(\mathcal{X},\mathcal{Y})} \leq \|T_n\|_{\mathcal{B}(\mathcal{X},\mathcal{Y})} + \varepsilon < \infty$. This means T is a bounded linear operator mapping $\mathcal{X}$ to $\mathcal{Y}$. □

Note that many linear operators are *not* bounded, i. e. they need not necessarily be defined on the entire normed space $\mathcal{X}$, but only on a dense subspace $\mathcal{D} \subseteq \mathcal{X}$,

$$T : \mathcal{D}(T) \subseteq \mathcal{X} \longrightarrow \mathcal{Y}.$$

Borrowing terminology from the analysis of functions, $\mathcal{D}$ is referred to as domain. An example of an unbounded operator is the Laplacian $-\Delta$ from above. Without specifying a domain, the definition of an unbounded operator is incomplete. Unless specifically mentioned otherwise, we shall always assume that T is densely defined.

There are also cases where it is initially easier to define an operator on a dense subspace, even though the operator is bounded. An example is the Fourier transform

$$(\mathcal{F}f)(\xi) := \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}x \, \mathrm{e}^{-\mathrm{i}x\cdot\xi} \, f(x), \tag{3.1.2}$$

on $L^2(\mathbb{R}^d)$. *A priori* this integral only makes sense for integrable functions, i. e. on $L^1(\mathbb{R}^d)$. However, there are integrable functions, which are not square integrable and vice versa; that is, we neither have the inclusion $L^1(\mathbb{R}^d) \not\subseteq L^2(\mathbb{R}^d)$ nor the opposite $L^2(\mathbb{R}^d) \not\subseteq L^1(\mathbb{R}^d)$. Fortunately, the intersection $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ is dense in $L^2(\mathbb{R}^d)$ and $\mathcal{F}$ is bounded on that dense subset. Consequently, the following theorem tells us the Fourier transform $\mathcal{F} : L^2(\mathbb{R}^d) \longrightarrow L^2(\mathbb{R}^d)$ defines a bounded operator, which extends (3.1.2) to all of $L^2(\mathbb{R}^d)$.

Theorem 3.1.6 *Suppose we are given a densely defined, bounded linear operator $T : \mathcal{D}(T) \subseteq \mathcal{X} \longrightarrow \mathcal{Y}$ between a normed space $\mathcal{X}$ and a Banach space $\mathcal{Y}$. Then there exists a unique bounded linear extension $\tilde{T} : \mathcal{X} \longrightarrow \mathcal{Y}$ and*

$$\|\tilde{T}\| = \|T\|_{\mathcal{D}} := \sup_{x \in \mathcal{D}\setminus\{0\}} \frac{\|Tx\|_{\mathcal{Y}}}{\|x\|_{\mathcal{X}}} \tag{3.1.3}$$

Proof We construct $\tilde{T}$ explicitly: let $x \in \mathcal{X}$ be arbitrary. Since $\mathcal{D}$ is dense in $\mathcal{X}$, there exists a sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{D}$ which converges to x. Then we set

$$\tilde{T}x := \lim_{n\to\infty} Tx_n.$$

First of all, we note $\tilde{T}$ inherits the linearity from T. It is also well-defined: $(Tx_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in the target space $\mathcal{Y}$,

$$\big\|Tx_n - Tx_k\big\|_{\mathcal{Y}} \leq \|T\|_{\mathcal{D}} \, \|x_n - x_k\|_{\mathcal{X}} \xrightarrow{n,k\to\infty} 0,$$

where the norm of T is defined by the right-hand side of equation (3.1.3). This Cauchy sequence in $\mathcal{Y}$ converges to some unique $y \in \mathcal{Y}$ as the target space is complete. Let $(x_n')_{n\in\mathbb{N}}$ be a second sequence in $\mathcal{D}$ that converges to x and assume the sequence $(Tx_n')_{n\in\mathbb{N}}$ converges to some $y' \in \mathcal{Y}$. We define a third sequence $(z_n)_{n\in\mathbb{N}}$ which alternates between elements of the first sequence $(x_n)_{n\in\mathbb{N}}$ and the second sequence $(x_n')_{n\in\mathbb{N}}$, i. e.

$$\begin{aligned} z_{2n-1} &:= x_n \\ z_{2n} &:= x_n' \end{aligned}$$

for all $n \in \mathbb{N}$. Then $(z_n)_{n\in\mathbb{N}}$ also converges to x and $\big(Tz_n\big)_{n\in\mathbb{N}}$ forms a Cauchy sequence that converges to, say, $\zeta \in \mathcal{Y}$. Subsequences of convergent sequences are also convergent, and they must converge to the same limit point. Since the limits of the sequence and its even and odd subsequences converge to the same limit, we conclude that

$$\begin{aligned} \zeta &= \lim_{n\to\infty} Tz_n = \lim_{n\to\infty} Tz_{2n} = \lim_{n\to\infty} Tx_n = y \\ &= \lim_{n\to\infty} Tz_{2n-1} = \lim_{n\to\infty} Tx_n' = y' \end{aligned}$$

all agree. Consequently, $\tilde{T}x$ does not depend on the particular choice of sequence which approximates x in $\mathcal{D}$. It remains to show the equality of the norms $\|\tilde{T}\| = \|T\|_{\mathcal{D}}$: we can calculate the norm of $\tilde{T}$ on the dense subset $\mathcal{D}$ and use that $\tilde{T}|_{\mathcal{D}} = T$ to obtain

$$\begin{aligned} \|\tilde{T}\| &= \sup_{\substack{x \in \mathcal{X} \\ \|x\|=1}} \|\tilde{T}x\| = \sup_{x \in \mathcal{X}\setminus\{0\}} \frac{\|\tilde{T}x\|}{\|x\|} = \sup_{x \in \mathcal{D}\setminus\{0\}} \frac{\|\tilde{T}x\|}{\|x\|} \\ &= \sup_{x \in \mathcal{D}\setminus\{0\}} \frac{\|Tx\|}{\|x\|} = \|T\|_{\mathcal{D}} \,. \end{aligned}$$

This means the norm of the extension $\tilde{T}$ is equal to the norm of the original operator T. This completes the proof. □

3.2 The spectrum of an operator

One of the fundamental characteristics an $N \times N$ matrix A possesses are its eigenvalues. These are obtained as the zeros of the characteristic polynomial

$$\chi(\lambda) := \det\big(\lambda - A\big) = \prod_{j=1}^{n} (\lambda - \mu_j)^{n_j} \,,$$

and their collection $\sigma(A) := \{\mu_1, \ldots, \mu_n\}$ is the *spectrum* of the matrix A. At these zeros, the eigenvalue equation

$$Av = \mu_j v$$

must have a non-trivial solution, and we may instead view $\sigma(A)$ as the set of eigenvalues of A. Note that due to the potential presence of Jordan blocks, the dimensionality of the eigenspace to each of the zeros μ_j of the characteristic polynomial $\chi(\lambda)$ may be smaller than the algebraic multiplicity n_j.

The generalization to infinite dimensions is more delicate, and here not all elements of the spectrum need to be due to eigenvalues.

Definition 3.2.1 (Spectrum) *Let $T \in \mathcal{B}(\mathcal{X})$ be a bounded linear operator on a Banach space $\mathcal{X}$. We define:*

(1) The resolvent set *of T is the set $\rho(T) := \big\{ z \in \mathbb{C} \;\big|\; T - z \text{ is bijective} \big\}$.*

(2) The spectrum *$\sigma(T) := \mathbb{C} \setminus \rho(T)$ is the complement of $\rho(T)$ in $\mathbb{C}$.*

(3) The set of all eigenvalues is called point spectrum

$$\sigma_{\mathrm{p}}(T) := \big\{ z \in \mathbb{C} \;\big|\; T - z \text{ is not injective} \big\}.$$

(4) The continuous spectrum *is defined as*

$$\sigma_{\text{cont}}(T) := \big\{ z \in \mathbb{C} \ \big| \ T - z \textit{ is injective, } \operatorname{im}(T-z) \subseteq \mathcal{X} \textit{ dense} \big\}.$$

(5) The remainder of the spectrum is called residual spectrum*,*

$$\sigma_{\text{res}}(T) := \big\{ z \in \mathbb{C} \ \big| \ T - z \textit{ is injective, } \operatorname{im}(T-z) \subseteq \mathcal{X} \textit{ not dense} \big\}$$

The spectrum motivates the introduction of the *resolvent operator*

$$(T-z)^{-1} \in \mathcal{B}(\mathcal{X}), \qquad\qquad z \in \rho(T),$$

that plays a central role in many results of functional analysis.

As the inverse of a continuous bijection, the operator $(T-z)^{-1}$ is necessarily continuous, thus, bounded by the *Inverse Mapping Theorem* [RS72, Theorem III.10]; the latter is a consequence of the *Open Mapping Theorem* [RS72, Theorem III.11].

The resolvent $z \mapsto (T-z)^{-1}$ is analytic on the resolvent set: for any $z_0 \in \rho(T)$ the power series

$$(T-z)^{-1} = (T-z_0)^{-1} \sum_{n=0}^{\infty} (z-z_0)^n \left((T-z_0)^{-1} \right)^n$$

of the resolvent at z converges on a small disc

$$\Big\{ z \in \mathbb{C} \ \Big| \ |z - z_0| < \big\| (T-z_0)^{-1} \big\| \Big\} \subseteq \rho(T),$$

centered around z_0. Note that as usual we have set $\left((T-z)^{-1} \right)^0 := \mathrm{id}$ in the above power series. That extends the invertibility of $T - z_0$ from the point z_0 to a small disc of positive radius; consequently, the resolvent set is open and its complement, the spectrum $\sigma(T) \subset \mathbb{C}$, is closed. In fact,

$$\sigma(T) \subseteq \big\{ z \in \mathbb{C} \ \big| \ |z| \leq \|T\| \big\},$$

is contained in a closed disc by [RS72, Theorem VI.6], and as a bounded, closed subset of $\mathbb{C}$ it is *compact*. 2020.11.11

3.3 The adjoint operator

For a normed space $\mathcal{X}$ we have previously introduced $\mathcal{X}'$, the space of bounded linear functionals on $\mathcal{X}$. Then an operator $T : \mathcal{X} \longrightarrow \mathcal{Y}$ between two normed spaces naturally defines an operator between their duals, the *adjoint operator* $T' : \mathcal{Y}' \longrightarrow \mathcal{X}'$, via the prescription

$$(T'L)(x) := L(Tx), \qquad \forall x \in \mathcal{X},\ L \in \mathcal{Y}'. \tag{3.3.1}$$

In case of Hilbert spaces, one can associate the *Hilbert space adjoint*. We will almost exclusively work with the latter and thus drop "Hilbert space" most of the time and just speak of the adjoint.

Definition 3.3.1 (The Hilbert space adjoint) *Let $\mathcal{H}$ be a Hilbert space and $A \in \mathcal{B}(\mathcal{H})$ be a bounded linear operator on the Hilbert space $\mathcal{H}$. Then for any $\varphi \in \mathcal{H}$, the equation*

$$\langle A\psi, \varphi \rangle = \langle \psi, \phi \rangle \qquad \forall \psi \in \mathcal{D}(A)$$

defines a vector ϕ. For each $\varphi \in \mathcal{H}$, we set $A^* \varphi := \phi$ and A^* is called the (Hilbert space) adjoint *of the operator A.*

The Banach space adjoint A' is related to the Hilbert space adjoint

$$A^* = C^{-1} A' C$$

via conjugation with the antilinear map

$$C\psi := \langle \psi, \cdot \rangle = L_\psi.$$

In the physics literature the adjoint of the operator A is usually denoted with $A^\dagger$ rather than A^*, because z^* is reserved for complex conjugation on $\mathbb{C}$. The map $C : |\psi\rangle \mapsto \langle\psi|$ implements the duality between kets and bras, which is how physicists refer to vectors of and functionals on a Hilbert space.

Example (The adjoint of a rank-1 operator) Suppose we pick two arbitrary vectors $\varphi_0, \psi_0 \in \mathcal{H}$ and define the rank-1 operator

$$T\psi := |\psi_0\rangle\langle\varphi_0|\,\psi := \langle\varphi_0, \psi\rangle\,\psi_0.$$

The Cauchy-Schwartz inequality ensures the boundedness of T,

$$\begin{aligned} \|T\| &= \sup_{\substack{\psi\in\mathcal{H} \\ \|\psi\|=1}} \big\|T\psi\big\| \\ &\leq \sup_{\substack{\psi\in\mathcal{H} \\ \|\psi\|=1}} \big(\|\varphi_0\|\,\|\psi_0\|\,\|\psi\|\big) = \|\varphi_0\|\,\|\psi_0\|. \end{aligned}$$

Rank-1 operators are often expressed in the eminently useful bra-ket notation, which makes computing the adjoint particularly simple. Indeed, we can verify

$$T^* = \big(|\psi_0\rangle\langle\varphi_0|\big)^* = |\varphi_0\rangle\langle\psi_0|$$

from a straightforward computation that starts from first principles,

$$\begin{aligned}\big\langle T\varphi, \psi \big\rangle &= \big\langle \langle \varphi_0, \varphi \rangle \, \psi_0, \psi \big\rangle = \langle \varphi, \varphi_0 \rangle \, \langle \psi_0, \psi \rangle \\ &= \big\langle \varphi, \langle \psi_0, \psi \rangle \varphi_0 \big\rangle = \big\langle \varphi, T^* \psi \big\rangle.\end{aligned}$$

Let us collect some facts about adjoints of bounded operators.

Proposition 3.3.2 *Let $A, B \in \mathcal{B}(\mathcal{H})$ be two bounded linear operators on a Hilbert space $\mathcal{H}$ and $\alpha \in \mathbb{C}$. Then the following holds true:*

(1) $(A+B)^* = A^* + B^*$

(2) $(\alpha A)^* = \overline{\alpha} \, A^*$

(3) $(AB)^* = B^* A^*$

(4) $\|A^*\| = \|A\|$

(5) $A^{**} = A$

(6) $\|A^* A\| = \|A \, A^*\| = \|A\|^2$

(7) $\ker A = (\mathrm{im}\, A^*)^\perp$, $\ker A^* = (\mathrm{im}\, A)^\perp$

Proof Properties (1)–(3) follow directly from the definition of the adjoint.

To show (4), we note that $\|A\| \leq \|A^*\|$ follows from

$$\begin{aligned}\|A\varphi\| = \Big| \Big\langle \tfrac{A\varphi}{\|A\varphi\|}, A\varphi \Big\rangle \Big| &\overset{\star}{=} \sup_{\|L\|_* = 1} |L(A\varphi)| \\ &= \sup_{\|\psi_L\| = 1} |\langle A^* \psi_L, \varphi \rangle| \leq \|A^*\| \, \|\varphi\|\end{aligned}$$

where in the step marked with $\star$, we have used that we can calculate the norm from picking the functional associated to $\frac{A\varphi}{\|A\varphi\|}$: for a functional with norm 1, $\|L\|_* = 1$, the norm of $L(A\varphi)$ cannot exceed that of $A\varphi$

$$\big|L(A\varphi)\big| = \big|\langle \psi_L, A\varphi \rangle\big| \leq \|\psi_L\| \|A\varphi\| = \|A\varphi\|.$$

Here, ψ_L is the vector such that $L = \langle \psi_L, \cdot \, \rangle$ which exists and is unique by virtue of Theorem **??**. This theorem also ensures $\|L\|_* = \|\psi_L\|$.

On the other hand, from

$$\begin{aligned}
\left\|A^* \psi_L\right\| = \left\|L_{A^* \psi_L}\right\|_* &= \sup_{\|\varphi\|=1} \left|\left\langle A^* \psi_L, \varphi \right\rangle\right| \\
&\leq \sup_{\|\varphi\|=1} \|\psi_L\| \, \|A\varphi\| = \|A\| \, \|L\|_* = \|A\| \, \|\psi_L\|
\end{aligned}$$

we conclude $\|A^*\| \leq \|A\|$. Hence, the norms coincide, $\|A^*\| = \|A\|$.

(5) is clear. For (6), we rewrite the square of the operator norm in terms of the scalar product and estimate from above,

$$\begin{aligned}
\|A\|^2 = \sup_{\|\varphi\|=1} \|A\varphi\|^2 &= \sup_{\|\varphi\|=1} \left\langle \varphi, A^* A \varphi \right\rangle \\
&\leq \sup_{\|\varphi\|=1} \|A^* A \varphi\| = \|A^* A\| \, .
\end{aligned}$$

But the right-hand side can be estimated from *above* by the lower bound $\|A\|^2$,

$$\|A\|^2 \leq \|A^* A\| \leq \|A^*\| \, \|A\| = \|A\|^2 \, ,$$

i. e. $\|A\|^2 = \|A^* A\|$ necessarily agree. Combined with (4), we obtain the same type of squeezing estimate

$$\|A\|^2 = \|A^*\|^2 \leq \|A \, A^*\| \leq \|A\| \, \|A^*\| = \|A\|^2 \, ,$$

that yields $\|A \, A^*\| = \|A\|^2 = \|A^* A\|$.

(7) is left as an exercise. □

We next give a simple zoology of bounded operators:

Definition 3.3.3 *Let $\mathcal{H}$ be a Hilbert space and $A \in \mathcal{B}(\mathcal{H})$. Then A is called*

(1) normal *if $A^* A = A \, A^*$,*

(2) selfadjoint *(or hermitian) if $A^* = A$,*

(3) unitary *if $A^* = A^{-1}$,*

(4) an orthogonal projection *if $A^2 = A$ and $A^* = A$, and*

(5) positive semidefinite *(or non-negative) if and only if $\langle \varphi, A\varphi \rangle \geq 0$ for all $\varphi \in \mathcal{H}$ and positive (definite) if the inequality is strict.*

The notions of selfadjoint and non-negative operators generalize to unbounded operators; we will detail some of the technical minutiæ later in this chapter.

3.4 Continuous operators between Fréchet spaces

A norm is one way to define a notion of distance between vectors; and this notion of distance $d(x_1, x_2) = \|x_1 - x_2\|$ defines a topology that is generated by open balls

$$B_r(x_0) := \{x \in \mathcal{X} \mid d(x, x_0) < r\}.$$

And once we have a topology in hand, we can define a notion of continuity (cf. Definition 2.2.10). But there are other ways to define a topology, and hence, continuity. A topological vector space is a vector space with a topology such that the vector space operations are compatible with the given topology.

Definition 3.4.1 (Topological vector space) *A topological vector space is a linear space $\mathcal{X}$ equipped with a topology $\mathcal{T}$ (cf. Definition 2.2.5) so that*

(a) vector addition $(x_1, x_2) \mapsto x_1 + x_2$ and

(b) scalar multiplication $(\alpha, x) \mapsto \alpha x$

are continuous (cf. Definition 2.2.10).

A common way to define a topology is with the help of a family of seminorms.

Definition 3.4.2 (Seminorm) *Let $\mathcal{X}$ be a linear space. A seminorm $p : \mathcal{X} \longrightarrow [0, \infty)$ is a function that satisfies*

(a) $p(\alpha x) = |\alpha| \, p(x)$ and

(b) $p(x_1 + x_2) \leq p(x_1) + p(x_2)$ (triangle inequality)

for all $\alpha \in \mathbb{C}$ and $x, x_1, x_2 \in \mathcal{X}$.

In case $p : \mathcal{X} \longrightarrow [0, \infty)$ is only a seminorm, but not a norm, then "Cauchy sequences" with respect to this seminorm need not converge. That is because $p(x) = 0$ does not imply $x = 0$.

A locally convex space is a topological vector space equipped with a *family* of seminorms, which in aggregate ensure convergence.

Definition 3.4.3 (Locally convex space) *Let $\mathcal{X}$ be a topological vector space whose topology is defined by a family of seminorms $p \in \mathcal{P}$. Then $\mathcal{X}$ is called a locally convex space if and only if the intersection*

$$\bigcap_{p \in \mathcal{P}} \{x \in \mathcal{X} \mid p(x) = 0\} = \{0\}$$

is trivial.

The condition says that while *individual* seminorms are not positive *definite* (in the sense of Definition 2.3.1 (a)), their *collection* is. This condition implies locally convex spaces are *Hausdorff*, i. e. we can separate points by neighborhoods: for any two distinct points $x_1, x_2 \in \mathcal{X}$, there are open neighborhoods $U_1, U_2 \in \mathcal{T}$ of these points with $U_1 \cap U_2 = \emptyset$. Equivalently, there exists a seminorm p with $p(x_1 - x_2) = \varepsilon > 0$, and we may pick

$$U_j := \bigl\{ x \in \mathcal{X} \;\; | \;\; p(x - x_j) < \varepsilon/2 \bigr\} \tag{3.4.1}$$

as our open neighborhoods. You may picture them as "pancakes" in the vector space that are semi-infinite in the indeterminate directions ("where $p(x) = 0$") and ε-thin in the other.

There are now three cases to consider:

(1) The collection of seminorms is *uncountable*.

(2) When the collection of seminorms is *finite*, the locally convex space can be made into a normed space by summing up all (finitely many) seminorms,

$$\lVert x \rVert := \sum_{p \in \mathcal{P}} p(x).$$

Consequently, we are dealing with just another normed vector space, which is covered by the previous sections in this chapter.

(3) The third case is when the family of seminorms is countably infinite. While there exists no norm, the space is still *metrizable:*

$$d(x_1, x_2) := \sum_{n=0}^{\infty} \frac{1}{2^n} \frac{p(x_1 - x_2)}{1 + p(x_1 - x_2)}. \tag{3.4.2}$$

Clearly, $d(x_1, x_2) = 0$ holds if and only if the two points $x_1 = x_2$ agree. Perhaps somewhat unusually, no two points are further than distance 2 apart, since $d(x_1, x_2) \leq \sum_{n=0}^{\infty} 2^{-n} = 2$.

Metrizable spaces have many very nice characteristics, which is why one defines the notion of

Definition 3.4.4 (Fréchet space) *A topological vector space $\mathcal{X}$ with a family of seminorms $\mathcal{P}$ is a Fréchet space if and only if the following three conditions are satisfied:*

(a) $\mathcal{X}$ is a locally convex space.

(b) Its topology can be induced by a countable *family of seminorms.*

(c) $\mathcal{X}$ is complete with respect to the family of seminorms.

Note that one can also *define* a countable family of seminorms once a basis for the topology is given.

A common example are the Schwartz functions and tempered distributions, which will be treated in Chapter **??**. Another standard example are the smooth functions.

Example ($\mathcal{C}^\infty_{\mathrm{b}}(\mathbb{R}^d, \mathbb{C})$) The space $\mathcal{C}^\infty(\mathbb{R}^d, \mathbb{C})$ of smooth, bounded, complex-valued functions with bounded derivatives to any order is a Fréchet space. The family of seminorms is just the sup norm,

$$\|f\|_a := \sup_{x \in \mathbb{R}^d} \bigl|\partial_x^a f(x)\bigr|$$

where $a = (a_1, \ldots, a_d) \in \mathbb{N}_0^d$ is a multiindex and $\partial_x^a := \partial_{x_1}^{a_1} \cdots \partial_{x_d}^{a_d}$ a partial derivative of order $|a| := a_1 + \ldots + a_d$.

Equivalently, we can define the family of seminorms via

$$\|f\|_n := \max_{|a|=n} \|f\|_a$$

or

$$p_n(f) := \max_{k=0,\ldots,n} \|f\|_k .$$

Convergence with respect to $\|f\|_n$ ensures the continuity and boundedness of all the nth order derivatives of the function f. Convergence in p_n controls continuity of the function and all of its derivatives up to nth order. 2020.11.18

A continuous operator between topological vector spaces is just a continuous function in the sense of Definition 2.2.10 that is also linear. Given that topologies are closed under *finite* intersections, continuity means we need to control finitely many seminorms at a time. In the context of $\mathcal{C}^\infty_{\mathrm{b}}(\mathbb{R}^d, \mathbb{C})$ this means that we do not need to have *uniform* control of all derivatives simultaneously, just finitely many at a time. Consequently, a continuous linear operator $T : \mathcal{X} \longrightarrow \mathcal{Y}$ between Fréchet spaces is a linear operator for which we can estimate the nth seminorm in $\mathcal{Y}$ via

$$\|Tx\|_{\mathcal{Y},n} \leq C_n \sum_{j \in \mathcal{I}} \|x\|_{\mathcal{X},j} \, ,$$

where $C_n > 0$ is a constant that only depends on n and $\mathcal{I}$ is a *finite* collection of indices. Of course, we have assumed without loss of generality that the index set is either $\mathbb{N}$ or a subset thereof.

When the target space is a normed space, then we only need to find *one* such estimate. A special case is when $\mathcal{Y} = \mathbb{C}$ is just the complex numbers and the only seminorm is the absolute value, i. e. the case of linear continuous functionals.

Chapter 4

4 Schwartz Functions & Tempered Distributions

Schwartz functions $\mathcal{S}(\mathbb{R}^d)$ are a space of test functions, i. e. a space of "very nicely behaved functions", on which many operations such as Fourier transform and differentiation are well-defined and very well-behaved. *Tempered distributions* $\mathcal{S}'(\mathbb{R}^d)$ make up the *dual* to this space of test functions (analogous to Definition 3.1.2). The concept of adjoint operator (cf. Chapter 3.3) then allows us to extend Fourier transform and derivatives to objects which may not even be functions.

Test functions and distributions are covered in every good text book on Fourier analysis, e. g. in [SS03, Chapter 5] or [Fol09, Chapter 9] to name just two.

4.1 Schwartz functions

Our class of test functions $\mathcal{S}(\mathbb{R}^d)$ has three defining properties:

(1) $\mathcal{S}(\mathbb{R}^d) \subseteq \mathcal{C}^\infty(\mathbb{R}^d)$ forms a vector space.

(2) *Stability under derivation*, $\partial_x^\alpha \mathcal{S}(\mathbb{R}^d) \subset \mathcal{S}(\mathbb{R}^d)$: for all multiindices $\alpha \in \mathbb{N}_0^d$ and $f \in \mathcal{S}(\mathbb{R}^d)$, we have $\partial_x^\alpha f \in \mathcal{S}(\mathbb{R}^d)$.

(3) *Stability under Fourier transform*, $\mathcal{F}\mathcal{S}(\mathbb{R}^d) \subseteq \mathcal{S}(\mathbb{R}^d)$: for all $f \in \mathcal{S}(\mathbb{R}^d)$, the Fourier transform

$$\mathcal{F}^{\pm 1} f : \xi \mapsto \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}x \, \mathrm{e}^{\mp \mathrm{i} x \cdot \xi} \, f(x) \in \mathcal{S}(\mathbb{R}^d) \tag{4.1.1}$$

is also a test function.

These relatively simple requirements have surprisingly rich implications:

(1) $\mathcal{S}(\mathbb{R}^d) \subseteq L^1(\mathbb{R}^d)$ forms a subset of $L^1(\mathbb{R}^d)$ since the Fourier transform of Schwartz functions is assumed to exist.

(2) Stability under derivation, $\partial_x^\alpha \mathcal{S}(\mathbb{R}^d) \subseteq \mathcal{S}(\mathbb{R}^d)$ implies not just $f \in \mathcal{S}(\mathbb{R}^d)$, but that all of its derivatives are integrable functions as well.

(3) $\mathcal{F} : \mathcal{S}(\mathbb{R}^d) \longrightarrow \mathcal{S}(\mathbb{R}^d)$ acts bijectively.

(4) For all $\alpha \in \mathbb{N}_0^d$, we have $\mathcal{F}\big((\mathrm{i}\partial_x)^\alpha f\big) = x^\alpha \, \mathcal{F}f \in \mathcal{S}(\mathbb{R}^d)$.

(5) Hence, for all $a, \alpha \in \mathbb{N}_0^d$, we have $x^a \partial_x^\alpha f \in \mathcal{S}(\mathbb{R}^d)$.

(6) Translations of Schwartz functions are again Schwartz functions; this follows from $\mathcal{F}f(\,\cdot\, - x_0) = \mathrm{e}^{-\mathrm{i}\xi \cdot x_0} \, \mathcal{F}f \in \mathcal{S}(\mathbb{R}^d)$ for all $x_0 \in \mathbb{R}^d$.

The only thing we are missing to a proper mathematical definition is to endow $\mathcal{S}(\mathbb{R}^d)$ with a suitable topology:

Definition 4.1.1 (Schwartz functions) *The space of Schwartz functions*

$$\mathcal{S}(\mathbb{R}^d) := \Big\{ f \in \mathcal{C}^\infty(\mathbb{R}^d) \;\Big|\; \forall a, \alpha \in \mathbb{N}_0^d : \lVert f \rVert_{a\alpha} < \infty \Big\}$$

is defined in terms of the family of seminorms[1] $\big\{ \lVert \cdot \rVert_{a\alpha} \big\}_{a,\alpha \in \mathbb{N}_0^d}$ *indexed by* $a, \alpha \in \mathbb{N}_0^d$*,*

$$\lVert f \rVert_{a\alpha} := \sup_{x \in \mathbb{R}^d} \big| x^a \partial_x^\alpha f(x) \big|. \tag{4.1.2}$$

The family of seminorms defines a so-called *Fréchet topology:* put in simple terms, to make sure that sequences in $\mathcal{S}(\mathbb{R}^d)$ converge to smooth functions that decay faster than polynomially (often referred to as *rapid* decay), we need to control infinitely many derivatives as well as their decay. This is also the reason why there is *no norm on* $\mathcal{S}(\mathbb{R}^d)$ which generates the same topology as the family of seminorms. However, $\lVert f \rVert_{a\alpha} = 0$ for all $a, \alpha \in \mathbb{N}_0^d$ ensures $f = 0$, all seminorms put together can distinguish functions in $\mathcal{S}(\mathbb{R}^d)$. Crucially, the notion of topology gives rise to a notion of continuity: to ensure continuity of linear functionals on $\mathcal{S}(\mathbb{R}^d)$ we need to control finitely many seminorms (cf. Proposition 4.2.2) rather than just one for normed spaces (cf. Chapter **??**).

Example Two simple examples of Schwartz functions in $d = 1$ are

$$f(x) = \mathrm{e}^{-a x^2}, \qquad a > 0,$$

[1] A seminorm has all properties of a norm except that $\lVert f \rVert = 0$ does not necessarily imply $f = 0$.

and

$$g(x) = \begin{cases} \mathrm{e}^{-\frac{1}{1-x^2}+1} & |x| < 1 \\ 0 & |x| \geq 1 \end{cases}.$$

The second one even has compact support.

The first major fact we will establish is completeness.

Theorem 4.1.2 *The space of Schwartz functions $\mathcal{S}(\mathbb{R}^d)$ endowed with the seminorms* (4.1.2) *and the metric*

$$\mathrm{d}(f,g) := \sum_{n=0}^{\infty} 2^{-n} \max_{|a|+|\alpha|=n} \frac{\|f-g\|_{a\alpha}}{1+\|f-g\|_{a\alpha}}$$

is complete, i. e. it is a Fréchet space.

Proof Evidently, d is positive and symmetric. It also satisfies the triangle inequality as $x \mapsto \frac{x}{1+x}$ is concave on $[0,\infty)$ and all of the seminorms satisfy the triangle inequality. Hence, $\big(\mathcal{S}(\mathbb{R}^d), \mathrm{d}\big)$ is a metric space.

To show completeness, take a Cauchy sequence $(f_n)_{n\in\mathbb{N}}$ with respect to d. By definition and positivity, this means $(f_n)_{n\in\mathbb{N}}$ is also a Cauchy sequence with respect to all of the seminorms $\|\cdot\|_{a\alpha}$. Each of the $\big(x^a \partial_x^\alpha f_n\big)_{n\in\mathbb{N}}$ converge to some $g_{a\alpha}$ as the space of bounded continuous functions $\mathcal{C}_{\mathrm{b}}(\mathbb{R}^d)$ endowed with the sup norm is complete.

It remains to show that the limits of the derivatives are the derivatives of the limits, i. e. $g_{a\alpha} = x^a \partial_x^\alpha g_{00}$. Clearly, only taking derivatives is problematic. We will initially prove this for $|\alpha| = 1$, i. e. the case of a single derivative and proceed by induction.

Assume we are interested in the sequence $\big(\partial_{x_k} f_n\big)_{n\in\mathbb{N}}$ of derivatives where $k \in \{1,\ldots,d\}$. We denote the associated multi index $\alpha_k := (0,\ldots,0,1,0,\ldots) \in \mathbb{N}_0^d$ and introduce the kth canonical basis vector $e_k := (0,\ldots,0,1,0,\ldots) \in \mathbb{R}^d$. The Fundamental Theorem of Calculus connects f_n to its derivative,

$$f_n(x) = f_n(x - x_k\, e_k) + \int_0^{x_k} \mathrm{d}s\, \partial_{x_k} f_n\big(x + (s - x_k)\, e_k\big).$$

Taking the limit $n \to \infty$ on both sides and exchanging limit and integral via a Dominated Convergence argument yields

$$g_{00}(x) = g_{00}(x - x_k\, e_k) + \int_0^{x_k} \mathrm{d}s\, g_{0\alpha_k}\big(x + (s - x_k)\, e_k\big),$$

since $f_n \to g_{00}$ and $\partial_{x_k} f_n \to g_{0\alpha_k}$ converge uniformly in x. Comparing these two and reading the Fundamental Theorem of Calculus in reverse allows us to conclude

$g_{00} \in \mathcal{C}^1(\mathbb{R}^d)$, and that the derivative $\partial_{x_k} g_{00} = g_{0\alpha_k} \in L^\infty(\mathbb{R}^d)$ indeed coincides with $g_{0\alpha_k}$.

We then proceed by induction to show $g_{00} \in \mathcal{C}^\infty_{\mathrm{b}}(\mathbb{R}^d)$. This means $\mathrm{d}(f_n, g_{00}) \to 0$ as $n \to \infty$ and $\mathcal{S}(\mathbb{R}^d)$ is complete. □

What makes Schwartz functions so versatile is that they automatically lie in a whole host of other spaces, most notably all the $L^p(\mathbb{R}^d)$ spaces. And importantly, the embeddings are continuous since the $L^p(\mathbb{R}^d)$ norm of each element in $\mathcal{S}(\mathbb{R}^d)$ can be dominated by a finite number of Fréchet seminorms.

Lemma 4.1.3 *Let $f \in \mathcal{S}(\mathbb{R}^d)$. Then for each $1 \leq p < \infty$, the L^p norm of f can be estimated by a finite number of seminorms,*

$$\|f\|_{L^p(\mathbb{R}^d)} \leq C_1(d) \, \|f\|_{00} + C_2(d) \max_{|a|=2n(d)} \|f\|_{a0} \, , \tag{4.1.3}$$

where the constants $C_1(d), C_2(d) \in \mathbb{R}^+$ and $n(d) \in \mathbb{N}_0$ only depend on the dimension of $\mathbb{R}^d$. Consequently, the Schwartz functions $\mathcal{S}(\mathbb{R}^d) \hookrightarrow L^p(\mathbb{R}^d)$ continuously embed into all the $L^p(\mathbb{R}^d)$ spaces.

Proof We split the integral on $\mathbb{R}^d$ into an integral over the unit ball centered at the origin and its complement, and then estimate both separately:

$$\begin{aligned} \|f\|_{L^p(\mathbb{R}^d)} &\leq \left(\int_{|x| \leq 1} \mathrm{d}x \, |f(x)|^p \right)^{1/p} + \left(\int_{|x|>1} \mathrm{d}x \, |f(x)|^p \right)^{1/p} \\ &\leq \|f\|_{00} \left(\int_{|x| \leq 1} \mathrm{d}x \, 1 \right)^{1/p} + \left(\int_{|x|>1} \mathrm{d}x \, \left| |x|^{2n} \, f(x) \right|^p \frac{1}{|x|^{2np}} \right)^{1/p} \\ &\leq \mathrm{Vol}\big(B_1(0)\big)^{1/p} \, \|f\|_{00} + \max_{|a|=2n} \|f\|_{a0} \left(\int_{|x|>1} \mathrm{d}x \, \frac{1}{|x|^{2np}} \right)^{1/p} \end{aligned}$$

□

If we choose $n \equiv n(d)$ large enough, $|x|^{-2np}$ is integrable away from the singularity at $x = 0$ and its integral can be computed explicitly. Therefore, we get the desired estimate (4.1.3). As indicated the minimal power $n(d)$ and the constants $C_1(d)$ and
2020.11.25 $C_2(d)$ only depend on the dimension d.

Lemma 4.1.4 *The smooth functions with compact support $\mathcal{C}^\infty_{\mathrm{c}}(\mathbb{R}^d)$ are dense in the space of Schwartz functions $\mathcal{S}(\mathbb{R}^d)$.*

Proof Take any $f \in \mathcal{S}(\mathbb{R}^d)$ and choose

$$g(x) = \begin{cases} \mathrm{e}^{-\frac{1}{1-x^2}+1} & |x| \leq 1 \\ 0 & |x| > 1 \end{cases} .$$

Then $f_n := g(\cdot/n)\, f$ converges to f in $\mathcal{S}(\mathbb{R}^d)$, i. e.

$$\lim_{n\to\infty} \big\lVert f_n - f \big\rVert_{a\alpha} = 0$$

holds for all multi indices $a, \alpha \in \mathbb{N}_0^d$. $\square$

The L^p-norm estimate from Lemma 4.1.3 and the density of $\mathcal{C}_{\mathrm{c}}^\infty(\mathbb{R}^d)$ in $L^p(\mathbb{R}^d)$, Lemma **??**, combine with the previous lemma to give us

Corollary 4.1.5 *$\mathcal{S}(\mathbb{R}^d)$ continuously embeds into $L^p(\mathbb{R}^d)$.*

Next, we will show that $\mathcal{F} : \mathcal{S}(\mathbb{R}^d) \longrightarrow \mathcal{S}(\mathbb{R}^d)$ is a continuous and bijective map from $\mathcal{S}(\mathbb{R}^d)$ onto itself.

Theorem 4.1.6 *The Fourier transform $\mathcal{F}$ as defined by equation* (4.1.1) *maps $\mathcal{S}(\mathbb{R}^d)$ continuously and bijectively onto itself. The inverse $\mathcal{F}^{-1}$ is continuous as well. Furthermore, for all $f \in \mathcal{S}(\mathbb{R}^d)$ and $a, \alpha \in \mathbb{N}_0^d$, we have*

$$\mathcal{F}\big(x^a (-\mathrm{i}\partial_x)^\alpha f\big) = (+\mathrm{i}\partial_\xi)^a \xi^\alpha \mathcal{F} f. \tag{4.1.4}$$

Proof We need to prove $\mathcal{F}\big(x^a (-\mathrm{i}\partial_x)^\alpha f\big) = (+\mathrm{i}\partial_\xi)^a \xi^\alpha \mathcal{F} f$ first: since $x^\alpha \partial_x^a f$ is integrable, its Fourier transform exists and is continuous by Dominated Convergence. For any $a, \alpha \in \mathbb{N}_0^d$, we have

$$\begin{aligned}
\Big(\mathcal{F}\big(x^a (-\mathrm{i}\partial_x)^\alpha f\big)\Big)(\xi) &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}x \, \mathrm{e}^{-\mathrm{i} x \cdot \xi} \, x^a \, (-\mathrm{i}\partial_x)^\alpha f(x) \\
&= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}x \, (+\mathrm{i}\partial_\xi)^a \mathrm{e}^{-\mathrm{i} x \cdot \xi} \, (-\mathrm{i}\partial_x)^\alpha f(x) \\
&\overset{\star}{=} \frac{1}{(2\pi)^{d/2}} (+\mathrm{i}\partial_\xi)^a \int_{\mathbb{R}^d} \mathrm{d}x \, \mathrm{e}^{-\mathrm{i} x \cdot \xi} \, (-\mathrm{i}\partial_x)^\alpha f(x).
\end{aligned}$$

In the step marked with $\star$, we have used Dominated Convergence to interchange integration and differentiation. Now we integrate partially $\lvert\alpha\rvert$ times and use that the boundary terms vanish,

$$\begin{aligned}
\Big(\mathcal{F}\big(x^a (-\mathrm{i}\partial_x)^\alpha f\big)\Big)(\xi) &= \frac{1}{(2\pi)^{d/2}} (+\mathrm{i}\partial_\xi)^a \int_{\mathbb{R}^d} \mathrm{d}x \, (+\mathrm{i}\partial_x)^\alpha \mathrm{e}^{-\mathrm{i} x \cdot \xi} \, f(x) \\
&= \frac{1}{(2\pi)^{d/2}} (+\mathrm{i}\partial_\xi)^a \int_{\mathbb{R}^d} \mathrm{d}x \, \xi^\alpha \mathrm{e}^{-\mathrm{i} x \cdot \xi} \, f(x) \\
&= \big((+\mathrm{i}\partial_\xi)^a \xi^\alpha \mathcal{F} f\big)(\xi).
\end{aligned}$$

To show $\mathcal{F}$ is continuous, we need to estimate the seminorms of $\mathcal{F}f$ by those of f: for any $a, \alpha \in \mathbb{N}_0^d$, it holds

$$\begin{aligned}\big\lVert \mathcal{F}f \big\rVert_{a\alpha} &= \sup_{\xi \in \mathbb{R}^d} \big\lvert \big(\xi^a \partial_\xi^\alpha \mathcal{F}f\big)(\xi) \big\rvert = \sup_{\xi \in \mathbb{R}^d} \Big\lvert \Big(\mathcal{F}\big(\partial_x^a x^\alpha f\big)\Big)(x) \Big\rvert \\ &\leq \frac{1}{(2\pi)^{d/2}} \big\lVert \partial_x^a x^\alpha f \big\rVert_{L^1(\mathbb{R}^d)}.\end{aligned}$$

In particular, this implies $\mathcal{F}f \in \mathcal{S}(\mathbb{R}^d)$. Since $\partial_x^a x^\alpha f \in \mathcal{S}(\mathbb{R}^d)$, we can apply Lemma 4.1.3 and estimate the right-hand side by a finite number of seminorms of f. Hence, $\mathcal{F}$ is continuous: if f_n is a Cauchy sequence in $\mathcal{S}(\mathbb{R}^d)$ that converges to f, then $\mathcal{F}f_n$ has to converge to $\mathcal{F}f \in \mathcal{S}(\mathbb{R}^d)$.

To show that $\mathcal{F}$ is a bijection with continuous inverse, we note that it suffices to prove $\mathcal{F}^{-1}\mathcal{F}f = f$ for functions f in a dense subset, for example $\mathcal{C}_\mathrm{c}^\infty(\mathbb{R}^d)$: pick f so that the support of is contained in a cube $[-n, +n]^d$ with sides of length $2n$. We can write f on $[-n, +n]^d$ as a uniformly convergent Fourier series,

$$f_n(x) = \sum_{\xi \in (\frac{\pi}{n}\mathbb{Z})^d} \hat{f}_n(\xi)\, \mathrm{e}^{\mathrm{i}x\cdot\xi},$$

which is computed from the Fourier components

$$\begin{aligned}\hat{f}_n(\xi) &= \frac{1}{\mathrm{Vol}\big([-n,+n]^d\big)} \int_{[-n,+n]^d} \mathrm{d}x\, \mathrm{e}^{-\mathrm{i}x\cdot\xi}\, f(x) \\ &= \frac{(2\pi)^{d/2}}{(2n)^d} \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}x\, \mathrm{e}^{-\mathrm{i}x\cdot\xi}\, f(x).\end{aligned}$$

The second equality holds if n is large enough so that $\mathrm{supp}\, f$ fits into the cube $[-n, +n]^d$. Hence, f_n can be expressed as

$$f_n(x) = \sum_{\xi \in (\frac{\pi}{n}\mathbb{Z})^d} \frac{1}{(2\pi)^{d/2}} \frac{\pi^d}{n^d}\, (\mathcal{F}f)(\xi)\, \mathrm{e}^{\mathrm{i}x\cdot\xi},$$

which is a Riemann sum that converges to

$$f(x) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}x\, \mathrm{e}^{\mathrm{i}x\cdot\xi}\, (\mathcal{F}f)(\xi) = \big(\mathcal{F}^{-1}\mathcal{F}f\big)(x)$$

as $\mathcal{F}f \in \mathcal{S}$. This concludes the proof. □

The stability of $\mathcal{S}(\mathbb{R}^d)$ under the Fourier transform was one of the defining properties that we posited Schwartz functions to have in the beginning of this chapter. The other defining property is item (1) of the next

Proposition 4.1.7 *The Schwartz functions have the following properties:*

(1) For all a, α, the map $f \mapsto x^a \partial_\xi^\alpha f$ is continuous on $\mathcal{S}(\mathbb{R}^d)$.

(2) With pointwise multiplication of functions $\cdot : \mathcal{S}(\mathbb{R}^d) \times \mathcal{S}(\mathbb{R}^d) \longrightarrow \mathcal{S}(\mathbb{R}^d)$ and complex conjugation, $\mathcal{S}(\mathbb{R}^d)$ forms a Fréchet $*$-algebra, i. e. multiplication and complex conjugation are continuous.

(3) For any $x_0 \in \mathbb{R}^d$, the map $\tau_{x_0} : f \mapsto f(\cdot - x_0)$ continuous on $\mathcal{S}(\mathbb{R}^d)$.

(4) For any $f \in \mathcal{S}(\mathbb{R}^d)$, $\frac{1}{s}\big(\tau_{se_k} f - f\big)$ converges to $\partial_{x_k} f$ as $s \to 0$ where e_k is the kth canonical base vector of $\mathbb{R}^d$.

The proofs are left as an exercise to the readers.

Theorem 4.1.8 *$\mathcal{S}(\mathbb{R}^d)$ is dense in $L^p(\mathbb{R}^d)$, $1 \leq p < \infty$.*

The fact that we can approximate any $L^p(\mathbb{R}^d)$ function arbitrarily well by test functions comes in handy when dealing with operators. For example, some of them have a “nice” definition on a dense subset such as $\mathcal{C}^\infty_{\mathrm{c}}(\mathbb{R}^d)$ or $\mathcal{S}(\mathbb{R}^d)$ and then we can extend their definition by density to larger sets of functions. One such example was the Fourier transform on $L^2(\mathbb{R}^d)$: while we could settle this question using a combination of Monotone and Dominated Convergence (cf. [LL01, Chapter 5]), the path was quite thorny and involved subtle arguments involving Dominated and Monotone Convergence. Schwartz functions allow for a much simpler and non-technical approach.

Theorem 4.1.9 *For all $f, g \in \mathcal{S}(\mathbb{R}^d)$, we have*

$$\int_{\mathbb{R}^d} \mathrm{d}x \, (\mathcal{F}f)(x) \, g(x) = \int_{\mathbb{R}^d} \mathrm{d}x \, f(x) \, (\mathcal{F}g)(x).$$

This implies $\langle \mathcal{F}f, g \rangle = \langle f, \mathcal{F}^{-1} g \rangle$ and $\langle \mathcal{F}f, \mathcal{F}g \rangle = \langle f, g \rangle$ where $\langle \cdot \, , \cdot \rangle$ is the usual scalar product on $L^2(\mathbb{R}^d)$.

Proof We start by writing out the Fourier transform on the left-hand side. Thanks to Fubini's Theorem we are free to reverse the order of integration,

$$\begin{aligned}
\int_{\mathbb{R}^d} \mathrm{d}x \, (\mathcal{F}f)(x) \, g(x) &= \int_{\mathbb{R}^d} \mathrm{d}x \, \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}\xi \, \mathrm{e}^{-\mathrm{i} x \cdot \xi} \, f(\xi) \, g(x) \\
&= \int_{\mathbb{R}^d} \mathrm{d}\xi \, f(\xi) \, \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}x \, \mathrm{e}^{-\mathrm{i} x \cdot \xi} g(x) \\
&= \int_{\mathbb{R}^d} \mathrm{d}\xi \, f(\xi) \, (\mathcal{F}g)(\xi).
\end{aligned}$$

To prove the second part, we remark that compared to the scalar product on $L^2(\mathbb{R}^d)$, we are missing a complex conjugation of the first function. Furthermore, $\overline{\mathcal{F}f} = \mathcal{F}^{-1}\bar{f}$ holds. From this, it follows that $\langle \mathcal{F}f, g \rangle = \langle f, \mathcal{F}^{-1}g \rangle$, and upon replacing g with $\mathcal{F}g$, that $\langle \mathcal{F}f, \mathcal{F}g \rangle = \langle f, \mathcal{F}^{-1}\mathcal{F}g \rangle = \langle f, g \rangle$. □

The previous statement gives another avenue to define the Fourier transform on $L^2(\mathbb{R}^d)$. Given that Schwartz functions are dense in $L^2(\mathbb{R}^d)$ (Theorem 4.1.8), we immediately deduce

Corollary 4.1.10 *The Fourier transform on $L^2(\mathbb{R}^d)$ is the unique continuous extension of $\mathcal{F} : \mathcal{S}(\mathbb{R}^d) \longrightarrow \mathcal{S}(\mathbb{R}^d)$.*

Since we know that the Fourier transform intertwines the convolution and the pointwise product of functions, Theorem 4.1.9 also implies that the convolution defines a commutative multiplication on the Schwartz functions:

Corollary 4.1.11 *The convolution $* : \mathcal{S}(\mathbb{R}^d) \times \mathcal{S}(\mathbb{R}^d) \longrightarrow \mathcal{S}(\mathbb{R}^d)$ defines a commutative product that is continuous in both arguments.*

Proof Let $f, g \in \mathcal{S}(\mathbb{R}^d)$. Because Schwartz functions are also integrable, $f * g$ exists in $L^1(\mathbb{R}^d)$ and satisfies $\mathcal{F}(f * g) = (2\pi)^{d/2}\mathcal{F}f\,\mathcal{F}g$. This means we can rewrite

$$f * g = (2\pi)^{d/2}\mathcal{F}^{-1}\big(\mathcal{F}f\,\mathcal{F}g\big) \tag{4.1.5}$$

as the Fourier transform of a product of Schwartz functions, and thus, $f * g \in \mathcal{S}(\mathbb{R}^d)$. Equation (4.1.5) also makes the commutativity $f * g = g * f$ evident.

The continuity follows from the continuity of the Fourier transform (Theorem 4.1.7) and the continuity of the pointwise product of Schwartz functions in the Fréchet topology (Proposition 4.1.7 (2)). □

4.2 Tempered distributions

After a thorough discussion of our space of test functions, we can now proceed to distributions, namely linear continuous functionals on $\mathcal{S}(\mathbb{R}^d)$.

Definition 4.2.1 (Tempered distributions) *The space of tempered distributions $\mathcal{S}'(\mathbb{R}^d)$ is comprised of continuous linear functionals on $\mathcal{S}(\mathbb{R}^d)$. When working with tempered distrubtions, we will often use* duality bracket

$$(L, f) := L(f) \qquad\qquad \forall L \in \mathcal{S}'(\mathbb{R}^d),\ f \in \mathcal{S}(\mathbb{R}^d).$$

Example (1) The δ distribution defined via

$$\delta(f) := f(0)$$

is a linear continuous functional on $\mathcal{S}(\mathbb{R}^d)$.

(2) Let $g \in L^p(\mathbb{R}^d)$, $1 \leq p < \infty$, then for $f \in \mathcal{S}(\mathbb{R}^d)$, we define

$$L_g(f) = \int_{\mathbb{R}^d} \mathrm{d}x \, g(x) \, f(x) =: (g, f). \tag{4.2.1}$$

As $f \in \mathcal{S}(\mathbb{R}^d) \subset L^q(\mathbb{R}^d)$ and $1/p + 1/q = 1$, by Hölder's inequality, we have

$$|(g, f)| \leq \|g\|_{L^p(\mathbb{R}^d)} \, \|f\|_{L^q(\mathbb{R}^d)} \, .$$

Since $\|f\|_q$ can be bounded by a finite linear combination of Fréchet seminorms of f, the linear map L_g is continuous, and the inclusion map $\imath : L^p(\mathbb{R}^d) \hookrightarrow \mathcal{S}'(\mathbb{R}^d)$ is continuous.

(3) Equation (4.2.1) is the *canonical way to interpret many "less nice" functions as distributions*: we identify a suitable function $g : \mathbb{R}^d \longrightarrow \mathbb{C}$ with the distribution L_g. For instance, polynomially bounded smooth functions (think of $g(x) = x^2$) define continuous linear functionals in this manner since for any $g \in \mathcal{C}^\infty_{\mathrm{pol}}(\mathbb{R}^d)$, there exists $n \in \mathbb{N}_0$ such that $\sqrt{1+x^2}^{\,-n} g(x)$ is bounded. Hence, for any $f \in \mathcal{S}(\mathbb{R}^d)$, Hölder's inequality yields

$$\begin{aligned}
|(g, f)| &= \left| \int_{\mathbb{R}^d} \mathrm{d}x \, g(x) \, f(x) \right| \\
&= \left| \int_{\mathbb{R}^d} \mathrm{d}x \left(\sqrt{1+x^2}^{\,-n} g(x) \right) \left(\sqrt{1+x^2}^{\,n} f(x) \right) \right| \\
&\leq \left\| \sqrt{1+x^2}^{\,-n} g(x) \right\|_{L^\infty(\mathbb{R}^d)} \left\| \sqrt{1+x^2}^{\,n} f(x) \right\|_{L^1(\mathbb{R}^d)} .
\end{aligned}$$

Later on, we will see that this point of view, interpreting "not so nice" functions as distributions, helps us extend operations from test functions to much broader classes of functions. 2020.12.02

Similar to the case of normed spaces, we see that continuity in Fréchet spaces implies boundedness in the following precise sense:

Proposition 4.2.2 *A linear functional $L : \mathcal{S}(\mathbb{R}^d) \longrightarrow \mathbb{C}$ is a tempered distribution (i. e. continuous) if and only if $\left|L(f)\right|$ can be estimated by finitely many seminorms. More precisely, there exist constants $C > 0$ and $N \in \mathbb{N}_0$ such that*

$$\left|L(f)\right| \leq C \sum_{|a|,|\alpha| \leq N} \|f\|_{a\alpha}$$

holds true for all $f \in \mathcal{S}(\mathbb{R}^d)$.

Even though we will not reproduce a proof here, let us at least sketch its idea: because one has no control over the growth or decay of the seminorms $\|f\|_{a\alpha}$, maxima or sums of seminorms are finite if and only if only finitely many of them enter the bound.

As mentioned before, we can interpret suitable functions g as tempered distributions. In particular, every Schwartz function $g \in \mathcal{S}(\mathbb{R}^d)$ defines a tempered distribution so that

$$\begin{aligned}\big(\partial_{x_k} g \, , \, f\big) &= \int_{\mathbb{R}^d} \mathrm{d}x \, \partial_{x_k} g(x) \, f(x) = - \int_{\mathbb{R}^d} \mathrm{d}x \, g(x) \, \partial_{x_k} f(x) \\ &= \big(g \, , \, -\partial_{x_k} f\big)\end{aligned}$$

holds for any $f \in \mathcal{S}(\mathbb{R}^d)$. In fact, the right-hand side extends the definition of the derivative to tempered distributions.

Definition 4.2.3 (Weak derivative) *For $\alpha \in \mathbb{N}_0^d$ and $L \in \mathcal{S}'(\mathbb{R}^d)$, we define the* weak *or* distributional derivative *of L as*

$$\big(\partial_x^\alpha L \, , \, f\big) := \big(L \, , \, (-1)^{|\alpha|} \partial_x^\alpha f\big) \qquad \qquad \forall f \in \mathcal{S}(\mathbb{R}^d).$$

Example (1) The weak derivative of δ is

$$\big(\partial_{x_k} \delta \, , \, f\big) = \big(\delta \, , \, -\partial_{x_k} f\big) = -\partial_{x_k} f(0).$$

(2) Let $g \in \mathcal{C}^\infty_{\mathrm{pol}}(\mathbb{R}^d)$. Then the weak derivative coincides with the usual derivative: by partial integration, we can push ∂_{x_k} to the right,

$$\begin{aligned}\big(\partial_{x_k} g \, , \, f\big) &= -\big(g \, , \, \partial_{x_k} f\big) = - \int_{\mathbb{R}^d} \mathrm{d}x \, g(x) \, \partial_{x_k} f(x) \\ &= + \int_{\mathbb{R}^d} \mathrm{d}x \, \partial_{x_k} g(x) \, f(x).\end{aligned}$$

Similarly, the Fourier transform can be extended to a bijection $\mathcal{S}'(\mathbb{R}^d) \longrightarrow \mathcal{S}'(\mathbb{R}^d)$. Theorem 4.1.9 tells us that for $g, f \in \mathcal{S}(\mathbb{R}^d)$ the Fourier transform in the duality bracket can be pushed to the other side,

$$\big(\mathcal{F} g \, , \, f\big) = \big(g \, , \, \mathcal{F} f\big).$$

If we replace g with an arbitrary tempered distribution, the right-hand side again serves as *definition* of the left-hand side:

Definition 4.2.4 (Fourier transform on $\mathcal{S}'(\mathbb{R}^d)$) *For any tempered distribution $L \in \mathcal{S}'(\mathbb{R}^d)$, we define its Fourier transform to be*

$$\big(\mathcal{F} L \, , \, f\big) := \big(L \, , \, \mathcal{F} f\big) \qquad \qquad \forall f \in \mathcal{S}(\mathbb{R}^d).$$

Example (1) The Fourier transform of the Dirac distribution δ is the constant function $\mathcal{F}\delta = (2\pi)^{-d/2}$,

$$\begin{aligned}\big(\mathcal{F}\delta \, , \, f\big) &= \big(\delta \, , \, \mathcal{F}f\big) = \mathcal{F}f(0) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}x \, f(x) \\ &= \big((2\pi)^{-d/2} \, , \, f\big).\end{aligned}$$

(2) The Fourier transform of x^2 makes sense as a tempered distribution on $\mathbb{R}$: x^2 is a polynomially bounded function and therefore defines a tempered distribution via equation (4.2.1):

$$\begin{aligned}\big(\mathcal{F}x^2, f\big) &= \big(x^2, \mathcal{F}f\big) = \int_{\mathbb{R}} \mathrm{d}x \, x^2 \, \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{d}\xi \, \mathrm{e}^{-\mathrm{i}x\cdot\xi} \, f(\xi) \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{d}\xi \int_{\mathbb{R}} \mathrm{d}x \, (+\mathrm{i})^2 \partial_\xi^2 \mathrm{e}^{-\mathrm{i}x\cdot\xi} \, f(\xi) \\ &= (-1)^2 \cdot (-1) \int_{\mathbb{R}} \mathrm{d}\xi \left(\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{d}x \, \mathrm{e}^{-\mathrm{i}x\cdot\xi} \right) \partial_\xi^2 f(\xi) \\ &= - \int_{\mathbb{R}} \mathrm{d}\xi \, \sqrt{2\pi} \, \delta(\xi) \, \partial_\xi^2 f(\xi) \\ &= -\sqrt{2\pi} \, \partial_\xi^2 f(0) = \big(\sqrt{2\pi}\delta \, , \, -\partial_\xi^2 f\big) = \big(-\sqrt{2\pi}\delta'' \, , \, f\big)\end{aligned}$$

This is consistent with what we have shown earlier in Theorem 4.1.6,

$$\mathcal{F}\big(x^2 f\big) = (+\mathrm{i}\partial_\xi)^2 \mathcal{F}f = -\partial_\xi^2 \mathcal{F}f.$$

We have just computed Fourier transforms of functions that do not have Fourier transforms in the usual sense and to tempered distributions that are not functions at all. Extending other operations to tempered distributions works in the same way: we first start with a continuous linear operator on $\mathcal{S}(\mathbb{R}^d)$ and then use the adjoint to extend it to tempered distributions. Before we can make that precise, though, we need to introduce the appropriate notion of continuity on the tempered distributions $\mathcal{S}'(\mathbb{R}^d)$.

Definition 4.2.5 (Weak-$*$ convergence) *Let $\mathcal{S}$ be a metric space with dual $\mathcal{S}'$. A sequence (L_n) in $\mathcal{S}'$ is said to converge to $L \in \mathcal{S}'$ in the weak-$*$ sense if*

$$L_n(f) \xrightarrow{n\to\infty} L(f)$$

holds for all $f \in \mathcal{S}$. We will write $\text{w-lim}_{n\to\infty} L_n = L$.

This notion of convergence implies a notion of continuity and clarifies how to think of continuity for the adjoint operator.

Theorem 4.2.6 *Let $A : \mathcal{S}(\mathbb{R}^d) \longrightarrow \mathcal{S}(\mathbb{R}^d)$ be a linear continuous map. Then for all $L \in \mathcal{S}'(\mathbb{R}^d)$, the map $A' : \mathcal{S}'(\mathbb{R}^d) \longrightarrow \mathcal{S}'(\mathbb{R}^d)$*

$$\big(A'L \, , \, f\big) := \big(L \, , \, Af\big), \qquad f \in \mathcal{S}(\mathbb{R}^d), \tag{4.2.2}$$

defines a weak-$*$ continuous linear map.

Put in the terms of Chapter 3.3, A' is the *adjoint* of A.

Proof First of all, A' is linear and well-defined, $A'L$ maps $f \in \mathcal{S}(\mathbb{R}^d)$ onto $\mathbb{C}$. To show continuity, let $(L_n)_{n\in\mathbb{N}}$ be a sequence of tempered distributions which converges in the weak-$*$ sense to $L \in \mathcal{S}'(\mathbb{R}^d)$. Then

$$\big(A'L_n \, , \, f\big) = \big(L_n \, , \, Af\big) \xrightarrow{n\to\infty} \big(L \, , \, Af\big) = \big(A'L \, , \, f\big)$$

holds for all $f \in \mathcal{S}(\mathbb{R}^d)$ and A' is weak-$*$ continuous. $\square$

As a last consequence, we can extend the convolution from $* : \mathcal{S}(\mathbb{R}^d)\times\mathcal{S}(\mathbb{R}^d) \longrightarrow \mathcal{S}(\mathbb{R}^d)$ to the case where *one* of the factors is a distribution,

$$\begin{aligned} * &: \mathcal{S}'(\mathbb{R}^d) \times \mathcal{S}(\mathbb{R}^d) \longrightarrow \mathcal{S}'(\mathbb{R}^d), \\ * &: \mathcal{S}(\mathbb{R}^d) \times \mathcal{S}'(\mathbb{R}^d) \longrightarrow \mathcal{S}'(\mathbb{R}^d). \end{aligned}$$

For any $f, g, h \in \mathcal{S}(\mathbb{R}^d)$, we can push the convolution from one argument of the duality bracket to the other,

$$\begin{aligned} \big(f * g \, , \, h\big) = \big(g * f \, , \, h\big) &= \int_{\mathbb{R}^d} \mathrm{d}y \, (f * g)(y) \, h(y) \\ &= \int_{\mathbb{R}^d} \mathrm{d}y \int_{\mathbb{R}^d} \mathrm{d}x \, f(x) \, g(y - x) \, h(y) \\ &= \int_{\mathbb{R}^d} \mathrm{d}x \, f(x) \, \big(g(- \, \cdot) * h\big)(x) = \big(f \, , \, g(- \, \cdot) * h\big). \end{aligned}$$

That computation now tells us how to define the convolution of a tempered distribution and a Schwartz function. To remain consistent with how we have defined the convolution of two functions, we need to apply spatial inversion $x \mapsto -x$ to one of the factors, though.

Definition 4.2.7 (Convolution on $\mathcal{S}'(\mathbb{R}^d)$) *Let $L \in \mathcal{S}'(\mathbb{R}^d)$ and $f \in \mathcal{S}(\mathbb{R}^d)$. Then the convolution of L and f is defined as*

$$\big(L * f \, , \, g\big) := \big(L \, , \, f(- \, \cdot) * g\big) \qquad \forall g \in \mathcal{S}(\mathbb{R}^d). \tag{4.2.3}$$

By Theorem 4.2.6, this extension of the convolution is weak-$*$ continuous. Moreover, the convolution has a neutral element in $\mathcal{S}'(\mathbb{R}^d)$, the delta distribution $\delta = \delta_0$: starting from the above definition and two Schwartz functions $f, g \in \mathcal{S}(\mathbb{R}^d)$, we move the convolution to the second argument of the duality bracket, apply the definition of the Dirac distribution and write the resulting integral as a duality bracket again,

$$\begin{aligned} \bigl(\delta * f \, , \, g\bigr) &= \bigl(\delta \, , \, f(- \, \cdot) * g\bigr) = \bigl(f(- \, \cdot) * g\bigr)(0) \\ &= \int_{\mathbb{R}^d} \mathrm{d}y \, f\bigl(-(0 - y)\bigr) \, g(y) = (f, g). \end{aligned}$$

Put succinctly, we have just verified

$$\delta * f = f. \tag{4.2.4}$$

In view of this, we can better understand why approximate identities are also referred to as *Dirac sequences*: equation (4.2.1) allows us to view any Dirac sequence (in $L^1(\mathbb{R}^d)$) as a sequence of tempered distributions. And because $\delta_\varepsilon * f$ converges to $f \in \mathcal{S}(\mathbb{R}^d) \subset L^p(\mathbb{R}^d)$ in L^p norm, the continuity the inclusion map $\imath : L^p(\mathbb{R}^d) \hookrightarrow \mathcal{S}'(\mathbb{R}^d)$ (Example 14 (2)) tells us that $\imath\bigl(\delta_\varepsilon * f\bigr)$ converges also in $\mathcal{S}'(\mathbb{R}^d)$ to f. Therefore, δ_ε converges to the Dirac distribution δ in the distributional sense (but not in $L^1(\mathbb{R}^d)$ as $\delta \not\in L^1(\mathbb{R}^d)$). 2020.12.09

Chapter 5

Oscillatory Integrals

The example (1.3.1) that we started this lecture out with is the matrix element

$$\bigl\langle \varphi, \mathrm{Op}(f)\psi \bigr\rangle := \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathrm{d}x \int_{\mathbb{R}^d} \mathrm{d}y \int_{\mathbb{R}^d} \mathrm{d}\eta \, \mathrm{e}^{-\mathrm{i}\eta\cdot(y-x)} \, \overline{\varphi(x)} \, f\bigl(\tfrac{1}{2}(x+y), \eta\bigr) \, \psi(y)$$

of the pseudodifferential operator $\mathrm{Op}(f)$. We know that *if* we can make sense of the integral on the right and give suitable L^2 estimates, this defines a bounded operator on $L^2(\mathbb{R}^d)$. We will postpone the second part until later and focus on making sense of this integral first.

Clearly, when f is nice enough, e. g. $f \in \mathcal{S}(\mathbb{R}^{2d})$, this double integral makes sense as an absolutely convergent integral:

(1) The invertible linear transformation $(x, y) \mapsto \bigl(\frac{1}{2}(x+y), y-x\bigr)$ maps Schwartz functions to Schwartz functions in a continuous fashion.

(2) The partial Fourier transform continuously maps Schwartz functions onto Schwartz functions, and the latter are absolutely integrable.

But of course, we would like to extend this integral to cases where f is not absolutely integrable.

Let us have a look at another instructive example that encapsulates the adjective "oscillatory" in its purest form, the Fourier transform. *Formally,* we can verify that the inverse Fourier transform $\mathcal{F}^{-1}$ is obtained by flipping the sign of ξ or x in the Fourier transform,

$$\begin{aligned}
\big(\mathcal{F}^{-1}\mathcal{F}f\big)(x) &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathrm{d}\xi \mathrm{e}^{+\mathrm{i}\xi\cdot x} \int_{\mathbb{R}^d} \mathrm{d}y \, \mathrm{e}^{-\mathrm{i}\xi\cdot y} \, f(y) \\
&= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathrm{d}y \, \left(\int_{\mathbb{R}^d} \mathrm{d}\xi \mathrm{e}^{-\mathrm{i}\xi\cdot(y-x)} \right) f(y) \\
&= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathrm{d}y \, (2\pi)^d \, \delta(x-y) \, f(y) = f(x).
\end{aligned}$$

Only the term at $y = x$ contributes, because, heuristically speaking, at this point the phase is constant and equal to 1 independently of our choice of ξ. The integral does not oscillate whatsoever. Whenever $y \neq x$, though, the phase factor oscillates quickly for large $|\xi|$ and intuitively it makes sense that these then average out to 0.

This equation points to a path to make this formal manipulation rigorous for functions which need not be absolutely integrable — we apply the theory of distributions. Many functions $F : \mathbb{R}^d \longrightarrow \mathbb{C}$ define a tempered distribution via

$$(f, g) := \int_{\mathbb{R}^d} \mathrm{d}x \, f(x) \, g(x), \qquad\qquad g \in \mathcal{S}(\mathbb{R}^d).$$

Two examples that come to mind are functions in $f \in L^2(\mathbb{R}^d) \subseteq \mathcal{S}'(\mathbb{R}^d)$ or in $f \in \mathcal{C}^\infty_{\mathrm{u,pol}}(\mathbb{R}^d) \subseteq \mathcal{S}'(\mathbb{R}^d)$, which indeed can be thought of as tempered distributions. For instance, the Fourier transform of the constant function $f(x) = 1$ is $(2\pi)^{d/2} \, \delta(x)$, the delta distribution centered at $x = 0$.

Before we proceed, let us extract the essential features of these two oscillatory integrals: simply put, we are dealing with a double integral

$$I_\vartheta(f, \psi) = \int_{\mathbb{R}^n} \mathrm{d}\xi \int_{\mathbb{R}^d} \mathrm{d}x \, \mathrm{e}^{+\mathrm{i}\vartheta(x,\xi)} \, f(x,\xi) \, \psi(x) \tag{5.0.1}$$

that contains a phase factor $\varphi : \mathbb{R}^d \times \mathbb{R}^n \longrightarrow \mathbb{R}$, a function $f \in \mathcal{C}^\infty(\mathbb{R}^d \times \mathbb{R}^n)$ and a test function $\psi \in \mathcal{S}(\mathbb{R}^d)$.

5.1 Definition of oscillatory integrals

One simple solution to define oscillatory integrals is by means of a limit. For example, if $\chi \in \mathcal{S}(\mathbb{R}^n)$ is a test function with $\chi(0) = 1$, then

$$f_\varepsilon(x,\xi) := \chi(\varepsilon\xi) \, f(x,\xi) \xrightarrow{\varepsilon \to 0} f(x,\xi)$$

converges pointwise to $f(x,\xi)$ for all $(x,\xi) \in \mathbb{R}^d \times \mathbb{R}^n$. For a wide range of functions the prefactor $\chi(\varepsilon\xi)$ makes f_ε absolutely integrable in ξ. Likewise, the absolute integrability of ψ then makes the product

$$(x,\xi) \mapsto f_\varepsilon(x,\xi)\,\psi(x)$$

absolutely integrable in ξ *and* x. This function is often called a *cutoff function*; you can picture it as a function $\chi \in \mathcal{C}^\infty(\mathbb{R}^n, [0,1])$ that takes values in $[0,1]$ for simplicity and equals 1 in a neighborhood of the origin. The latter conditions are not necessary, of course, we will only rely on $\chi(0) = 1$.

That suggests to define the *oscillatory integral* as the limit of the regularized integrals $I_\vartheta(f_\varepsilon, \psi)$.

Definition 5.1.1 (Oscillatory integral) *An oscillatory integral is defined through the limit*

$$I_\vartheta(f,\psi) := \lim_{\varepsilon \to 0} I_\vartheta(f_\varepsilon, \psi) = \lim_{\varepsilon \to 0} \int_{\mathbb{R}^n} \mathrm{d}\xi \int_{\mathbb{R}^d} \mathrm{d}x \, \mathrm{e}^{+\mathrm{i}\vartheta(x,\xi)} \, \chi(\varepsilon\xi) \, f(x,\xi) \, \psi(x), \qquad (5.1.1)$$

where $\chi \in \mathcal{S}(\mathbb{R}^n)$ is any Schwartz function with $\chi(0) = 1$ and provided that the limit exists for all $\psi \in \mathcal{S}(\mathbb{R}^d)$ is independent of the choice of cutoff function χ.

Basically we are exploiting that taking limits and integration in general do *not* commute. For if they did, then the right-hand side would frequently be ∞: the integrand of the regularized integral converges pointwise to $\mathrm{e}^{+\mathrm{i}\vartheta(x,\xi)} \, f(x,\xi)\,\psi(x)$, which is usually not absolutely integrable.

A few words on the structure: the reason we only regularize in ξ is that the presence of the Schwartz function ψ helps control the integral in x. So f and its derivatives need to be polynomially bounded in x. Key for the integrability in ξ is the phase factor $\mathrm{e}^{+\mathrm{i}\vartheta}$ and the regularization via the cutoff function χ.

On the basis of these considerations, we extract three necessary conditions on f:

(a) f must be such that for any $\chi \in \mathcal{S}(\mathbb{R}^n)$ with $\chi(0) = 1$ the function $(x,\xi) \mapsto f_\varepsilon(x,\xi)\,\psi(x) \in L^1(\mathbb{R}^d \times \mathbb{R}^n)$ is integrable for all $\psi \in \mathcal{S}(\mathbb{R}^d)$.

(b) f must be such that the limit (5.1.1) exists, which also places conditions on the phase factor ϑ.

(c) This limit must be independent of our choice of function χ.

5.1.1 Instructional example: the double Fourier transform

Before we develop the general theory, let us check whether we can verify $\mathcal{F}^{-1}\,\mathcal{F}f = f$ on the basis of definition (5.1.1). More precisely, we will be looking at

$$I_{\xi\cdot(x-y)}(f,\psi) := \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\mathrm{d}\xi\int_{\mathbb{R}^d}\mathrm{d}x\int_{\mathbb{R}^d}\mathrm{d}y\,\mathrm{e}^{-\mathrm{i}\xi\cdot(y-x)}\,f(y)\,\psi(x).$$

The space of functions we shall be interested in for the sake of argument are the smooth, *uniformly* polynomially bounded functions $\mathcal{C}^\infty_{\mathrm{u,pol}}(\mathbb{R}^q)$. They consist of smooth functions $f \in \mathcal{C}^\infty(\mathbb{R}^q)$ that satisfy the estimate

$$\bigl|\partial_X^a f(X)\bigr| \leq C_{a\alpha}\,\langle X\rangle^m := C_{a\alpha}\,\sqrt{1+X^2}^{\,m}$$

for some $m \geq 0$ and all multiindices $a \in \mathbb{N}_0^q$ and $X \in \mathbb{R}^q$. The point here is that the power of growth of f and *all* its derivatives is at most of mth degree. Compared with $|X|^m$ the prefactor $\langle X\rangle$ has the added advantage that we do not have to make separate estimates for $|X| \leq R$ (near the origin where $|X|^m$ vanishes) and $|X| \to \infty$. By considering

$$\mathcal{C}^\infty_{\mathrm{u,pol}}(\mathbb{R}^q) := \bigcup_{m\in\mathbb{R}}\mathcal{C}^\infty_{\mathrm{m,pol}}(\mathbb{R}^q)$$

as the union of smooth functions bounded by polynomials of mth degree, we can endow this space with the so-called inductive limit topology [cite].

Furthermore, we can show that $\mathcal{C}^\infty_{\mathrm{u,pol}}(\mathbb{R}^q) \subseteq \mathcal{S}'(\mathbb{R}^q)$ defines a tempered distribution via the integral (4.2.1).

The polynomial growth also tells us that f_ε is a Schwartz function, hence, integrable for all $\varepsilon > 0$: choosing a cutoff function χ and $m = 2n$ large enough and even, we verify that

$$f_\varepsilon(x,y,\xi)\,\psi(x) = \chi(\varepsilon\xi)\,\chi(\varepsilon y)\,f(y)\,\psi(x)$$

is a Schwartz function, hence, absolutely integrable in x, y and ξ. Comparing this with equation (5.1.1), we see that (y,ξ) play the role of ξ and x plays the role of x. Moreover, we *chose* a cutoff function that factors into a product.

Now let us verify that $I_{\xi\cdot(x-y)}(f,\psi) = (f,\psi)$ by hand. Our arguments just showed that f_ε is a Schwartz function. Therefore, we can actually change the order of integration and integrate it up to

$$\begin{aligned}
I_{\xi\cdot(x-y)}(f_\varepsilon,\psi) &= \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\mathrm{d}\xi\int_{\mathbb{R}^d}\mathrm{d}x\int_{\mathbb{R}^d}\mathrm{d}y\,\mathrm{e}^{-\mathrm{i}\xi\cdot(y-x)}\,\chi(\varepsilon\xi)\,\chi(\varepsilon y)\,f(y)\,\psi(x) \\
&= \frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d}\mathrm{d}y\,\chi(\varepsilon y)\,f(y)\int_{\mathbb{R}^d}\mathrm{d}\xi\,\mathrm{e}^{-\mathrm{i}\xi\cdot y}\,\chi(\varepsilon\xi)\,\left(\frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d}\mathrm{d}x\,\mathrm{e}^{+\mathrm{i}\xi\cdot x}\psi(x)\right) \\
&= \int_{\mathbb{R}^d}\mathrm{d}y\,\chi(\varepsilon y)\,f(y)\,\left(\frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d}\mathrm{d}\xi\,\mathrm{e}^{-\mathrm{i}\xi\cdot y}\,\chi(\varepsilon\xi)\,(\mathcal{F}^{-1}\psi)(\xi)\right).
\end{aligned}$$

The inner integral just concerns Schwartz functions, and we can show that

$$\psi_\varepsilon(y) := \frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d}\mathrm{d}\xi\,\mathrm{e}^{-\mathrm{i}\xi\cdot y}\,\chi(\varepsilon\xi)\,(\mathcal{F}^{-1}\psi)(\xi) \xrightarrow{\varepsilon\to 0} \psi(y)$$

converges to $\psi(y)$ pointwise for all $y \in \mathbb{R}^d$. Indeed, as a product of two Schwartz functions, $\xi \mapsto \chi(\varepsilon\xi)\,(\mathcal{F}^{-1}\psi)(\xi)$ is again a Schwartz function and we can bound their product independently of ε by the integrable function

$$\left|\chi(\varepsilon\xi)\,(\mathcal{F}^{-1}\psi)(\xi)\right| \leq \|\chi\|_{00}\,\left|(\mathcal{F}^{-1}\psi)(\xi)\right|.$$

Therefore, Dominated Convergence applies and we can exchange limit and integration to invoke Theorems 4.1.6 and 4.1.9,

$$\begin{aligned}
\lim_{\varepsilon\to 0}\frac{1}{(2\pi)^{d/2}}&\int_{\mathbb{R}^d}\mathrm{d}\xi\,\mathrm{e}^{-\mathrm{i}\xi\cdot y}\,\chi(\varepsilon\xi)\,(\mathcal{F}^{-1}\psi)(\xi) \\
&= \frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d}\mathrm{d}\xi\,\mathrm{e}^{-\mathrm{i}\xi\cdot y}\,\Bigl(\lim_{\varepsilon\to 0}\chi(\varepsilon\xi)\Bigr)\,(\mathcal{F}^{-1}\psi)(\xi) \\
&= \frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d}\mathrm{d}\xi\,\mathrm{e}^{-\mathrm{i}\xi\cdot y}\,(\mathcal{F}^{-1}\psi)(\xi) = \psi(y).
\end{aligned} \tag{5.1.2}$$

Note that the convergence of this integral is not just pointwise, but also in $\mathcal{S}(\mathbb{R}^d)$: the limit of the integrand

$$\xi \mapsto \chi(\varepsilon\xi)\,(\mathcal{F}^{-1}\psi)(\xi) \xrightarrow{\varepsilon\to 0} \mathcal{F}^{-1}\psi \in \mathcal{S}(\mathbb{R}^d)$$

follows from straightforward Fréchet seminorm estimates: derivatives in ξ give factors of $\varepsilon^k \leq 1$, which are harmless. And for the 00-seminorm we exploit that Schwartz functions decay faster than any polynomial towards ∞. So once we pick some $\delta > 0$, we may replace ψ_ε with ψ,

$$\left|\int_{\mathbb{R}^d}\mathrm{d}y\,\chi(\varepsilon y)\,f(y)\,\psi_\varepsilon(y) - \int_{\mathbb{R}^d}\mathrm{d}y\,\chi(\varepsilon y)\,f(y)\,\psi(y)\right| \leq \|f\|_{00}\,\|\chi\|_{00}\,\bigl\|\psi_\varepsilon - \psi\bigr\|_{L^1(\mathbb{R}^d)}.$$

Thanks to Lemma 4.1.3 we can estimate the L^1 norm in terms of Fréchet seminorms. The existence of the limit $\lim_{\varepsilon \to 0} \psi_\varepsilon = \psi$ in the Schwartz functions implies we can make the right-hand side smaller than δ, provided we choose ε small enough,

$$\left| \int_{\mathbb{R}^d} \mathrm{d}y \, \chi(\varepsilon y) \, f(y) \, \psi_\varepsilon(y) - \int_{\mathbb{R}^d} \mathrm{d}y \, \chi(\varepsilon y) \, f(y) \, \psi(y) \right| \leq C_\psi \, \delta.$$

The argument for the convergence of the integral of $\chi(\varepsilon y) \, f(y) \, \psi(y)$ is even easier: once again the limit

$$y \mapsto \chi(\varepsilon y) \, f(y) \, \psi(y) \xrightarrow{\varepsilon \to 0} f \, \psi \in \mathcal{S}(\mathbb{R}^d)$$

exists in $\mathcal{S}(\mathbb{R}^d)$ we can invoke Lemma 4.1.3 to estimate the right-hand side of

$$\left| \int_{\mathbb{R}^d} \mathrm{d}y \, \chi(\varepsilon y) \, f(y) \, \psi(y) - \int_{\mathbb{R}^d} \mathrm{d}y \, f(y) \, \psi(y) \right| \leq \bigl\| \bigl(\chi(\varepsilon \, \cdot \,) - 1 \bigr) \, f \, \psi \bigr\|_{L^1(\mathbb{R}^d)}$$

by finitely many Fréchet seminorms.

Putting everything together yields the existence of the limit

$$\begin{aligned} &\bigl| I_{\xi \cdot (x-y)}(f, \psi) - I_{\xi \cdot (x-y)}(f_\varepsilon, \psi) \bigr| \leq \\ &\qquad \leq \| f \|_{00} \, \| \chi \|_{00} \, \bigl\| \psi_\varepsilon - \psi \bigr\|_{L^1(\mathbb{R}^d)} + \bigl\| \bigl(\chi(\varepsilon \, \cdot \,) - 1 \bigr) \, f \, \psi \bigr\|_{L^1(\mathbb{R}^d)} \xrightarrow{\varepsilon \to 0} 0. \end{aligned}$$

The limit is evidently independent of our choice of cutoff function.

That was a lot of work for very simple statement of fact, and it is clear why it is preferable to develop a general theory rather than having to do these estimates again and again for each integral we encounter.

Before we do that, let us extract a few lessons from this example:

(1) The definition of the oscillatory integral involved a regularization procedure via a "cutoff function".

(2) We used Fubini's Theorem in several places and Dominated Convergence, which suggests that we can use Fubini's Theorem and Dominated Convergence in oscillatory integrals.

(3) Very often we are actually not interested in $I_\vartheta(f, \psi)$, but in the distribution

$$L : \psi \mapsto I_\vartheta(f, \psi).$$

And in the cases we are interested in $L = L_g$ is given in terms of a function, namely

2020.12.16

$$L(\psi) = (g, \psi) = \int_{\mathbb{R}^d} \mathrm{d}x \, g(x) \, \psi(x).$$

5.1.2 The method of L operators

The above showed how cumbersome it is to deal with the definition of oscillatory integrals directly. Fortunately, if we restrict ourselves to a smaller class of admissible functions and make some assumptions on the phase function ϑ more explicit, there exists another way that is frequently used in applications. The class of functions frequently used are called Hörmander symbols.

Definition 5.1.2 (Hörmander symbols) *Let $m \in \mathbb{R}$ and $0 \leq \delta < \rho \leq 1$ or $\rho = 0 = \delta$; m is referred to as the* degree *and (ρ, δ) as the* type*. Then the space of Hörmander symbols is*

$$S^m_{\rho,\delta} \equiv S^m_{\rho,\delta}(\mathbb{R}^d \times \mathbb{R}^n) := \big\{ f \in \mathcal{C}^\infty(\mathbb{R}^d \times \mathbb{R}^n) \;\mid\; \|f\|_{m,a\alpha} < \infty \; \forall a \in \mathbb{N}_0^d, \; \alpha \in \mathbb{N}_0^n \big\},$$

where the Fréchet seminorms are given by

$$\|f\|_{m,a\alpha} := \sup_{(x,\xi) \in \mathbb{R}^d \times \mathbb{R}^n} \Big(\langle \xi \rangle^{m - |\alpha|\rho + |a|\delta} \, \partial_x^a \partial_\xi^\alpha f(x,\xi) \Big), \qquad a \in \mathbb{N}_0^d, \, \alpha \in \mathbb{N}_0^n.$$

When $\rho = 1$ and $\delta = 0$, we abbreviate Hörmander classes with $S^m \equiv S^m(\mathbb{R}^d \times \mathbb{R}^n) := S^m_{1,0}(\mathbb{R}^d \times \mathbb{R}^n)$.

Evidently, Hörmander symbol classes are subclasses of $\mathcal{C}^\infty_{\mathrm{u,pol}}(\mathbb{R}^d \times \mathbb{R}^n) \subseteq \mathcal{S}'(\mathbb{R}^d \times \mathbb{R}^n)$, and therefore can be considered as tempered distributions via the identification (4.2.1). The reason Hörmander classes are useful is that derivatives in ξ will improve its decay properties, and they improve faster than derivatives in x will worsen it ($\delta < \rho$ by assumption). Moreover, for fixed ξ they are bounded functions in x.

A simple example of a Hörmander symbol is a polynomial in ξ,

$$f(x,\xi) = \sum_{k=0}^{m} \sum_{\substack{\alpha \in \mathbb{N}_0^n \\ |\alpha| = k}} c_\alpha \, \xi^\alpha .$$

Since this function is independent of x, for fixed ξ this is indeed trivially a bounded function of x. Each derivative in ξ will reduce the degree of the polynomial by 1, and therefore, we conclude $f \in S^m_{1,0}(\mathbb{R}^d \times \mathbb{R}^n)$.

As indicated in the definition, Hörmander symbols form a Fréchet space; in particular, it is complete with respect to Cauchy sequences defined with respect to the metric (3.4.2).

What is more, Hörmander classes are *nested* in several ways. When $m \leq m'$, then

$$S^m_{\rho,\delta} \subseteq S^{m'}_{\rho,\delta}$$

is naturally contained in the Hörmander class for the larger m'. Similarly, when $\rho' \leq \rho$ or $\delta' \geq \delta$, we have the inclusion

$$S^m_{\rho,\delta} \subseteq S^m_{\rho',\delta'}$$

as e. g. $\rho' \leq \rho$ relaxes conditions on the growth of derivatives of symbols. The above polynomial, for example, is contained in any $S^m_{\rho,\delta}$ where ρ and δ are arbitrary.

Products of Hörmander symbols are Hörmander symbols whose degree is the sum of the two degrees, i. e. the pointwise product of functions $(f, g) \mapsto f\,g$ defines a continuous map

$$S^{m_1}_{\rho,\delta} \times S^{m_2}_{\rho,\delta} \longrightarrow S^{m_1+m_2}_{\rho,\delta}.$$

The space of bounded smooth functions with bounded derivatives $\mathcal{C}^\infty_{\mathrm{b}}(\mathbb{R}^d \times \mathbb{R}^n) = S^0_{0,0}(\mathbb{R}^d \times \mathbb{R}^n)$ is another frequently used symbol class. All of these properties explain why bookkeeping the decay properties is much easier with Hörmander symbols. We will need one more property of Hörmander symbols:

Lemma 5.1.3 *Suppose $f \in S^m_{\rho,\delta}$ and $\chi \in \mathcal{S}(\mathbb{R}^n)$ is a test function with $\chi(0) = 1$. Then the functions $f_\varepsilon(x, \xi) := \chi(\varepsilon\xi)\, f(x, \xi)$ form a sequence $(f_\varepsilon)_{\varepsilon>0}$ in $S^{-\infty}_{\rho,\delta}$ that converges to f in the Fréchet topology of degree $m' > m$,*

$$f_\varepsilon \xrightarrow{\varepsilon\to 0} f \in S^{m'}_{\rho,\delta}.$$

Proof We abbreviate the scaled cutoff function with $\chi(\varepsilon\xi) =: \chi_\varepsilon(\xi)$. Evidently, f_ε decays faster than any polynomial in ξ thanks to the prefactor χ_ε and $f_\varepsilon \in S^{-|m|}_{\rho,\delta}$ for any $m < 0$ and $0 \leq \delta < \rho \leq 1$ or $\rho = 0 = \delta$. Moreover, f_ε converges pointwise to f, and all we need to show is convergence in Fréchet topology.

The proof rests on the idea that while

$$\begin{aligned} \bigl\|1 - \chi_\varepsilon\bigr\|_{0,00} &= \sup_{(x,\xi)\in\mathbb{R}^d\times\mathbb{R}^n} \bigl|1 - \chi(\varepsilon\xi)\bigr| = \sup_{(x,\xi)\in\mathbb{R}^d\times\mathbb{R}^n} \bigl|1 - \chi(\xi)\bigr| \\ &= \|1 - \chi\|_{0,00} \neq 0 \end{aligned}$$

never goes to 0, once we introduce $m > 0$ we actually do know that

$$\bigl\|1 - \chi_\varepsilon\bigr\|_{m,00} = \sup_{(x,\xi)\in\mathbb{R}^d\times\mathbb{R}^n} \Bigl(\langle\xi\rangle^{-m}\,\bigl|1 - \chi(\varepsilon\xi)\bigr|\Bigr) \xrightarrow{\varepsilon\to 0} 0$$

as $\xi \mapsto \langle\xi\rangle^{-m}\mathcal{C}_\infty(\mathbb{R}^n)$ vanishes at infinity. Making ε large enough, we can make $\xi \mapsto \langle\xi\rangle^{-m}\,\bigl(1 - \chi(\varepsilon\xi)\bigr)$ as small as we like since either the first or the second factor becomes small.

Now on to the other seminorms: when $a \neq 0 \in \mathbb{N}_0^d$, then $\|1 - \chi_\varepsilon\|_{m'-m,a\alpha} = 0$ vanishes identically as the function only depends on ξ.

The remaining seminorms are those where $a = 0$ and $\alpha \neq 0 \in \mathbb{N}_0^n$. In view of

$$\bigl(\partial_{\xi_j}\chi_\varepsilon\bigr)(x) = \varepsilon\,\partial_{\xi_j}\chi(\varepsilon x)$$

derivatives in ξ product factors of ε, which implies that these seminorms also vanish in the limit $\varepsilon \to 0$.

$$\bigl\|1 - \chi_\varepsilon\bigr\|_{m,0\alpha} = \|\chi_\varepsilon\|_{m,0\alpha} = \varepsilon^{|\alpha|}\,\|\chi\|_{m,0\alpha} \xrightarrow{\varepsilon\to 0} 0.$$

These arguments establish that $\chi_\varepsilon \to 1 \in S^m_{\rho,\delta}$ converges to the constant function in the Fréchet topology of any Hörmander symbol class of strictly positive order $m > 0$.

The fact that f_ε converges to f in $S^{m'}_{\rho,\delta}$, $m' > m$, follows from the continuity of pointwise multiplication of two Hörmander symbols. Alternatively, we may estimate Fréchet seminorms directly. The 00-seminorm in $S^{m'}_{\rho,\delta}$, $m' > m$ can now be estimated by

$$\|f - f_\varepsilon\|_{m',00} \leq \bigl\|1 - \chi_\varepsilon\bigr\|_{m'-m,00}\,\|f\|_{m,00} \xrightarrow{\varepsilon\to 0} 0,$$

which converges to 0 by the above.

Introducing derivatives in x is trivial, because χ_ε only depends on ξ. Derivatives in ξ distribute amongst the two factors, 2020.12.23

$$\begin{aligned}\bigl\|f - f_\varepsilon\bigr\|_{m,a\alpha} &\leq \sum_{\substack{\beta,\gamma\in\mathbb{N}_0^n\\ \beta+\gamma=\alpha}} \bigl\|1 - \chi_\varepsilon\bigr\|_{m'-m,a\beta}\,\|f\|_{m,a\gamma} \\ &\xrightarrow{\varepsilon\to 0} 0.\end{aligned}$$

□

The utility of Hörmander symbols reveals itself in the simple observation that if $m - |\alpha|\rho < -n$ the Hörmander symbol

$$\partial_\xi^\alpha f \in S^{m-|\alpha|\rho}_{\rho,\delta}$$

becomes absolutely integrable in ξ. Consequently, the product

$$(x,\xi) \mapsto \partial_\xi^\alpha f(x,\xi)\,\psi(x) \in L^1(\mathbb{R}^d \times \mathbb{R}^n)$$

becomes absolutely integrable. But how do we add derivatives to f in $I_\vartheta(f,\psi)$ without, well, simply considering a very different integral?

Let us take some inspiration from a simple phase factor, $\vartheta = x \cdot \xi$. Since the phase is a simple exponential, the operator

$$L = -\frac{\mathrm{i}}{\xi}\,\partial_x$$

satisfies the eigenvalue equation

$$L \mathrm{e}^{+\mathrm{i} x \cdot \xi} = \mathrm{e}^{+\mathrm{i} x \cdot \xi}.$$

Of course, there are plenty of other choices of operators. To avoid (ultimately harmless) singularities around 0, another common choice is

$$L = \langle \xi \rangle^{-1} \sqrt{1 - \Delta_x}.$$

or its square $L^2 = \langle \xi \rangle^{-2} (1 - \Delta_x)$. Such operators are often called "L operators" in the literature. They are highly non-unique, and choosing "wrong" ones like $L = -\frac{\mathrm{i}}{x} \partial_\xi$ (where the roles of x and ξ are reversed) can lead to dead ends.

The idea is now to use partial integration: formally, we can introduce the factor L^{2k} free of charge in the integral,

$$\begin{aligned} I_{x \cdot \xi}(f, \psi) &= \int_{\mathbb{R}^d} \mathrm{d}x \int_{\mathbb{R}^d} \mathrm{d}\xi \, \bigl(L^{2k} \mathrm{e}^{+\mathrm{i} x \cdot \xi} \bigr) \, f(x, \xi) \, \psi(x) \\ &= \int_{\mathbb{R}^d} \mathrm{d}x \int_{\mathbb{R}^d} \mathrm{d}\xi \, \mathrm{e}^{+\mathrm{i} x \cdot \xi} \, \langle \xi \rangle^{-2k} \, \bigl((1 - \Delta_x)^k f(x, \xi) \bigr) \, \psi(x). \end{aligned}$$

This does two things, and we start with the bad news: the function $(1-\Delta_x)^k f \in S^{m+2k\delta}_{\rho,\delta}$ actually has *worse* growth in ξ as $|\xi| \to \infty$. However, this is more than compensated for by the factor $\langle \xi \rangle^{-2k}$: because $\delta < 1$ by assumption, the product

$$\langle \xi \rangle^{-2k} \, (1 - \Delta_x)^k f \in S^{m - 2k(1-\delta)}_{\rho,\delta} \subseteq S^{m - 2k(\rho - \delta)}_{\rho,\delta}$$

is a Hörmander symbol of strictly lower degree. So as soon as $m - 2k(1 - \delta) < -n$, the above becomes absolutely integrable in ξ. Given that the presence of ψ ensures integrability in x, we can use the right-hand side

$$I_{x \cdot \xi}(f, \psi) = \int_{\mathbb{R}^d} \mathrm{d}x \int_{\mathbb{R}^d} \mathrm{d}\xi \, \mathrm{e}^{+\mathrm{i} x \cdot \xi} \, \langle \xi \rangle^{-2k} \, \bigl((1 - \Delta_x)^k f(x, \xi) \bigr) \, \psi(x)$$

to *define* the oscillatory integral — provided we can show that the right-hand side agrees with Definition 5.1.1. Of course, this can be established once we specify the phase factor.

Theorem 5.1.4 *Suppose the phase function ϑ admits an L operator,*

$$L \mathrm{e}^{+\mathrm{i} \vartheta} = \mathrm{e}^{+\mathrm{i} \vartheta}$$

with coefficients $g_j \in S^0_{1,0}$ and $q_k, r \in S^{-1}_{1,0}$ of the form

$$L = \sum_{j=1}^{d} g_j \, \partial_{\xi_j} + \sum_{k=1}^{n} q_k \, \partial_{x_k} + r.$$

Then for all $f \in S^m_{\rho,\delta}$, $0 \leq \delta < \rho \leq 1$, the absolutely convergent integral on the right

$$I_\vartheta(f,\psi) = \int_{\mathbb{R}^d} \mathrm{d}x \int_{\mathbb{R}^n} \mathrm{d}\xi \, \mathrm{e}^{+\mathrm{i}\vartheta(x,\xi)} \, (L^{\mathrm{t}})^k \big(f(x,\xi) \, \psi(x) \big) \tag{5.1.3}$$

agrees with the oscillatory integral (5.1.1)*, where the transpose of the L operator*

$$L^{\mathrm{t}} = \sum_{j=1}^{d} g'_j \, \partial_{\xi_j} + \sum_{k=1}^{n} q'_k \, \partial_{x_k} + r'$$

has the coefficients

$$\begin{aligned} g'_j &= -g_j , \\ q'_k &= -q_k , \\ r' &= r - \sum_{j=1}^{d} \partial_{\xi_j} g_j - \sum_{k=1}^{n} \partial_{x_k} q_k . \end{aligned}$$

Proof The prefactors of the adjoint operator L^{t} can be deduced by formally partially integrating; the name transpose stems from the connection between the (bilinear) duality bracket and the sesquilinear scalar product — the transpose is the complex conjugate of the “Hilbert space” adjoint operator. Another avenue, which justifies calling $L^{\mathrm{t}} = \big(\bar{L}\big)^*$ the transpose operator at the same time, is to rewrite $I_\vartheta(f,\psi)$ as a sesquilinear form akin to a scalar product and then define the transpose operator as the complex conjugate of the adjoint.

The derivatives of L^{t} (which acts on $f \, \psi$) distribute to the two factors, f_ε and ψ. But importantly, f is either derived with respect to ξ, which lowers the degree by ρ, or it is multiplied by a symbol in $S^{-1}_{0,0}$, which lowers the degree by 1. The terms with the worst behavior are those with $q_k \, \partial_{x_k} f_\varepsilon \in S^{m-(1-\delta)}$.

For example, we have

$$L^{\mathrm{t}} f \in S^{m-\mu}_{\rho,\delta} , \qquad \mu = \min(\rho, 1-\delta),$$

is a Hörmander symbol of strictly smaller degree. Choosing $k \in \mathbb{N}_0$ large enough so that $m - k\mu < -n$ makes the right-hand side of (5.1.3) an absolutely convergent integral. The other mixed terms are treated analogously.

It remains to show that this indeed coincides with the fundamental definition (5.1.1) of oscillatory integrals. Let $\chi \in \mathcal{S}(\mathbb{R}^n)$ be a cutoff function with $\chi(0) = 1$ and $\chi_\varepsilon(\xi) := \chi(\varepsilon\xi)$ its ε-scaled version. Then by Lemma 5.1.3 the renormalized symbol $f_\varepsilon = \chi_\varepsilon \, f$ converges to f in $S^{m'}_{\rho,\delta}$ as long as $m' > m$.

What is more, f_ε is integrable in ξ and vanishes at ∞. Therefore, the boundary terms vanish when we integrate

$$\begin{aligned} I_\vartheta(f_\varepsilon, \psi) &= \int_{\mathbb{R}^n} \mathrm{d}\xi \int_{\mathbb{R}^d} \mathrm{d}x \, \mathrm{e}^{+\mathrm{i}\vartheta(x,\xi)} \, f_\varepsilon(x,\xi) \, \psi(x) \\ &= \int_{\mathbb{R}^n} \mathrm{d}\xi \int_{\mathbb{R}^d} \mathrm{d}x \, \bigl(L^k \mathrm{e}^{+\mathrm{i}\vartheta(x,\xi)}\bigr) \, f_\varepsilon(x,\xi) \, \psi(x) \\ &= \int_{\mathbb{R}^n} \mathrm{d}\xi \int_{\mathbb{R}^d} \mathrm{d}x \, \mathrm{e}^{+\mathrm{i}\vartheta(x,\xi)} \, (L^{\mathrm{t}})^k \bigl(f_\varepsilon(x,\xi) \, \psi(x)\bigr) \end{aligned}$$

by parts after introducing the L operator.

The above arguments on the existence of the integral as an absolutely convergent integral still apply verbatim after replacing f with f_ε. And thanks to Lemma 5.1.3 we can exploit the convergence of f_ε and its derivatives to f and its derivatives in the Fréchet topology of $S^{m'-k\mu}_{\rho,\delta}$.

What is more, we can bound $(L^{\mathrm{t}})^k \bigl(f_\varepsilon(x,\xi) \, \psi(x)\bigr)$ (with ε) in absolute value independently of ε in terms of the absolute values of the terms of $(L^{\mathrm{t}})^k \bigl(f(x,\xi) \, \psi(x)\bigr)$ (same term without ε).

Choosing k large enough makes all of the terms absolutely integrable, and we can invoke Dominated Convergence to exchange limit and integration,

$$\begin{aligned} I_\vartheta(f, \psi) &= \lim_{\varepsilon \to 0} I_\vartheta(f_\varepsilon, \psi) \\ &= \lim_{\varepsilon \to 0} \int_{\mathbb{R}^n} \mathrm{d}\xi \int_{\mathbb{R}^d} \mathrm{d}x \, \mathrm{e}^{+\mathrm{i}\vartheta(x,\xi)} \, f_\varepsilon(x,\xi) \, \psi(x) \\ &= \lim_{\varepsilon \to 0} \int_{\mathbb{R}^n} \mathrm{d}\xi \int_{\mathbb{R}^d} \mathrm{d}x \, \mathrm{e}^{+\mathrm{i}\vartheta(x,\xi)} \, (L^{\mathrm{t}})^k \bigl(f_\varepsilon(x,\xi) \, \psi(x)\bigr) \\ &= \int_{\mathbb{R}^n} \mathrm{d}\xi \int_{\mathbb{R}^d} \mathrm{d}x \, \mathrm{e}^{+\mathrm{i}\vartheta(x,\xi)} \, \lim_{\varepsilon \to 0} (L^{\mathrm{t}})^k \bigl(f_\varepsilon(x,\xi) \, \psi(x)\bigr) \\ &= \int_{\mathbb{R}^n} \mathrm{d}\xi \int_{\mathbb{R}^d} \mathrm{d}x \, \mathrm{e}^{+\mathrm{i}\vartheta(x,\xi)} \, (L^{\mathrm{t}})^k \bigl(f(x,\xi) \, \psi(x)\bigr). \end{aligned}$$

This not only shows that the existence of the oscillatory integral and that the value is independent of our choice of χ, but also that it equals the right-hand side of equation (5.1.3). This finishes the proof. $\square$

Example Let us evaluate the oscillatory integral (1.3.1) from the introduction for the special case $f(x,\xi) = \xi^2$ and $d = 1$.Since an operator is uniquely defined by its matrix elements (cf. Sheet 06, Problem 22), the integral $\langle \varphi, \mathrm{Op}(f)\psi \rangle$ serves as a definition for the operator $\mathrm{Op}(f)$.

To set our expectation we use formal manipulations to show that $\mathrm{Op}(\xi^2) = -\partial_x^2$ is just the Laplacian. The function $\xi \mapsto \xi^2 \in \mathcal{C}^\infty_{\mathrm{u,pol}}(\mathbb{R})$ is smooth and polynomially bounded.

Therefore, it defines a tempered distribution via the prescription (4.2.1).The Fourier transform of the tempered distribution ξ^2 is perfectly well-defined and computes to

$$(\mathcal{F}\xi^2)(x-y) = -\sqrt{2\pi}\,\partial_x^2\delta(x-y).$$

The delta distribution eliminates one of the integrals, e. g. $\int_{\mathbb{R}^d} \mathrm{d}y$, takes the second derivative of ψ in y and sets $y = x$,

$$\begin{aligned}
\bigl\langle \varphi, \mathrm{Op}(\xi^2)\psi \bigr\rangle &= \frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{d}x \int_{\mathbb{R}} \mathrm{d}y \int_{\mathbb{R}} \mathrm{d}\eta \, \mathrm{e}^{-\mathrm{i}\eta\cdot(y-x)} \, \overline{\varphi(x)} \, \eta^2 \, \psi(y) \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{d}x \int_{\mathbb{R}} \mathrm{d}y \, \overline{\varphi(x)} \, \bigl(\mathcal{F}\eta^2\bigr)(y-x) \, \psi(y) \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{d}x \int_{\mathbb{R}} \mathrm{d}y \, \overline{\varphi(x)} \, \bigl(-\sqrt{2\pi}\bigr) \, \partial_x^2\delta(y-x) \, \psi(y) \\
&= \int_{\mathbb{R}^d} \mathrm{d}x \, \overline{\varphi(x)} \, \bigl(-\partial_x^2\psi(x)\bigr) = \bigl\langle \varphi, -\partial_x^2\psi \bigr\rangle \, .
\end{aligned}$$

Making these considerations mathematically precise is very easy once we introduce the right L operator, namely

$$L = \langle \eta \rangle^{-1} \, \sqrt{1 - \Delta_y}.$$

Given that we are in $d = 1$, the decay for large $\lvert \eta \rvert$ must be faster than $1/\lvert\eta\rvert$, which can be achieved by applying L four times,

$$\begin{aligned}
\bigl\langle \varphi, \mathrm{Op}(\xi^2)\psi \bigr\rangle &= \frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{d}x \int_{\mathbb{R}} \mathrm{d}y \int_{\mathbb{R}} \mathrm{d}\eta \, \bigl(\langle \eta \rangle^{-4} \, (1-\Delta_y)^2 \mathrm{e}^{-\mathrm{i}\eta\cdot(y-x)}\bigr) \, \overline{\varphi(x)} \, \eta^2 \, \psi(y) \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{d}x \int_{\mathbb{R}} \mathrm{d}y \int_{\mathbb{R}} \mathrm{d}\eta \, \mathrm{e}^{-\mathrm{i}\eta\cdot(y-x)} \, \overline{\varphi(x)} \, \eta^2 \, \langle \eta \rangle^{-4} \, \bigl((1-\Delta_y)^2\psi\bigr)(y).
\end{aligned}$$

Hence, $\bigl\langle \varphi, \mathrm{Op}(\xi^2)\psi \bigr\rangle$ exists as an oscillatory integral. As we shall see in what follows, this suffices to make the above formal manipulations rigorous.

For example, the factor of η^2 can be obtained from deriving the phase factor,

$$\eta^2 \, \mathrm{e}^{-\mathrm{i}\eta\cdot y} = -\partial_y^2 \mathrm{e}^{-\mathrm{i}\eta\cdot y}.$$

These derivatives in y can be pushed over to ψ by partial integration,

$$\ldots = -\frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{d}x \int_{\mathbb{R}} \mathrm{d}y \int_{\mathbb{R}} \mathrm{d}\eta \, \mathrm{e}^{-\mathrm{i}\eta\cdot(y-x)} \, \overline{\varphi(x)} \, \langle \eta \rangle^{-4} \, \bigl((1-\Delta_y)^2 \partial_y^2 \psi\bigr)(y).$$

Note that all integrals have remained integrable since $\psi \in \mathcal{S}(\mathbb{R}^d)$ and $\eta \mapsto \langle \eta \rangle^{-4}$ decays even more rapidly.

The proof shifts the difficulty from proving the existence of the limit (5.1.1) and showing it is independent of our choice of cutoff function χ to finding an L operator.

Proposition 5.1.5 *Suppose the phase* $\vartheta(x, \xi)$

(a) is real-valued,

(b) positively homogeneous of degree 1 *with respect to* ξ*,*

(c) $\vartheta \in \mathcal{C}^\infty$ *for* $\xi \neq 0$*, and*

(d) ϑ *has no critical points with* $\xi \neq 0$*.*

Then there exists an L *operator with the properties enumerated in Theorem 5.1.4.*

Corollary 5.1.6 *Under the above hypotheses on* ϑ *the oscillatory integral* (5.1.1) *exists for all* $f \in S^m_{\rho,\delta}$, $0 \leq \delta < \rho \leq 1$.

Proof By hypothesis the function

$$\varphi := \xi^2 \sum_{j=1}^n \bigl(\partial_{\xi_j} \vartheta\bigr)^2 + \sum_{k=1}^d \bigl(\partial_{x_k} \vartheta\bigr)^2$$

is homogeneous of degree 2 in ξ and $\neq 0$ when $\xi \neq 0$. Consequently, φ^{-1} is homogenous of degree -2 in ξ and $\mathcal{C}^\infty$ away from $\xi = 0$.

To deal with the potential singularity of the derivative at $\xi = 0$, we pick out a cutoff function $\chi \in \mathcal{C}^\infty_c(\mathbb{R}^n)$ with $\chi(x) = 1$ in a neighborhood of $\xi = 0$.

With that we propose to use the L operator,

$$L = \sum_{j=1}^n a_j \, \partial_{\xi_j} + \sum_{k=1}^d b_k \, \partial_{x_k} + c,$$

whose coefficients

$$\begin{aligned}
a_j(x, \xi) &= -\mathrm{i}\bigl(1 - \chi(\xi)\bigr) \, \varphi(x, \xi)^{-1} \, \xi^2 \, \partial_{\xi_j} \vartheta(x, \xi) \in S^0_{1,0}, \\
b_k(x, \xi) &= -\mathrm{i}\bigl(1 - \chi(\xi)\bigr) \, \varphi(x, \xi)^{-1} \, \partial_{x_k} \vartheta(x, \xi) \in S^{-1}_{1,0}, \\
c(x, \xi) &= \chi(\xi) \in S^{-1}_{1,0},
\end{aligned}$$

are defined in terms of the cutoff function and φ. The presence of $1 - \chi$ in the first two terms avoids singularities in the derivatives and ensures that a_j and b_k are smooth functions on *all* of $\mathbb{R}^d \times \mathbb{R}^n$.

Furthermore, for large $|\xi|$ where $1 - \chi(\xi) = 1$ the a_js remains bounded: the quadratic growth caused by ξ^2 is compensated for by the presence of φ^{-1} that decays like $|\xi|^{-2}$.

And each derivative in ξ improves decay by a factor of $|\xi|^{-1}$: derivatives of a_j are sums of products of polynomials in ξ, inverse powers of φ and derivatives of φ. Derivatives in ξ improve the decay of each of these terms, so the claim $a_j \in S^0_{1,0}$ follows from the product rule.

Similarly, the b_ks decay like $|\xi|^{-1}$ as the linear growth of $\partial_{x_k}\vartheta$ is balanced out by the $|\xi|^{-2}$ decay of φ^{-1}. The fact that each derivative in ξ improves decay follows from the same arguments as before.

Lastly, $c = \chi$ is smooth and has compact support; therefore, it belongs to $S^m_{1,0}$ for any $m \in \mathbb{R}$, positive or negative, including $m = -1$.

Consequently, the prefactors satisfy the conditions enumerated in Theorem 5.1.1. It remains to show that $L\mathrm{e}^{+\mathrm{i}\vartheta} = \mathrm{e}^{+\mathrm{i}\vartheta}$, which follows from direct computation: the first two terms combine to give φ and in aggregate cancel the factor φ^{-1},

$$\begin{aligned} L\mathrm{e}^{+\mathrm{i}\vartheta} &= -\mathrm{i}\,(1-\chi)\,\varphi(x,\xi)^{-1}\,\left(\xi^2 \sum_{j=1}^{n} \partial_{\xi_j}\vartheta\,(+\mathrm{i})\,\partial_{\xi_j}\vartheta + \sum_{k=1}^{d} \partial_{x_k}\vartheta\,(+\mathrm{i})\,\partial_{x_k}\vartheta\right)\,\mathrm{e}^{+\mathrm{i}\vartheta} + \\ &\quad + \chi\,\mathrm{e}^{+\mathrm{i}\vartheta} \\ &= (1-\chi)\,\varphi^{-1}\,\varphi\,\mathrm{e}^{+\mathrm{i}\varphi} + \chi\,\mathrm{e}^{+\mathrm{i}\varphi} = \mathrm{e}^{+\mathrm{i}\varphi}. \end{aligned}$$

This concludes the proof. $\square$

5.2 Basic properties of oscillatory integrals

With classical integrals we had to be careful when exchanging limits and integrals or exchanging the order of integration, it was not automatic that we were able to do that without changing the value of the integral. The nice thing about oscillatory integrals is that the necessary conditions are baked right into the definition. Additional assumptions are unnecessary. We could make very precise mathematical statements, but we opt to give a "meta theorem" instead.

Theorem 5.2.1 *For oscillatory integrals we are free to*

(1) change the order of integration (Fubini's Theorem),

(2) exchange limits and oscillatory integration, and

(3) integrate by parts. 2021.01.06

The important thing is that we may use these techniques to make sense of an oscillatory integral, to *define it*.

Remark 5.2.2 ("Meta corollary") The validity of this theorem leads to a few more consequences: under the conditions we ordinarily place on the phase ϑ and the function f, the following holds true:

(1) For oscillatory integrals we are free to derive under the integral sign.

(2) If the phase ϑ_λ and f_λ depend on a parameter λ in a k times continuously differentiable fashion, then so does the oscillatory integral.

The formulation may be vague, e. g. it does not stipulate with respect to which topology $\lambda \mapsto f_\lambda$ is continuous. Nevertheless, they are consequences of us being able to exchange limits and oscillatory integrals.

On the other hand, the proof below explains why: properties of the regularized integral extend to the oscillatory integral. And while for a generic sequence of integrals, the limit need not exist, here it does by assumption.

Proof (1) The proofs of all these statements just go back to the definition: the regularized integral exists as an absolutely convergent integral, so for fixed $\varepsilon > 0$ Fubini's Theorem applies and we may exchange the order of integration, e. g.

$$\begin{aligned}\int_{\mathbb{R}^{n_1}} \mathrm{d}\xi_1 \int_{\mathbb{R}^{n_2}} \mathrm{d}\xi_2 \int_{\mathbb{R}^d} \mathrm{d}x \, \mathrm{e}^{+\mathrm{i}\varphi(x,\xi_1,\xi_2)} \, \chi(\varepsilon\xi_1, \varepsilon\xi_2) \, f(x,\xi_1,\xi_2) \, \psi(x) = \\ = \int_{\mathbb{R}^{n_2}} \mathrm{d}\xi_2 \int_{\mathbb{R}^{n_1}} \mathrm{d}\xi_1 \int_{\mathbb{R}^d} \mathrm{d}x \, \mathrm{e}^{+\mathrm{i}\varphi(x,\xi_1,\xi_2)} \, \chi(\varepsilon\xi_1, \varepsilon\xi_2) \, f(x,\xi_1,\xi_2) \, \psi(x).\end{aligned}$$

For any sequence $\varepsilon_k \to 0$ as $k \to \infty$ the sequences computed from the original order of integration and the alternate order of integration are identical. And the limit $\varepsilon \to 0$ is assumed exist as we are dealing with an oscillatory integral. Consequently, the integral with the reversed order of integration is also an oscillatory integral that converges to the same limit.

(2) That follows from applying the Dominated Convergence Theorem to the regularized integral; since the limit exists by assumption, this then extends as above to the oscillatory integral. Note that if the limit of the integrand is *not* oscillatorily integrable, the same holds for the limit of the oscillatory integrals.

(3) Also here, if we can integrate the regularized integral by parts, then this is inherited by the oscillatory integral. □

To illustrate how to deal with parameter-dependent oscillatory integral, we turn back to (1.3.1) from the introduction.

Example (Weyl quantizations associated with $f \in S^m_{\rho,\delta}$) Equation (1.3.1) defines the matrix elements of the operator $\mathrm{Op}(f)$. The question is whether there exists a more direct way to define

$$\mathrm{Op}(f)\psi(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathrm{d}y \int_{\mathbb{R}^d} \mathrm{d}\eta \, \mathrm{e}^{-\mathrm{i}\eta \cdot (y-x)} \, f\big(\tfrac{1}{2}(x+y), \eta\big) \, \psi(y)$$

as an oscillatory integral, where we regard $x \in \mathbb{R}^d$ as a parameter, and study its properties. An obvious question is whether $\mathrm{Op}(f)\psi \in \mathcal{S}(\mathbb{R}^d)$ is a Schwartz function whenever $\psi \in \mathcal{S}(\mathbb{R}^d)$ is, and whether $\mathrm{Op}(f) : \mathcal{S}(\mathbb{R}^d) \longrightarrow \mathcal{S}(\mathbb{R}^d)$ is a continuous map. We will not answer all these questions at once and just start with one.

Existence for $x \in \mathbb{R}^d$. Evidently, the integrand depends on x in a $\mathcal{C}^\infty$ fashion, and for fixed x the function

$$(y, \eta) \mapsto f\big(\tfrac{1}{2}(x+y), \eta\big) \in S^m_{\rho,\delta}$$

is still a Hörmander symbol of the same order and type as f. Moreover, taking derivatives in x is the same as taking derivatives in y, and the supremum in y to compute the the seminorms is independent of x. Putting this together shows what the x-dependent *linear* coordinate transformation

$$(x, y) \mapsto \tfrac{1}{2}(x+y)$$

can be viewed as a continuous map on $S^m_{\rho,\delta}$. The same is true for the phase, which depends linearly on x, y and η.

We can apply the method of L operators. Here, it is convenient to choose

$$L := \langle \eta \rangle^{-1} \sqrt{1 - \Delta_y}.$$

Each time we add L, we improve decay by $1/|\eta|$ for large η. Derivatives with respect to y worsen decay by $\delta < 1$, but that is more than made up for by the factor $\langle \eta \rangle^{-1}$.

We could have picked derivatives with respect to x instead, but those would distribute amongst $\mathrm{e}^{-\mathrm{i}\xi \cdot (y-x)}$ and f rather than f and ψ.

This L operator does not fit into the scheme of Theorem 5.1.4 directly, we can nevertheless use the exact same arguments as in the proof. In fact, the L operator above is probably amongst the most common choices, not least because the transpose operator $L^{\mathrm{t}} = L$ coincides with L itself. Choosing k large enough, we can make the integrand *absolutely* integrable,

$$\begin{aligned}
\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathrm{d}y \int_{\mathbb{R}^d} \mathrm{d}\eta \, \big(L^{2k} \mathrm{e}^{-\mathrm{i}\eta \cdot y}\big) \, \mathrm{e}^{+\mathrm{i}\eta \cdot x} \, f\big(\tfrac{1}{2}(x+y), \eta\big) \, \psi(y) = \\
= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathrm{d}y \int_{\mathbb{R}^d} \mathrm{d}\eta \, \mathrm{e}^{-\mathrm{i}\eta \cdot y} \, \mathrm{e}^{+\mathrm{i}\eta \cdot x} \, \langle \eta \rangle^{-2k} \, (1 - \Delta_y)^k \Big(f\big(\tfrac{1}{2}(x+y), \eta\big) \, \psi(y) \Big).
\end{aligned}$$

A more detailed argument would involve a regularizing cutoff function, which can be inserted and then removed with the help of Lemma 5.1.3.

The derivatives with respect to y distribute. But for our arguments it suffices to identify the term with the worst behavior. Derivatives of Schwartz functions remain Schwartz functions, so taking derivatives of ψ is harmless. Each derivative of f with respect to y worsens decay by δ. Therefore, the term with the worst behavior is the one where all the derivatives act on f. The assumption $f \in S^m_{\rho,\delta}$ implies that the term

$$(y, \eta) \mapsto \langle \eta \rangle^{-2k} \, \Delta_y^k f\big(\tfrac{1}{2}(x+y), \eta\big) \in S^{m-2k(1-\delta)}_{\rho,\delta}$$

is in a symbol class of order strictly smaller than m. Choosing k so that

$$m - 2k(1-\delta) < -d$$

holds makes the integral absolutely integrable in η.

The nesting of Hörmander classes ensures that also the better-behaved terms with less derivatives in y belong to the same symbol class.

Consequently, we conclude that the oscillatory integral exists. Furthermore, it is continuous and uniformly bounded in the parameter x. The latter follows from the fact that we can estimate the integrand

$$\Big| \langle \eta \rangle^{-2k} \, \partial_y^a f\big(\tfrac{1}{2}(x+y), \eta\big) \, \partial_y^b \psi(y) \Big| \leq C_a \langle \eta \rangle^{m-2k(1-\delta)} \, \big| \partial_y^b \psi(y) \big|$$

independently of x, which shows that we can estimate the supremum in x from above by finitely many seminorms,

$$\big\| \mathrm{Op}(f)\psi \big\|_{00} \leq C \sum_{|a| \leq n(d,m,\delta)} \|\psi\|_{0a},$$

where $n(d, m, \delta)$ is a constant that only depends on the dimension of $\mathbb{R}^d$, the order of the symbol m and $\delta < 1$.

Furthermore, the oscillatory integral inherits the continuity of the integrand in the parameter x. We will continue with this example later on to derive similar estimates for all seminorms of $\mathcal{S}(\mathbb{R}^d)$.

Put another way, our example above shows that

$$\mathrm{Op}(f)\psi : \varphi \mapsto \big(\mathrm{Op}(f)\psi, \varphi\big)$$

is not just a tempered distribution, but is also a continuous bounded function of x. Proving smoothness can (and will be) done with explicit arguments later, but as usual in mathematics it tends to be better to extract more general statements.

5.3 Functions defined by oscillatory integrals and the abbreviated L operator method

Almost always we are not interested in the oscillatory integral itself, but in functions that are defined by them. Let us be more precise: we usually would like to know whether

$$L : \psi \mapsto I_\vartheta(f, \psi)$$

defines a tempered distribution. Not only that, typically this tempered distribution is defined through a smooth, polynomially bounded function $g \in \mathcal{C}^\infty_{\mathrm{pol}}(\mathbb{R}^d)$,

$$I_\vartheta(f, \psi) = (L, \psi) = (g, \psi) \qquad \qquad \forall \psi \in \mathcal{S}(\mathbb{R}^d).$$

Proposition 5.3.1 *Under the assumptions of Theorem 5.1.4 the map*

$$L : \psi \mapsto I_\vartheta(f, \psi) \in \mathcal{S}'(\mathbb{R}^d)$$

defines a tempered distribution.

Proof Theorem 5.1.4 tells us that $I_\vartheta(f, \psi)$ exists as an oscillatory integral for all $\psi \in \mathcal{S}(\mathbb{R}^d)$.

Therefore, it remains to show that we can estimate

$$\bigl|(L, \psi)\bigr| \leq \sum_{a,\alpha \in \mathbb{N}_0^d} C_{a\alpha} \, \lVert \psi \rVert_{a\alpha}$$

in terms of finitely many Fréchet seminorms. These terms come from writing out the *finite* sum

$$\begin{aligned} (L^{\mathrm{t}})^k \bigl(f(x,\xi) \, \psi(x) \bigr) &= \bigl((L^{\mathrm{t}})^k f(x,\xi) \bigr) \, \psi(x) + \text{lower order} && (5.3.1) \\ &= \sum_{\substack{a,b \in \mathbb{N}_0^d \\ \alpha \in \mathbb{N}_0^n \\ |a|+|\alpha|+|\beta| \leq k}} c_{a,b,\alpha}(x,\xi) \, \partial_x^a \partial_\xi^\alpha f(x,\xi) \; \partial_x^b \psi(x), && (5.3.2) \end{aligned}$$

where the coefficients $c_{a,b,\alpha}$ are functions of x and ξ that can be computed from the coefficients of the L^{t} operator and their derivatives. Revisiting the arguments from the proof of Theorem 5.1.4 and exploiting that Hörmander classes are nested, we see that each of the terms belongs to

$$c_{a,b,\alpha} \, \partial_x^a \partial_\xi^\alpha f \in S^{m-k\mu}_{\rho,\delta}$$

where $\mu = \min(\rho, 1 - \delta)$. Put another way, we may estimate this term by

$$\Bigl| c_{a,b,\alpha}(x,\xi) \, \partial_x^a \partial_\xi^\alpha f(x,\xi) \Bigr| \leq C_{a,b,\alpha} \, \langle \xi \rangle^{m - k\mu}. \tag{5.3.3}$$

By our choice of k, this term on the right is integrable. Consequently, we may estimate the oscillatory integral by

$$\bigl| I_\vartheta(f,\psi) \bigr| \leq \sum_{\substack{a,b \in \mathbb{N}_0^d \\ \alpha \in \mathbb{N}_0^n \\ |a| + |b| + |\alpha| \leq N}} C_{a,b,\alpha} \, C_{k,d} \, \lVert \partial_x^\alpha \psi \rVert_{L^1(\mathbb{R}^d)}$$

for some appropriate choice of $N \in \mathbb{N}_0$. Estimating $\lVert \partial_x^\alpha \psi \rVert_{L^1(\mathbb{R}^d)}$ by Fréchet seminorms via Lemma 4.1.3 now yields the claim. □

Very often, we would like to know more, not only that L is a tempered distribution, but a tempered distribution defined in terms of function g. In that case we abuse notation and also call

$$g(x) = \int_{\mathbb{R}^n} \mathrm{d}\xi \, \mathrm{e}^{+\mathrm{i}\vartheta(x,\xi)} \, f(x,\xi) \tag{5.3.4}$$

an oscillatory integral. As usual it is implied that g is related to the original oscillatory integral (5.0.1) via

$$(g, \psi) = I_\vartheta(f, \psi) \qquad \qquad \forall \psi \in \mathcal{S}(\mathbb{R}^d). \tag{5.3.5}$$

This was precisely the strategy spelled out in Chapter 5.1.1: we constructed an oscillatory integral of the form (5.0.1) by writing out the definition of

$$\mathcal{F}^{-1} \mathcal{F} f = f \in \mathcal{S}'(\mathbb{R}^d)$$

on the level of distributions, i. e.

$$\bigl(\mathcal{F}^{-1} \mathcal{F} f \, , \, \psi \bigr) = (f, \psi) \qquad \qquad \forall \psi \in \mathcal{S}(\mathbb{R}^d).$$

In the business of oscillatory integrals, one would then like to establish properties of g. Typically, $g \in \mathcal{C}^\infty_{\mathrm{pol}}(\mathbb{R}^d)$ is a polynomially bounded, smooth function g, e. g. a Hörmander symbol. For the instructional example it was really easy to verify that $\mathcal{F}^{-1} \mathcal{F} f = f \in \mathcal{C}^\infty_{\mathrm{pol}}(\mathbb{R}^d)$ since $\mathcal{F}^{-1} \mathcal{F} = \mathrm{id}_{\mathcal{S}'(\mathbb{R}^d)}$ reduces to the identity. Consequently, we may read off directly that f is smooth and polynomially bounded.

But normally, we need to estimate Fréchet seminorms in order to show that (5.3.4) belongs to some class of functions (e. g. some Hörmander symbol class). The default strategy is to employ the method of L operators in an abbreviated form. Looking at the proof of Theorem 5.1.4, we arrive at the integral

$$I_\vartheta(f,\psi) = \int_{\mathbb{R}^n} \mathrm{d}\xi \int_{\mathbb{R}^d} \mathrm{d}x \, \mathrm{e}^{+\mathrm{i}\vartheta(x,\xi)} \, (L^{\mathrm{t}})^k \big(f(x,\xi) \, \psi(x) \big)$$

where repeated application of the transpose L operators gives us a finite sum

$$(L^{\mathrm{t}})^k \big(f(x,\xi) \, \psi(x) \big) = \big((L^{\mathrm{t}})^k f(x,\xi) \big) \, \psi(x) + \text{lower order}. \tag{5.3.6}$$

We see that the worst term is the first one where the L^{t} operators all act on f. In the context of that theorem, the critical question is whether we can pick $k \in \mathbb{N}_0$ large enough so that

$$\int_{\mathbb{R}^n} \mathrm{d}\xi \, \mathrm{e}^{+\mathrm{i}\vartheta(x,\xi)} \, (L^{\mathrm{t}})^k f(x,\xi)$$

exists as an ordinary Lebesgue integral. In that case the answer was yes.

This generalizes to many situations where the term with the worst behavior is the one where L^{t} acts only on f, and showing the existence of the oscillatory integral from Theorem 5.1.4 is equivalent to proving that for $k \in \mathbb{N}_0$ large enough

$$x \mapsto g(x) = \int_{\mathbb{R}^n} \mathrm{d}\xi \, \mathrm{e}^{+\mathrm{i}\vartheta(x,\xi)} \, (L^{\mathrm{t}})^k f(x,\xi) \in \mathcal{C}(\mathbb{R}^d)$$

exists as an ordinary, absolutely convergent integral. Our assumptions on f further guarantee that we may use a Dominated Convergence argument to commute the limit $x \to x_0$ and the integral in ξ for any $x_0 \in \mathbb{R}^d$; this shows that the integral is continuous in x. What is more, the estimate (5.3.3) is uniform in x, which means that we may bound the supremum of g in x by summing up the terms.

Under assumptions (a)–(c) (excluding (d)) from Proposition 5.1.5 we can show that derivatives in x and the integral commute: at least on the open set

$$X_\vartheta := \big\{ x \in \mathbb{R}^d \; \mid \; \xi \mapsto \vartheta(x,\xi) \text{ has no critical points for } \xi \neq 0 \big\}$$

the oscillatory integral is smooth, i. e. $g \in \mathcal{C}^\infty(X_\vartheta)$. So when we add assumption (d), then $X_\vartheta = \mathbb{R}^d$ and $g \in \mathcal{C}^\infty(\mathbb{R}^d)$ is smooth everywhere.

Corollary 5.3.2 *Suppose the phase function satisfies the assumptions in Proposition 5.1.5 and the function $f \in S^m_{\rho,\delta}$ is a Hörmander symbol of order m with $0 \leq \delta < \rho \leq 1$. Then the tempered distribution*

$$L : \psi \mapsto I_\vartheta(f,\psi) = (g,\psi) \in \mathcal{S}'(\mathbb{R}^d)$$

defined by the oscillatory integral is a smooth, bounded function $g \in \mathcal{C}^\infty(\mathbb{R}^d)$ (but need not *have bounded derivatives).* 2021.01.13

There are situations where we need to estimate other seminorms, e. g. to prove $g \in S^m_{\rho,\delta}$. In the abstract, the strategy looks as follows: suppose we would like to prove that the function $g \in S \subseteq \mathcal{C}^\infty_{\mathrm{pol}}(\mathbb{R}^d)$ defined by (5.3.4), where S is some Fréchet space; the seminorms are typically suprema of derivatives of g times some polynomially bounded function (e. g. $\langle x \rangle$).

(1) Use the (abbreviated) L operator method to show that (5.3.4) exists as long as $k \in \mathbb{N}_0$ is chosen large enough. Note that it is sometimes useful to choose L operators that are of a different form, a common choice is

$$L = \langle \xi \rangle^{-1} \sqrt{1 - \Delta_x}.$$

Estimate the integrand (at least locally) uniformly in x; then Dominated Convergence proves continuity in x.

(2) Write out the seminorms for g, which are typically of the form $x^a \, \partial_x^\alpha g$. Repeat the above arguments for this function. Note that sometimes you need to convert x^a to derivatives of f by exploiting that monomials of x can be obtained by deriving the phase function,

$$x^a \, \mathrm{e}^{+\mathrm{i}\vartheta(x,\xi)} = c(a) \, \partial_\xi^a \mathrm{e}^{+\mathrm{i}\vartheta(x,\xi)}, \qquad\qquad a \in \mathbb{N}_0^d, \; c(a) \in \mathbb{C},$$

and partially integrating (which you may by definition according to Theorem 5.2.1). This is not always possible, of course, but often phase functions take simple forms like $\vartheta(x, \xi) = x \cdot \xi$.

5.4 Asymptotic expansions

Very often oscillatory integrals cannot be computed exactly, and it is useful to develop some systematic approximation techniques. The idea is to expand certain functions

$$f \asymp \sum_{n=0}^{\infty} f_n$$

into an infinite sum. The symbol $\asymp$ emphasizes that the sum on the right need not converge in any sense — we are speaking of *asymptotic expansion*.

In the context of Hörmander symbols, the terms f_n need to be better and better behaved. More precisely, there must exist a non-increasing sequence $m(n) \to -\infty$ that diverges to $-\infty$, which

(a) captures the decay of the terms in ξ,

$$f_n \in S^{m(n)}_{\rho,\delta},$$

(b) and correctly predicts the symbol class of the remainder

$$R_n := f - \sum_{k=0}^{n} f_k \in S^{m(n+1)}_{\rho,\delta}.$$

Very often we need to include a small perturbation parameter ε that orders the term "by size",

$$f \asymp \sum_{n=0}^{\infty} \varepsilon^n f_n.$$

Taylor expansions are a good example of an asymptotic expansion: we can expand any smooth function in terms of its Taylor series. But *a priori* we have no idea whether the Taylor series converges. This is only the case for analytic functions — and smooth functions need not be analytic. Here, the distance $\varepsilon := |x - x_0|$ from the expansion point plays the role of the small parameter, and e. g. in $d = 1$ we may write

$$f(x) \asymp \sum_{k=0}^{n} \frac{1}{k!} \partial_x^k f(x_0)\,(x - x_0)^k + \mathcal{O}\big((x - x_0)^{n+1}\big)$$

But there are cases where $f \neq 0$ even though all derivatives vanish in a point such as

$$f(x) = \begin{cases} \mathrm{e}^{-\frac{1}{1-x^2}} & x \in (-1, +1) \\ 0 & \end{cases}$$

at $x = \pm 1$. For these functions the remainder $R_n(x) = f(x)$ coincide with the original function for all $n \in \mathbb{N}_0$. Such functions are evidently not analytic and the terms of the Taylor expansion do not allow us to approximate the behavior of the function near the expansion point.

In that case, the function $f(x) = \mathcal{O}\big((x - x_0)^\infty\big)$ meaning

$$f(x) = \mathcal{O}\big((x - x_0)^n\big) \quad \forall n \in \mathbb{N}_0.$$

Asymptotic expansions of Hörmander symbols are inherently non-unique, the best we can hope for is

$$f - \sum_{n=0}^{\infty} f_n \in S^{-\infty}_{\rho,\delta}$$

where $S^{-\infty}_{\rho,\delta} := \bigcap_{m \in \mathbb{R}} S^m_{\rho,\delta}$ is the space of smoothing symbols. Its elements belong to *any* of the symbol classes $S^m_{\rho,\delta}$.

We may also read this in reverse: given an asymptotic expansion $\sum_{n=0}^{\infty} f_n$ we can *find* a symbol f, which agrees with the infinite series up to a smoothing symbol.

Proposition 5.4.1 (Proposition 1.1.9 in [Hör71]) *Suppose we are given a series of symbols $(f_n)_{n\in\mathbb{N}_0}$ of decreasing order $(m_n)_{n\in\mathbb{N}_0}$, $m_n \to -\infty$ as $n \to \infty$. Then for each $k \in \mathbb{N}_0$ we can find a symbol $f \in S^{m'_0}_{\rho,\delta}$ of order $m'_0 := \max_{k\in\mathbb{N}_0} m_k$ so that the remainder*

$$f - \sum_{n=0}^{k} f_n \in S^{m'_k}_{\rho,\delta}$$

belongs to the Hörmander symbol class of order $m'_k := \max_{k\leq j} m_j$.

The function f is uniquely determined modulo $S^{-\infty}_{\rho,\delta}$ and has the same property relative to any rearrangement of the series $f \asymp \sum_{n=0}^{\infty} f_n$.

Up until now everything was fairly abstract, so let us get to an example. Consider the Weyl product

$$(f\sharp g)(x,\xi) := \frac{1}{(2\pi)^{2d}} \int_{\mathbb{R}^d} \mathrm{d}y \int_{\mathbb{R}^d} \mathrm{d}\eta \int_{\mathbb{R}^d} \mathrm{d}z \int_{\mathbb{R}^d} \mathrm{d}\zeta \, \mathrm{e}^{+\mathrm{i}\xi\cdot(y+z)} \, \mathrm{e}^{-\mathrm{i}x\cdot(\eta+\zeta)} \, \mathrm{e}^{\mathrm{i}\frac{\varepsilon}{2}(\eta\cdot z - y\cdot\zeta)} \, \cdot$$
$$\cdot \, (\mathcal{F}_\sigma f)(y,\eta) \, (\mathcal{F} g)(z,\zeta) \qquad (5.4.1)$$

of two suitable functions $f, g : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{C}$, where

$$(\mathcal{F}_\sigma f)(x,\xi) := \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathrm{d}x' \int_{\mathbb{R}^d} \mathrm{d}\xi' \, \mathrm{e}^{+\mathrm{i}(\xi\cdot x' - x\cdot\xi')} \, f(x',\xi')$$

is a convenient variant of the Fourier transform that is its own inverse, $\mathcal{F}_\sigma^2 = \mathrm{id}$. We will not show how to obtain an asymptotic expansion of the Weyl product from rather elementary considerations.

Proposition 5.4.2 *Suppose $f \in S^{m_1}_{\rho,\delta}$ and $g \in S^{m_2}_{\rho,\delta}$ are two Hörmander symbols. Then their Weyl product $f\sharp g \in S^{m_1+m_2}_{\rho,\delta}$ has an asymptotic expansion to any order,*

$$f\sharp g = \sum_{k=0}^{n} \varepsilon^k (f\sharp g)_{(k)} + \varepsilon^{n+1} R_n, \qquad (5.4.2)$$

where each of the terms lies in symbol class $S^{m_1+m_2-2n(\rho-\delta)}$ and is explicitly given by

$$(f\sharp g)_{(k)}(x,\xi) = \frac{\mathrm{i}^k}{2^k} \frac{1}{k!} \big(\nabla_\eta \cdot \nabla_z - \nabla_y \cdot \nabla_\zeta\big)^k f(y,\eta)\, g(z,\zeta)\Big|_{\substack{(y,\eta)=(x,\xi)\\(z,\zeta)=(x,\xi)}}. \qquad (5.4.3)$$

The remainder $R_n \in S^{m_1+m_2-2(n+1)(\rho-\delta)}$ is of $\mathcal{O}(1)$ in ε.

Because the purpose of this section is to study asymptotic expansions, we will factor out one important statement:

Lemma 5.4.3 *For each* $m_1, m_2 \in \mathbb{R}$ *and* $0 \leq \delta < \delta \leq 1$ *the Weyl product is a continuous map* $\sharp : S^{m_1}_{\rho,\delta} \times S^{m_2}_{\rho,\delta} \longrightarrow S^{m_1+m_2}_{\rho,\delta}$.

Even though we can prove it with the method of L operators from Chapter 5.1.2, we will postpone that until Chapter 6.

Proof First off, by Lemma 5.4.3 the Weyl product is well-defined as an oscillatory integral and maps Hörmander symbols onto Hörmander symbols.

To obtain the asymptotic expansion, we Taylor expand the "twister" term

$$\mathrm{e}^{\mathrm{i}\frac{\varepsilon}{2}(\eta \cdot z - y \cdot \zeta)} = \sum_{k=0}^{n} \varepsilon^k \, \frac{\mathrm{i}^k}{2^k} \, \frac{1}{k!} \, (\eta \cdot z - y \cdot \zeta)^k + \varepsilon^{n+1} \, r_n(y, \eta, z, \zeta)$$

up to nth order in ε; for the explicit form of the remainder, we may pick e. g. the integral form of the remainder of Taylor series,

$$\begin{aligned} r_n(y, \eta, z, \zeta) &= \frac{1}{\varepsilon^{n+1}} \, \frac{1}{n!} \int_0^1 \mathrm{d}\tau \, (1 - \tau)^n \, \partial_\tau^{n+1} \mathrm{e}^{\tau \, u} \big|_{u = \mathrm{i}\frac{\varepsilon}{2}(\eta \cdot z - y \cdot \zeta)} \\ &= \frac{1}{n!} \, \frac{\mathrm{i}^{n+1}}{2^{n+1}} \, (\eta \cdot z - y \cdot \zeta)^{n+1} \int_0^1 \mathrm{d}\tau \, (1 - \tau)^n \, \mathrm{e}^{\tau \, \mathrm{i}\frac{\varepsilon}{2}(\eta \cdot z - y \cdot \zeta)} = \mathcal{O}(1). \end{aligned}$$

The kth order term of the Taylor expansion yields the kth-order term of the asymptotic expansion,

$$\begin{aligned} (f \sharp g)_{(k)}(x, \xi) := \frac{1}{(2\pi)^{2d}} \int_{\mathbb{R}^d} \mathrm{d}y \int_{\mathbb{R}^d} \mathrm{d}\eta \int_{\mathbb{R}^d} \mathrm{d}z \int_{\mathbb{R}^d} \mathrm{d}\zeta \, \mathrm{e}^{+\mathrm{i}\xi \cdot (y+z)} \, \mathrm{e}^{-\mathrm{i}x \cdot (\eta+\zeta)} \, \cdot \\ \cdot \, \frac{\mathrm{i}^k}{2^k} \, \frac{1}{k!} \, (\eta \cdot z - y \cdot \zeta)^k \, \cdot \\ \cdot \, (\mathcal{F}_\sigma f)(y, \eta) \, (\mathcal{F} g)(z, \zeta) \end{aligned}$$

and the remainder term is defined in terms of the remainder of the Taylor expansion,

$$\begin{aligned} R_n(x, \xi) := \frac{1}{(2\pi)^{2d}} \int_{\mathbb{R}^d} \mathrm{d}y \int_{\mathbb{R}^d} \mathrm{d}\eta \int_{\mathbb{R}^d} \mathrm{d}z \int_{\mathbb{R}^d} \mathrm{d}\zeta \, \mathrm{e}^{+\mathrm{i}\xi \cdot (y+z)} \, \mathrm{e}^{-\mathrm{i}x \cdot (\eta+\zeta)} \, r_n(y, \eta, z, \zeta) \, \cdot \\ \cdot \, (\mathcal{F}_\sigma f)(y, \eta) \, (\mathcal{F} g)(z, \zeta). \end{aligned}$$

The kth-order term can be computed explicitly: up to constants, which are not important for the existence, $(f \sharp g)_{(k)}$ the the linear combination of monomials in y, η, z and ζ, which can be converted into derivatives with respect to η, y, ζ and z (position variables become momenta and vice versa),

$$\begin{aligned}\frac{1}{(2\pi)^{2d}} \int_{\mathbb{R}^d} \mathrm{d}y \int_{\mathbb{R}^d} \mathrm{d}\eta \int_{\mathbb{R}^d} \mathrm{d}z \int_{\mathbb{R}^d} \mathrm{d}\zeta \, \mathrm{e}^{+\mathrm{i}\xi\cdot(y+z)} \, \mathrm{e}^{-\mathrm{i}x\cdot(\eta+\zeta)} \, y^a \, \eta^\alpha \, z^b \, \zeta^\beta \, \cdot \\ \cdot \, (\mathcal{F}_\sigma f)(y,\eta) \, (\mathcal{F} g)(z,\zeta) = \\ = \frac{1}{(2\pi)^{2d}} \int_{\mathbb{R}^d} \mathrm{d}y \int_{\mathbb{R}^d} \mathrm{d}\eta \int_{\mathbb{R}^d} \mathrm{d}z \int_{\mathbb{R}^d} \mathrm{d}\zeta \, \mathrm{e}^{+\mathrm{i}\xi\cdot(y+z)} \, \mathrm{e}^{-\mathrm{i}x\cdot(\eta+\zeta)} \, \cdot \\ \cdot \, (-\mathrm{i})^{|a|+|b|} \, (+\mathrm{i})^{|\alpha|+|\beta|} \, \cdot \\ \cdot \, (\mathcal{F}_\sigma \partial_y^\alpha \partial_\eta^a f)(y,\eta) \, (\mathcal{F} \partial_z^\beta \partial_\zeta^b g)(z,\zeta).\end{aligned}$$

Of the four multiindices, only two are independent. When taking powers of

$$\eta \cdot y - y \cdot \zeta = \sum_{j=1}^d \bigl(\eta_j \, z_j - z_j \, \zeta_j \bigr)$$

we get monomials for which the multiindices $a = \beta$ and $b = \alpha$ are necessarily the same. A nice side benefit is that the prefactor

$$(-\mathrm{i})^{|a|+|\alpha|} \, (+\mathrm{i})^{|\alpha|+|a|} = (-\mathrm{i}^2)^{|\alpha|+|a|} = 1.$$

is always 1.

Refactoring the phase factors we see that this oscillatory integral becomes the Fourier transform of the Fourier transform,

$$\begin{aligned}\ldots &= \frac{1}{(2\pi)^{2d}} \int_{\mathbb{R}^d} \mathrm{d}y \int_{\mathbb{R}^d} \mathrm{d}\eta \int_{\mathbb{R}^d} \mathrm{d}z \int_{\mathbb{R}^d} \mathrm{d}\zeta \, \mathrm{e}^{+\mathrm{i}(\xi\cdot y - x\cdot\eta)} \, \mathrm{e}^{+\mathrm{i}(\xi\cdot z - x\cdot\zeta)} \, \cdot \\ &\qquad \cdot \, (\mathcal{F}_\sigma \partial_y^\alpha \partial_\eta^a f)(y,\eta) \, (\mathcal{F} \partial_z^a \partial_\zeta^\alpha g)(z,\zeta) \\ &= \partial_x^\alpha \partial_\xi^a f(x,\xi) \; \partial_x^a \partial_\xi^\alpha g(x,\xi)\end{aligned}$$

which in this case just reduces to the identity. Consequently, according to our arguments in Chapter 5.1.1 the oscillatory integral reduces to the pointwise product of derivatives of f and g.

All that remains is to insert the proper coefficients and sum up. That gives us the claimed form of the kth-order term. Moreover, as pointwise products of Hörmander symbols, $(f \sharp g)_{(k)}$ is again a Hörmander symbol; counting derivatives, we indeed confirm that

$$(f \sharp g)_{(k)} \in S^{m_1+m_2-2k(\rho-\delta)}_{\rho,\delta}$$

is a symbol of order that is $2k(\rho - \delta)$ lower than the 0th-order term.

We postpone a proof that the remainder belongs to the correct symbol class. But we can already see that this integral can be brought into the same form as the original Weyl product, save for an extra integral over τ — we merely have to convert the monomial into derivatives and identify ε with $\tau\varepsilon$.

The integral over τ does not change the story since we may estimate the integrand in absolute value by 1.

Invoking the auxiliary Lemma 5.4.3 and counting the number of derivatives with respect to x and ξ once more yields

$$R_n \in S^{m_1+m_2-2(n+1)(\rho-\delta)}_{\rho,\delta}.$$

This concludes the proof. $\square$

Let us reiterate the utility of this proof: the asymptotic expansion allows us to *compute* $f \sharp g$ order-by-order up to an error of higher order, and it turns out that the oscillatory integral reduces to the *pointwise product* of functions,

$$f \sharp g = \sum_{k=0}^{n} \sum_{\substack{a,b \in \mathbb{N}_0^d \\ |a|+|b|=2k}} c(a,b)\, \partial_x^a \partial_\xi^b f \; \partial_x^b \partial_\xi^a g.$$

Here, $c(a,b) \in \mathbb{C}$ is just a collection of complex coefficients.

When it comes to deriving asymptotic expansions, choosing a suitable formula for the oscillatory integral is very important. For instance, with a little bit of algebra we can verify that we may *equivalently* write the Weyl product as

$$\begin{aligned}(f \sharp g)(x,\xi) &= \\ &= \frac{1}{(\pi\varepsilon)^{2d}} \int_{\mathbb{R}^d} \mathrm{d}y \int_{\mathbb{R}^d} \mathrm{d}\eta \int_{\mathbb{R}^d} \mathrm{d}z \int_{\mathbb{R}^d} \mathrm{d}\zeta \, \mathrm{e}^{-\mathrm{i}\frac{2}{\varepsilon}((\eta-\xi)\cdot(z-x)-(y-x)\cdot(\zeta-\xi))} \, f(y,\eta)\, g(z,\zeta)\end{aligned}$$

by writing out the Fourier transforms and integrating out the right variables. This second formula for the Weyl product is less conducive to an asymptotic expansion — although it is not impossible. 2021.01.20

Chapter 6

6 Pseudodifferential operators and Weyl calculus

The name “pseudodifferential operators” stems from the desire to define and understand operators like

$$\sqrt{1 - \Delta_x}$$

and fractional derivatives like $|\nabla_x|^{1/2}$. While these simple operators can be defined via functional calculus for the so-called *momentum operator*

$$P = -\mathrm{i}\varepsilon\nabla_x$$

on $L^2(\mathbb{R}^d)$, pseudodifferential theory allows us to connect properties of certain associated functions to properties of the resulting operator. We introduce the parameter ε already here in anticipation of the asymptotic expansions we will study in Chapter 6.3.

There is also a second motivation that comes from the so-called *quantization problem* from physics: the idea is to generate quantum mechanical observables, selfadjoint operators on a Hilbert space, from classical observables, functions on phase space. Observables are mathematical representations of quantities that can be measured in experiment.

In the cases we are considering here, the relevant Hilbert space is $L^2(\mathbb{R}^d)$ and phase space is $\mathbb{R}^d \times \mathbb{R}^d$. For simple classical observables such as classical angular momentum

$$l(q, p) = q \times p,$$

we can replace position q_j by the *position operator*

$$Q_j : \varphi \mapsto (Q_j\varphi)(x) = x_j\, \varphi(x)$$

and classical momentum p_j by $P_j = -\mathrm{i}\varepsilon\nabla_j$. For brevity, we will skip over technical complications that arise due to the unboundedness of Q_j and P_j; these can be solved. This heuristic procedure only works for very simple classical observables, e. g. those of the form $f(q,p) = k(p) + V(q)$ or polynomials of low order. The quantum mechanical angular momentum operator

$$L = l(Q,P) = Q \times P = L^*$$

is of this category, once we insist that suitable real-valued functions are mapped onto selfadjoint operators, this mapping *should* be unique.

This is how quantum hamiltonians (the energy observable) can by systematically "guessed" from classical hamilton functions; these are particularly important, because the energy observable generates the time evolution. While the quantization of the non-relativistic Hamilton function

$$h(q,p) = \tfrac{1}{2m} p^2 + V(q)$$

is unambiguous and yields the Schrödinger operator

$$\begin{aligned} H = h(Q,P) &= \tfrac{1}{2m} P^2 + V(Q) \\ &= \frac{\varepsilon^2}{2m} \Delta_x + V(\hat{x}), \end{aligned}$$

there are physical systems where this is not obvious. An example is the Hamilton function for a semirelativistic particle in a magnetic field,

$$h(q,p) = \sqrt{m^2 + \big(p - A(q)\big)^2} + V(q).$$

That is because the operators Q and P do not commute,

$$\mathrm{i}\big[Q_j, Q_k\big] = 0 \qquad \mathrm{i}\big[P_j, P_k\big] = 0 \qquad \mathrm{i}\big[P_j, Q_k\big] = \varepsilon\delta_{jk}.$$

To extend this prescription to more general functions, we will use a variant of the Fourier transform called the *symplectic Fourier transform,*

$$(\mathcal{F}_\sigma f)(X) := \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathrm{d}X' \, \mathrm{e}^{\mathrm{i}\sigma(X,X')} \, f(X')$$

defined with the help of the so-called symplectic form

$$\sigma(X,Y) := \sigma((x,\xi),(y,\eta)) := \xi \cdot y - x \cdot \eta$$

that has the advantage that it is its own inverse, $\mathcal{F}_\sigma^2 = \mathrm{id}$. To simplify notation, we will label phase space coordinates such as $X = (x,\xi) \in \mathbb{R}^d \times \mathbb{R}^d$, $Y = (y,\eta)$ and $Z = (z,\zeta)$

with capital letters, positions with small letters (x, y and z) and momenta with greek letters. These pairings makes it easier to understand integrals.

Now we propose the solution

$$f(Q,P) := \mathrm{Op}(f) := \frac{1}{(2\pi)^d} \int_{\mathbb{R}^{2d}} \mathrm{d}X \, \mathrm{e}^{-\mathrm{i}\sigma(X,(Q,P))} \, (\mathcal{F}_\sigma f)(X) \tag{6.0.1}$$

which *formally* just corresponds to writing out $\mathcal{F}_\sigma^2 f$ and "evaluating it" at (Q,P). Of course, Q and P are operators, so the expression on the right actually *defines* $f(Q,P)$.

Note that we already know to make sense of the integral on the left at least when $f \in L^1(\mathbb{R}^{2d})$ and its symplectic Fourier transform are integrable: the operator

$$W(X) := \mathrm{e}^{-\mathrm{i}\sigma(X,(Q,P))} \tag{6.0.2}$$

turns out to be unitary and therefore has norm 1. Consequently, the integral on the right-hand side of equation (6.0.1) is a *Bochner integral* (cf. Sheet xyz, Problem xyz).

The purpose of this section is to extend Weyl quantization (6.0.1) in a systematic fashion from Schwartz functions to more general functions such as $h(q,p) = p^2$. We do this by duality as in Chapter 4. Then we will show how to use the theory of oscillatory integrals when we discuss pseudodifferential operators associated to Hörmander symbols.

6.1 Weyl calculus on $\mathcal{S}$

- Derive explicit expression for $f \in \mathcal{S}(\mathbb{R}^d \times \mathbb{R}^d)$ rigorously
- Introduce the operator kernel
- Explain the other option: matrix elements approach
- Derive Weyl product

We will first develop Weyl quantization on $\mathcal{S}(\mathbb{R}^d \times \mathbb{R}^d)$. As already mentioned for Schwartz functions we can define $\mathrm{Op}(f)$ as a Bochner integral. However, this approach is a dead end if we intend to extend it to functions that are not Schwartz functions (or are integrable with an integrable Fourier transform). This is in anticipation of extending it to $\mathcal{S}'(\mathbb{R}^d \times \mathbb{R}^d)$ by duality.

6.1.1 The Weyl system

The first step is to study the action of the Weyl system $W(X)$ on L^2 functions. Instead of working with Q_ε and P, each of which defines a selfadjoint operator on its maximal domain

$$\begin{aligned}\mathcal{D}(Q_{\varepsilon l}) &= \bigl\{\varphi \in L^2(\mathbb{R}^d) \mid Q_{\varepsilon l}\varphi \in L^2(\mathbb{R}^d)\bigr\} \supset \mathcal{S}(\mathbb{R}^d)\\ \mathcal{D}(P_l) &= \bigl\{\varphi \in L^2(\mathbb{R}^d) \mid P_l\varphi \in L^2(\mathbb{R}^d)\bigr\} \supset \mathcal{S}(\mathbb{R}^d),\end{aligned}$$

we work with the corresponding evolution groups, namely translations in momentum

$$V(\eta) := \mathrm{e}^{-\mathrm{i}\eta\cdot Q_\varepsilon},\ \bigl(V(\eta)\varphi\bigr)(x) = \mathrm{e}^{-\mathrm{i}\eta\cdot x}\varphi(x), \qquad \varphi \in L^2(\mathbb{R}^d),$$

and position

$$U(y) := \mathrm{e}^{-\mathrm{i}y\cdot P},\ \bigl(U(y)\varphi\bigr)(x) = \varphi(x - \varepsilon y), \qquad \varphi \in L^2(\mathbb{R}^d).$$

These are not *one*-parameter groups, but *strongly continuous* d-parameter groups: one can check that

$$\begin{aligned} V(\xi)V(\eta) &= V(\xi+\eta), & V(\xi)^* &= V(-\xi),\\ U(x)U(y) &= U(x+y), & U(x)^* &= U(-x),\end{aligned}$$

hold as all the different components of position and momentum commute. But what about combinations of U and V? Clearly the origin of the fact

$$\begin{aligned}\bigl(U(y)V(\eta)\varphi\bigr)(x) &= \bigl(V(\eta)\varphi\bigr)(x-\varepsilon y) = \mathrm{e}^{-\mathrm{i}\eta\cdot(x-\varepsilon y)}\varphi(x-\varepsilon y)\\ &\neq \mathrm{e}^{-\mathrm{i}\eta\cdot x}\varphi(x-\varepsilon y) = \bigl(V(\eta)U(y)\varphi\bigr)(x)\end{aligned}$$

can be traced back to the noncommutativity of the generators of U and V. Instead, we have just shown

$$U(y)V(\eta) = \mathrm{e}^{+\mathrm{i}\varepsilon y\cdot\eta}\, V(\eta)U(y) \qquad \forall y \in \mathbb{R}^d, \eta \in \mathbb{R}^d.$$

It turns out that picking an operator ordering can be rephrased into picking an ordering of U and V. Since we would like real-valued functions to be quantized to potentially selfadjoint operators, we choose the "symmetrically" defined

Definition 6.1.1 (Weyl system) *For all $X \in T^*\mathbb{R}^d$, we define*

$$W(X) := \mathrm{e}^{-\mathrm{i}(\xi\cdot Q_\varepsilon - x\cdot P)} =: \mathrm{e}^{-\mathrm{i}\sigma((x,\xi),(Q_\varepsilon,P))}$$

where $\sigma(X,Y) := \xi\cdot y - x\cdot\eta$.

We note that the small parameter ε is contained in the definition of the operators Q_ε and P. The next Lemma tells us how this operator acts on wave functions:

Lemma 6.1.2 *For all $Y \in T^*\mathbb{R}^d$ and $\varphi \in L^2(\mathbb{R}^d)$, $\bigl(W(Y)\varphi\bigr)(x) = \mathrm{e}^{-\mathrm{i}\eta\cdot(x+\frac{\varepsilon}{2}y)}\varphi(x+\varepsilon y)$ holds. The map $Y \mapsto W(Y)$ is strongly continuous.*

Proof We use the Trotter product formula ([RS72, Theorem VIII.31]) to write $W(Y)$ as

$$\begin{aligned}
&\operatorname*{s-lim}_{n\to\infty}\Bigl(\mathrm{e}^{-\frac{\mathrm{i}}{n}\eta\cdot Q_\varepsilon}\mathrm{e}^{+\frac{\mathrm{i}}{n}y\cdot P}\Bigr)^n = \\
&\quad = \operatorname*{s-lim}_{n\to\infty}\Bigl(\mathrm{e}^{-\frac{\mathrm{i}}{n}\eta\cdot Q_\varepsilon}\mathrm{e}^{+\frac{\mathrm{i}}{n}y\cdot P}\mathrm{e}^{-\frac{\mathrm{i}}{n}\eta\cdot Q_\varepsilon}\mathrm{e}^{+\frac{\mathrm{i}}{n}y\cdot P}\cdots\mathrm{e}^{-\frac{\mathrm{i}}{n}\eta\cdot Q_\varepsilon}\mathrm{e}^{+\frac{\mathrm{i}}{n}y\cdot P}\mathrm{e}^{-\frac{\mathrm{i}}{n}\eta\cdot Q_\varepsilon}\mathrm{e}^{+\frac{\mathrm{i}}{n}y\cdot P}\Bigr) \\
&\quad = \operatorname*{s-lim}_{n\to\infty}\Bigl(\mathrm{e}^{-\frac{\mathrm{i}}{n}\eta\cdot Q_\varepsilon}\mathrm{e}^{+\frac{\mathrm{i}}{n}y\cdot P}\mathrm{e}^{-\frac{\mathrm{i}}{n}\eta\cdot Q_\varepsilon}\mathrm{e}^{-\frac{\mathrm{i}}{n}y\cdot P}\mathrm{e}^{+\mathrm{i}\frac{2}{n}y\cdot P}\mathrm{e}^{-\frac{\mathrm{i}}{n}\eta\cdot Q_\varepsilon}\mathrm{e}^{-\mathrm{i}\frac{2}{n}y\cdot P}\mathrm{e}^{+\mathrm{i}\frac{3}{n}y\cdot P}\cdots \\
&\qquad\qquad \cdots\mathrm{e}^{+\mathrm{i}\frac{n-1}{n}y\cdot P}\mathrm{e}^{-\frac{\mathrm{i}}{n}\eta\cdot Q_\varepsilon}\mathrm{e}^{-\mathrm{i}\frac{n-1}{n}y\cdot P}\mathrm{e}^{+\mathrm{i}\frac{n}{n}y\cdot P}\Bigr).
\end{aligned}$$

The terms can be simplified using

$$\begin{aligned}
\Bigl(\mathrm{e}^{+\mathrm{i}\frac{k}{n}y\cdot P}\mathrm{e}^{-\frac{\mathrm{i}}{n}\eta\cdot Q_\varepsilon}\mathrm{e}^{-\mathrm{i}\frac{k}{n}y\cdot P}\varphi\Bigr)(x) &= \Bigl(\mathrm{e}^{-\frac{\mathrm{i}}{n}\eta\cdot Q_\varepsilon}\mathrm{e}^{-\mathrm{i}\frac{k}{n}y\cdot P}\varphi\Bigr)\bigl(x+\varepsilon\tfrac{k}{n}y\bigr) \\
&= \mathrm{e}^{-\frac{\mathrm{i}}{n}\eta\cdot(x+\varepsilon\frac{k}{n}y)}\,\bigl(\mathrm{e}^{-\mathrm{i}\frac{k}{n}y\cdot P}\varphi\bigr)\bigl(x+\varepsilon\tfrac{k}{n}y\bigr) \\
&= \mathrm{e}^{-\frac{\mathrm{i}}{n}\eta\cdot(x+\varepsilon\frac{k}{n}y)}\,\varphi(x) = \bigl(\mathrm{e}^{-\frac{\mathrm{i}}{n}\eta\cdot(Q_\varepsilon+\varepsilon\frac{k}{n}y)}\varphi\bigr)(x)
\end{aligned}$$

where $k \in \{0,1,\ldots,n-1\}$. This means the above expression can be rewritten as

$$\begin{aligned}
&\operatorname*{s-lim}_{n\to\infty}\Bigl(\mathrm{e}^{-\mathrm{i}\frac{1}{n}\eta\cdot Q_\varepsilon}\mathrm{e}^{-\mathrm{i}\frac{1}{n}\eta\cdot(Q_\varepsilon+\varepsilon\frac{1}{n}y)}\cdots\mathrm{e}^{-\mathrm{i}\frac{1}{n}\eta\cdot(Q_\varepsilon+\varepsilon\frac{n-1}{n}y)}\mathrm{e}^{+\mathrm{i}y\cdot P}\Bigr) = \\
&\qquad = \operatorname*{s-lim}_{n\to\infty}\mathrm{e}^{-\mathrm{i}\frac{1}{n}\eta\cdot\sum_{k=0}^{n-1}(Q_\varepsilon+\varepsilon\frac{k}{n}y)}\mathrm{e}^{+\mathrm{i}y\cdot P}.
\end{aligned}$$

The sum in the exponential is a Riemann sum that converges to

$$\frac{1}{n}\sum_{k=0}^{n-1}\bigl(x+\varepsilon\tfrac{k}{n}y\bigr) \xrightarrow{n\to\infty} \int_0^1 \mathrm{d}s\,\bigl(x+s\varepsilon y\bigr) = x+\tfrac{\varepsilon}{2}y$$

so that we get

$$W(Y) = \mathrm{e}^{-\mathrm{i}\eta\cdot(Q_\varepsilon+\frac{\varepsilon}{2}y)}\mathrm{e}^{+\mathrm{i}y\cdot P}.$$

The strong continuity of $Y \mapsto W(Y)$ follows from the strong continuity of $y \mapsto U(y)$ and $\eta \mapsto V(\eta)$. □

Remark 6.1.3 Another way to see this is via the Baker-Campbell-Hausdorff formula: formally, we have

$$\begin{aligned}
W(Y) = \mathrm{e}^{-\mathrm{i}(\eta\cdot Q_\varepsilon-y\cdot P)} &= \mathrm{e}^{-\mathrm{i}\eta\cdot Q_\varepsilon}\mathrm{e}^{+\mathrm{i}y\cdot P}\mathrm{e}^{-\frac{i^2}{2}[\eta\cdot Q_\varepsilon,-y\cdot P]} = \mathrm{e}^{-\mathrm{i}\eta\cdot Q_\varepsilon}\mathrm{e}^{+\mathrm{i}y\cdot P}\mathrm{e}^{-\mathrm{i}\frac{\varepsilon}{2}\eta\cdot y} \\
&= \mathrm{e}^{-\mathrm{i}\eta\cdot(Q_\varepsilon+\frac{\varepsilon}{2}y)}\mathrm{e}^{+\mathrm{i}y\cdot P}.
\end{aligned}$$

However, it is a bit tricky to rigorously control commutators of unbounded operators [RS72, Chapter VIII.5]. 2009.12.08

The Weyl system combines translations in real and reciprocal space. If one of the two components is 0, then W reduces to U or V,

$$\begin{aligned} W(x,0) &= U(-x) = U(x)^*, \\ W(0,\xi) &= V(\xi). \end{aligned}$$

The Weyl system completely encodes the commutation relations:

Proposition 6.1.4 *Let $Y, Z \in T^*\mathbb{R}^d$. Then, the Weyl system obeys the following composition law:*

$$W(Y)W(Z) = \mathrm{e}^{+\mathrm{i}\frac{\varepsilon}{2}\sigma(Y,Z)}\, W(Y+Z)$$

This contains the commutation relations: if we combine a translation along e_k in real space and e_j in reciprocal space,

$$W(e_k,0)W(0,e_j) = \mathrm{e}^{-\mathrm{i}\frac{\varepsilon}{2}\delta_{kj}}\, W(e_k,e_j),$$

then the extra factor reduces to the exponential of half of $[Q_{\varepsilon k}, P_j] = -\mathrm{i}\varepsilon\delta_{kj}$.

Proof This follows from direct calculation: for all $\varphi \in L^2(\mathbb{R}^d)$, we have

$$\begin{aligned} \bigl(W(Y)W(Z)\varphi\bigr)(x) &= \mathrm{e}^{-\mathrm{i}(x+\frac{\varepsilon}{2}y)\cdot\eta}\bigl(W(Z)\varphi\bigr)(x+\varepsilon y) \\ &= \mathrm{e}^{-\mathrm{i}(x+\frac{\varepsilon}{2}y)\cdot\eta}\mathrm{e}^{-\mathrm{i}(x+\varepsilon y+\frac{\varepsilon}{2}z)\cdot\zeta}\varphi(x+\varepsilon y+\varepsilon z) \\ &= \mathrm{e}^{+\mathrm{i}\frac{\varepsilon}{2}(-y\cdot\zeta+z\cdot\eta)}\mathrm{e}^{-\mathrm{i}(x+\frac{\varepsilon}{2}(y+z))\cdot(\eta+\zeta)}\varphi\bigl(x+\varepsilon(y+z)\bigr) \\ &= \mathrm{e}^{+\mathrm{i}\frac{\varepsilon}{2}\sigma(Y,Z)}\,\bigl(W(Y+Z)\varphi\bigr)(x). \end{aligned}$$

Hence, the claim follows. □

6.1.2 Weyl quantization

To define the Weyl quantization, we first introduce a convenient variant of the Fourier transform, the symplectic Fourier transform on $T^*\mathbb{R}^d$

$$(\mathcal{F}_\sigma f)(X) := \frac{1}{(2\pi)^d}\int_{T^*\mathbb{R}^d} \mathrm{d}Y\, \mathrm{e}^{+\mathrm{i}\sigma(X,Y)}\, f(Y), \qquad f \in \mathcal{S}(T^*\mathbb{R}^d).$$

$\mathcal{F}_\sigma$ has the nice property of being an involution, i. e. $\mathcal{F}_\sigma^2 = \mathrm{id}_{\mathcal{S}}$. If we pretend for a moment that Q_ε and P are *variables* and *not* operators, we have

$$\begin{aligned}
f(Q_\varepsilon, P) = (\mathcal{F}_\sigma^2 f)(Q_\varepsilon, P) &= \frac{1}{(2\pi)^d} \int_{T^*\mathbb{R}^d} \mathrm{d}X \, \mathrm{e}^{+\mathrm{i}\sigma((Q_\varepsilon,P),X)} \, \frac{1}{(2\pi)^d} \int_{T^*\mathbb{R}^d} \mathrm{d}Y \, \mathrm{e}^{+\mathrm{i}\sigma(X,Y)} \, f(Y) \\
&= \frac{1}{(2\pi)^d} \int_{T^*\mathbb{R}^d} \mathrm{d}X \, (\mathcal{F}_\sigma f)(X) \, \mathrm{e}^{-\mathrm{i}\sigma(X,(Q_\varepsilon,P))} \\
&= \frac{1}{(2\pi)^{2d}} \int_{T^*\mathbb{R}^d} \mathrm{d}X \int_{T^*\mathbb{R}^d} \mathrm{d}Y \, \mathrm{e}^{+\mathrm{i}\sigma(X,Y-(Q_\varepsilon,P))} \, f(Y) = \int_{T^*\mathbb{R}^d} \mathrm{d}Y \, \delta\big(Y - (Q_\varepsilon, P)\big) \, f(Y).
\end{aligned}$$

Even though the above does not make sense if Q_ε and P are operators, it gives an intuition why we define Weyl quantization the way we do:

Definition 6.1.5 (Weyl quantization) *Let $f \in \mathcal{S}(T^*\mathbb{R}^d)$. Then the Weyl quantization of f is defined as*

$$\mathrm{Op}(f)\varphi := \frac{1}{(2\pi)^d} \int_{T^*\mathbb{R}^d} \mathrm{d}X \, (\mathcal{F}_\sigma f)(X) \, W(X)\varphi, \qquad \forall \varphi \in \mathcal{S}(\mathbb{R}^d).$$

Lemma 6.1.2 essentially already tells us how $\mathrm{Op}(f)$ acts on wave functions:

Proposition 6.1.6 *Let $f \in \mathcal{S}(T^*\mathbb{R}^d)$ and $\varphi \in L^2(\mathbb{R}^d)$. Then $\mathrm{Op}(f)$ defines a bounded operator on $L^2(\mathbb{R}^d)$ and its action on φ is given by*

$$\begin{aligned}
\big(\mathrm{Op}(f)\varphi\big)(x) &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathrm{d}y \int_{\mathbb{R}^d} \mathrm{d}\eta \, \mathrm{e}^{-\mathrm{i}(y-x)\cdot\eta} \, f\big(\tfrac{1}{2}(x+y), \varepsilon\eta\big) \, \varphi(y) \qquad (6.1.1) \\
&= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}y \, \varepsilon^{-d} \, (\mathcal{F}_2 f)\big(\tfrac{1}{2}(x+y), \tfrac{y-x}{\varepsilon}\big) \, \varphi(y) \\
&=: \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}y \, K_f(x,y) \, \varphi(y) =: \big(\mathfrak{Int}(K_f)\varphi\big)(x).
\end{aligned}$$

Proof Let $\varphi \in \mathcal{S}(\mathbb{R}^d) \subset L^2(\mathbb{R}^d)$. Then we can bound $\mathrm{Op}(f)\varphi$ by

$$\begin{aligned}
\big\|\mathrm{Op}(f)\varphi\big\| &\leq \frac{1}{(2\pi)^d} \int_{T^*\mathbb{R}^d} \mathrm{d}X \, \big|(\mathcal{F}_\sigma f)(X)\big| \, \big\|W(X)\varphi\big\| = \left(\frac{1}{(2\pi)^d} \int_{T^*\mathbb{R}^d} \mathrm{d}X \, \big|(\mathcal{F}_\sigma f)(X)\big| \right) \|\varphi\| \\
&= (2\pi)^{-d} \, \|\mathcal{F}_\sigma f\|_{L^1(T^*\mathbb{R}^d)} \, \|\varphi\| \, .
\end{aligned}$$

The L^1-norm of $\mathcal{F}_\sigma f$ is certainly bounded as $\mathcal{F}_\sigma f$ is a Schwartz function and thus integrable. Hence, $\mathrm{Op}(f)$ defines a bounded operator on $\mathcal{S}(\mathbb{R}^d) \subset L^2(\mathbb{R}^d)$. Since the Schwartz functions are dense in $L^2(\mathbb{R}^d)$ by Lemma **??**, we invoke Theorem 3.1.6 which ensures the existence of a unique extension of $\mathrm{Op}(f)$ to all of $L^2(\mathbb{R}^d)$.

Equation (6.1.1) follows from direct computation and Lemma 6.1.2: for any $\varphi \in L^2(\mathbb{R}^d)$, we have

$$\begin{aligned}\big(\mathrm{Op}(f)\varphi\big)(x) &= \frac{1}{(2\pi)^{2d}} \int_{T^*\mathbb{R}^d} \mathrm{d}Y \int_{T^*\mathbb{R}^d} \mathrm{d}Z \, \mathrm{e}^{+\mathrm{i}\sigma(Y,Z)} \, f(Z) \, \big(W(Y)\varphi\big)(x) \\ &= \frac{1}{(2\pi)^{2d}} \int_{\mathbb{R}^d} \mathrm{d}y \int_{\mathbb{R}^d} \mathrm{d}\eta \int_{\mathbb{R}^d} \mathrm{d}z \int_{\mathbb{R}^d} \mathrm{d}\zeta \, \mathrm{e}^{+\mathrm{i}(\eta \cdot z - y \cdot \zeta)} \, f(z,\zeta) \cdot \\ &\qquad \cdot \mathrm{e}^{-\mathrm{i}\eta \cdot (x + \frac{\varepsilon}{2} y)} \, \varphi\big(x + \varepsilon y\big) \\ &= \frac{\varepsilon^{-d}}{(2\pi)^{2d}} \int_{\mathbb{R}^d} \mathrm{d}y \int_{\mathbb{R}^d} \mathrm{d}\eta \int_{\mathbb{R}^d} \mathrm{d}z \int_{\mathbb{R}^d} \mathrm{d}\zeta \, \mathrm{e}^{+\mathrm{i}\eta \cdot (z - \frac{1}{2}(x+y))} \mathrm{e}^{-\frac{i}{\varepsilon}(y-x)\cdot\zeta} \, f(z,\zeta) \, \varphi(y)\end{aligned}$$

If we integrate over η, we get a δ that can be used to kill the integral with respect to z.

$$\begin{aligned}\big(\mathrm{Op}(f)\varphi\big)(x) &= \frac{\varepsilon^{-d}}{(2\pi)^{2d}} \int_{\mathbb{R}^d} \mathrm{d}y \int_{\mathbb{R}^d} \mathrm{d}z \int_{\mathbb{R}^d} \mathrm{d}\zeta \, (2\pi)^d \delta\big(z - \tfrac{1}{2}(x+y)\big) \, \mathrm{e}^{-\frac{i}{\varepsilon}(y-x)\cdot\zeta} \cdot \\ &\qquad \cdot f(z,\zeta) \, \varphi(y) \\ &= \frac{1}{(2\pi\varepsilon)^d} \int_{\mathbb{R}^d} \mathrm{d}y \int_{\mathbb{R}^d} \mathrm{d}\zeta \, \mathrm{e}^{-\frac{i}{\varepsilon}(y-x)\cdot\zeta} \, f\big(\tfrac{1}{2}(x+y), \zeta\big) \, \varphi(y)\end{aligned}$$

Identifying the integral with respect to ζ as partial Fourier transform, we get the second line of equation (6.1.1). A variable substitution $y' := \varepsilon y$ yields the first line. □

Example Although we cannot justify this example rigorously yet, we will only do a quick sanity check whether the quantization of $h(x,\xi) = \frac{1}{2m}\xi^2 + V(x)$ gives the expected result. By linearity, we can consider each of the terms in turn: we have already computed the Fourier transform of $\frac{1}{2m}\xi^2$ in the distributional sense in Chapter **??**, for all $\varphi \in \mathcal{S}(\mathbb{R}^d)$

$$\begin{aligned}\big(\mathrm{Op}\big(\tfrac{1}{2m}\xi^2\big)\varphi\big)(x) &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}y \, \frac{1}{2m} \, \big(\mathcal{F}(\varepsilon^2\eta^2)\big)(y-x) \, \varphi(y) \\ &= \frac{1}{2m} \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}y \, (2\pi)^{d/2} (+i)^2 \varepsilon^2 \Delta_y \delta(y-x) \, \varphi(y) \\ &= \tfrac{1}{2m} \big((+i)^2 \varepsilon^2 \Delta_x \delta_x, \varphi\big) = \tfrac{1}{2m}(-i)^2 \varepsilon^2 \Delta_x \varphi(x) = -\tfrac{\varepsilon^2}{2m} \Delta_x \varphi(x)\end{aligned}$$

holds. We see that this calculation hinges on φ being a test function. The other term can also be calculated using $\mathcal{F}\mathrm{e}^{+ix\cdot\eta} = (2\pi)^d \delta_x$ where $\delta_x(f) := f(x)$ is the shifted Dirac distribution,

$$\begin{aligned}\big(\mathrm{Op}(V)\varphi\big)(x) &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}y \, (\mathcal{F}\,1)(y-x) \, V\big(\tfrac{1}{2}(x+y)\big) \, \varphi(y) \\ &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}y \, (2\pi)^{d/2} \delta(y-x) \, V\big(\tfrac{1}{2}(x+y)\big) \, \varphi(y) = V(x) \, \varphi(x).\end{aligned}$$

Hence, the quantization of h yields

$$\mathsf{Op}(h) = \tfrac{1}{2m}(-i\varepsilon\nabla_x)^2 + V(\hat{x}) = \tfrac{1}{2m}P^2 + V(Q_\varepsilon),$$

exactly what we have expected. We see that it was crucial for this argument to work that $\varphi \in \mathcal{S}(\mathbb{R}^d)$ and we have to work to extend the integral formula to other $\varphi \in L^2(\mathbb{R}^d) \setminus \mathcal{S}(\mathbb{R}^d)$. If V is bounded, for instance, then $\mathsf{Op}(V) = V(Q_\varepsilon)$ defines a bounded operator and we can use Theorem 3.1.6 to extend it to all of $L^2(\mathbb{R}^d)$. P^2, however, is an unbounded operator and thus does not extend to a *bounded* operator on all of $L^2(\mathbb{R}^d)$

In Proposition 6.1.6, we have introduced the notion of operator kernel: the operator kernel is a function or distribution that tells us how the operator acts on wave functions. It can be useful in calculating things, e. g. if one wants to compute a trace.

Every bounded operator and many unbounded ones have a distributional operator kernel. In general, it is not a function, but only distribution on $\mathbb{R}^d \times \mathbb{R}^d$. The operator kernel of $-i\partial_{x_l}$ is $+i(2\pi)^{d/2}\partial_{x_l}\delta(x-y)$ since for all $\varphi \in \mathcal{S}(\mathbb{R}^d) \subset L^2(\mathbb{R}^d)$

$$\begin{aligned}(-i\partial_{x_l}\varphi)(x) &= \big((+i)\partial_{x_l}\delta_x, \varphi\big) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}y\, i(2\pi)^{d/2}\partial_{x_l}\delta(x-y)\,\varphi(y) \\ &= \Big(\mathfrak{Int}\big(+i(2\pi)^{d/2}\partial_{x_l}\delta(x-y)\big)\varphi\Big)(x)\end{aligned}$$

holds. Similarly, the operator kernel associated to $T = |\psi\rangle\langle\varphi| = \langle\varphi, \cdot\,\rangle\;\psi$, $\varphi, \psi \in \mathcal{S}(\mathbb{R}^d)$ is

$$\begin{aligned}(T\phi)(x) &= \langle\varphi, \phi\rangle\,\psi(x) = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}y\,(2\pi)^{d/2}\,\langle\varphi, \phi\rangle\,\delta(x-y)\,\psi(y) \\ &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}y \int_{\mathbb{R}^d} \mathrm{d}z\,(2\pi)^{d/2}\,\varphi^*(z)\,\phi(z)\,\delta(x-y)\,\psi(y) \\ &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}z \bigg((2\pi)^{d/2} \int_{\mathbb{R}^d} \mathrm{d}y\,\varphi^*(z)\,\delta(x-y)\,\psi(y)\bigg)\,\phi(z) \\ &=: \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}z\,K_T(x,z)\,\phi(z).\end{aligned}$$

Hence, even this well-behaved operator has a distributional operator kernel! 2009.12.09

To get back to the properties of quantization procedures: Weyl quantization is *linear*, i. e.

$$\mathsf{Op}(f + \alpha g) = \mathsf{Op}(f) + \alpha\mathsf{Op}(g)$$

holds for all $f, g \in \mathcal{S}(T^*\mathbb{R}^d)$ and $\alpha \in \mathbb{C}$. Furthermore, we can compute the quantization of the constant function 1 to be

$$\begin{aligned}\big(\mathrm{Op}(f)\varphi\big)(x) &= \frac{\varepsilon^{-d}}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}y \, (\mathcal{F}_2 \, 1)\big(\tfrac{1}{2}(x+y), \tfrac{y-x}{\varepsilon}\big) \, \varphi(y) \\ &= \frac{\varepsilon^{-d}}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}y \, \varepsilon^d \, (2\pi)^{d/2} \, \delta(y-x) \, \varphi(y) \\ &= \varphi(x) = \big(\mathrm{id}_{L^2}\varphi\big)(x).\end{aligned}$$

In the tutorials, we will show that Weyl quantization intertwines complex conjugation and taking adjoints,

$$\mathrm{Op}(f^*) = \mathrm{Op}(f)^*.$$

This fact has the important consequence that real-valued functions are potentially mapped onto selfadjoint operators.

Theorem 6.1.7 *Weyl quantization is linear, maps the constant function* 1 *to* id_{L^2} *and intertwines complex conjugation with taking adjoints in the sense of operators, i. e.* $\mathrm{Op}(f^*) = \mathrm{Op}(f)^*$ *holds for all* $f \in \mathcal{S}(T^*\mathbb{R}^d)$.

6.1.2.1 The Weyl product

The Weyl product $\sharp$ emulates the operator product on the level of functions on phase space, i. e. it satisfies

$$\mathrm{Op}\big(f \sharp g\big) = \mathrm{Op}(f) \, \mathrm{Op}(g)$$

for suitable $f, g : T^*\mathbb{R}^d \longrightarrow \mathbb{C}$. It can be derived from the composition law of the Weyl system (Proposition 6.1.4).

Theorem 6.1.8 *For* $f, g \in \mathcal{S}(T^*\mathbb{R}^d)$*, the distribution* $f \sharp g$ *which satisfies* $\mathrm{Op}\big(f \sharp g\big) = \mathrm{Op}(f) \, \mathrm{Op}(g)$ *is a Schwartz function given by*

$$\begin{aligned}(f \sharp g)(X) &= \frac{1}{(2\pi)^{2d}} \int_{T^*\mathbb{R}^d} \mathrm{d}Y \int_{T^*\mathbb{R}^d} \mathrm{d}Z \, \mathrm{e}^{+\mathrm{i}\sigma(X,Y+Z)} \, \mathrm{e}^{+\mathrm{i}\frac{\varepsilon}{2}\sigma(Y,Z)} \, (\mathcal{F}_\sigma f)(Y) \, (\mathcal{F}_\sigma g)(Z) \qquad (6.1.2) \\ &= \frac{1}{(\pi\varepsilon)^{2d}} \int_{T^*\mathbb{R}^d} \mathrm{d}Y \int_{T^*\mathbb{R}^d} \mathrm{d}Z \, \mathrm{e}^{-\mathrm{i}\frac{2}{\varepsilon}\sigma(X-Y,X-Z)} \, f(Y) \, g(Z).\end{aligned}$$

Before we can prove this statement, we need an auxiliary result: take two operators T and S whose operator kernels K_T and K_S are in $\mathcal{S}(\mathbb{R}^d \times \mathbb{R}^d)$. Then the operator kernel of TS is given by

$$(K_T \diamond K_S)(x,y) := \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}z \, K_T(x,z) \, K_S(z,y).$$

Lemma 6.1.9 *For any $K_T, K_S \in \mathcal{S}(\mathbb{R}^d \times \mathbb{R}^d)$, the product $K_T \diamond K_S$ is also in $\mathcal{S}(\mathbb{R}^d \times \mathbb{R}^d)$, i. e. $\diamond : \mathcal{S}(\mathbb{R}^d \times \mathbb{R}^d) \times \mathcal{S}(\mathbb{R}^d \times \mathbb{R}^d) \longrightarrow \mathcal{S}(\mathbb{R}^d \times \mathbb{R}^d)$.*

Proof We need to estimate the seminorms of $K_T \diamond K_S$: let $a, \alpha, b, \beta \in \mathbb{N}_0^d$ be multiindices and for simplicity define $\Phi : (x, y, z) \mapsto (2\pi)^{-d/2} \, K_T(x, z) \, K_S(z, y)$. Then $\Phi \in \mathcal{S}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ is a Schwartz function in all three variables. First, we need to show we can exchange differentiation with respect to x and y and integration with respect to z, i. e. that for fixed x and y

$$\begin{aligned} x^a y^b \partial_x^\alpha \partial_y^\beta (K_T \diamond K_S)(x, y) &= x^a y^b \partial_x^\alpha \partial_y^\beta \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}z \, K_T(x, z) \, K_S(z, y) \\ &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}z \, x^a y^b \partial_x^\alpha \partial_y^\beta \big(K_T(x, z) \, K_S(z, y) \big) = \int_{\mathbb{R}^d} \mathrm{d}z \, x^a y^b \partial_x^\alpha \partial_y^\beta \Phi(x, y, z) \end{aligned} \tag{6.1.3}$$

holds. We will do this by estimating $\big| x^a y^b \partial_x^\alpha \partial_y^\beta \Phi(x, y, z) \big|$ uniformly in x and y by an integrable function $G(z)$ and then invoking Dominated Convergence. For fixed x and y, we estimate the L^1 norm of $x^a y^b \partial_x^\alpha \partial_y^\beta \Phi(x, y, \cdot)$ from above by a finite number of seminorms of $\Phi(x, y, \cdot)$ with the help of Lemma **??**,

$$\begin{aligned} \int_{\mathbb{R}^d} \mathrm{d}z \, \big| x^a y^b \partial_x^\alpha \partial_y^\beta \Phi(x, y, z) \big| &= \big\| x^a y^b \partial_x^\alpha \partial_y^\beta \Phi(x, y, \cdot) \big\|_{L^1(\mathbb{R}^d)} \\ &\leq C_1 \sup_{z \in \mathbb{R}^d} \big| x^a y^b \partial_x^\alpha \partial_y^\beta \Phi(x, y, z) \big| + C_2 \max_{|c| = 2n} \sup_{z \in \mathbb{R}^d} \big| x^a y^b \partial_x^\alpha \partial_y^\beta \Phi(x, y, z) \big| \\ &= C_1 \big\| x^a y^b \partial_x^\alpha \partial_y^\beta \Phi(x, y, \cdot) \big\|_{00} + C_2 \max_{|c| = 2n} \big\| x^a y^b \partial_x^\alpha \partial_y^\beta \Phi(x, y, \cdot) \big\|_{c0} . \end{aligned}$$

Now we interchange sup and integration with respect to z,

$$\begin{aligned} \sup_{x, y \in \mathbb{R}^d} \int_{\mathbb{R}^d} \mathrm{d}z \, \big| x^a y^b \partial_x^\alpha \partial_y^\beta \Phi(x, y, z) \big| &\leq \int_{\mathbb{R}^d} \mathrm{d}z \sup_{x, y \in \mathbb{R}^d} \big| x^a y^b \partial_x^\alpha \partial_y^\beta \Phi(x, y, z) \big| \\ &= \Big\| \sup_{x, y \in \mathbb{R}^d} \big| x^a y^b \partial_x^\alpha \partial_y^\beta \Phi(x, y, \cdot) \big| \Big\|_{L^1(\mathbb{R}^d)} , \end{aligned}$$

which can be estimated from above by

$$\begin{aligned} \Big\| \sup_{x, y \in \mathbb{R}^d} \big| x^a y^b \partial_x^\alpha \partial_y^\beta \Phi(x, y, \cdot) \big| \Big\|_{L^1(\mathbb{R}^d)} &\leq C_1 \sup_{x, y \in \mathbb{R}^d} \sup_{z \in \mathbb{R}^d} \big| x^a y^b \partial_x^\alpha \partial_y^\beta \Phi(x, y, z) \big| + \\ &\quad + C_2 \max_{|c| = 2n} \sup_{x, y \in \mathbb{R}^d} \sup_{z \in \mathbb{R}^d} \big| x^a y^b \partial_x^\alpha \partial_y^\beta \Phi(x, y, z) \big| \\ &= C_1 \, \|\Phi\|_{a\alpha b\beta 00} + C_2 \max_{|c| = 2n} \|\Phi\|_{a\alpha b\beta c0} < \infty . \end{aligned}$$

Here $\left\{ \|\cdot\|_{a\alpha b\beta c\gamma} \right\}_{a,\alpha,b,\beta,c,\gamma \in \mathbb{N}_0^d}$ is the family of seminorms associated to $\mathcal{S}(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ which are defined by

$$\|\Phi\|_{a\alpha b\beta c\gamma} := \sup_{x,y,z \in \mathbb{R}^d} \left| x^a y^b z^c \partial_x^\alpha \partial_x^\beta \partial_x^\gamma \Phi(x,y,z) \right|.$$

This means we have found an *integrable* function

$$G(z) := \sup_{x,y \in \mathbb{R}^d} \left| x^a y^b \partial_x^\alpha \partial_y^\beta \Phi(x,y,z) \right|$$

which dominates $x^a y^b \partial_x^\alpha \partial_y^\beta \Phi(x,y,z)$ for all $x, y \in \mathbb{R}^d$. Hence, exchanging differentiation and integration in equation (6.1.3) is possible and we can bound the $\|\cdot\|_{a\alpha b\beta}$ seminorm on $\mathcal{S}(\mathbb{R}^d \times \mathbb{R}^d)$ by

$$\begin{aligned} \left\| K_T \diamond K_S \right\|_{a\alpha b\beta} &= \sup_{x,y \in \mathbb{R}^d} \left| x^a y^b \partial_x^\alpha \partial_y^\beta (K_T \diamond K_S)(x,y) \right| \\ &\leq C_1 \left\| \Phi \right\|_{a\alpha b\beta 00} + C_2 \max_{|c|=2n} \left\| \Phi \right\|_{a\alpha b\beta c0} < \infty. \end{aligned}$$

This means $K_T \diamond K_S \in \mathcal{S}(\mathbb{R}^d \times \mathbb{R}^d)$. □

Proof (Theorem 6.1.8) Using the definition of Op, we get

$$\begin{aligned} \mathrm{Op}(f)\,\mathrm{Op}(g) &= \frac{1}{(2\pi)^{2d}} \int_{T^*\mathbb{R}^d} \mathrm{d}Y \int_{T^*\mathbb{R}^d} \mathrm{d}Z \, (\mathcal{F}_\sigma f)(Y) \, (\mathcal{F}_\sigma g)(Z) \, W(Y) \, W(Z) \\ &= \frac{1}{(2\pi)^{2d}} \int_{T^*\mathbb{R}^d} \mathrm{d}Y \int_{T^*\mathbb{R}^d} \mathrm{d}Z \, (\mathcal{F}_\sigma f)(Y) \, (\mathcal{F}_\sigma g)(Z) \, \mathrm{e}^{+\mathrm{i}\frac{\varepsilon}{2}\sigma(Y,Z)} \, W(Y+Z) \\ &= \frac{1}{(2\pi)^{d}} \int_{T^*\mathbb{R}^d} \mathrm{d}Z \left(\frac{1}{(2\pi)^d} \int_{T^*\mathbb{R}^d} \mathrm{d}Y \, \mathrm{e}^{+\mathrm{i}\frac{\varepsilon}{2}\sigma(Y,Z-Y)} \, (\mathcal{F}_\sigma f)(Y) \, (\mathcal{F}_\sigma g)(Z-Y) \right) W(Z). \end{aligned}$$

We recognize the inner integral as $\big(\mathcal{F}_\sigma(f \sharp g)\big)(Z)$ and thus we add a Fourier transform to obtain the first of the two equivalent forms of the product formula:

$$\begin{aligned} (f \sharp g)(X) &= \mathcal{F}_\sigma \left(\frac{1}{(2\pi)^d} \int_{T^*\mathbb{R}^d} \mathrm{d}Y \, \mathrm{e}^{+\mathrm{i}\frac{\varepsilon}{2}\sigma(Y,\,\cdot\,-Y)} \, (\mathcal{F}_\sigma f)(Y) \, (\mathcal{F}_\sigma g)(\,\cdot\,-Y) \right)(X) \\ &= \frac{1}{(2\pi)^{2d}} \int_{T^*\mathbb{R}^d} \mathrm{d}Z \int_{T^*\mathbb{R}^d} \mathrm{d}Y \, \mathrm{e}^{+\mathrm{i}\sigma(X,Z)} \, \mathrm{e}^{+\mathrm{i}\frac{\varepsilon}{2}\sigma(Y,Z-Y)} \, (\mathcal{F}_\sigma f)(Y) \, (\mathcal{F}_\sigma g)(Z-Y) \\ &= \frac{1}{(2\pi)^{2d}} \int_{T^*\mathbb{R}^d} \mathrm{d}Y \int_{T^*\mathbb{R}^d} \mathrm{d}Z \, \mathrm{e}^{+\mathrm{i}\sigma(X,Y+Z)} \, \mathrm{e}^{+\mathrm{i}\frac{\varepsilon}{2}\sigma(Y,Z)} \, (\mathcal{F}_\sigma f)(Y) \, (\mathcal{F}_\sigma g)(Z). \end{aligned}$$

It remains to show that $f \sharp g$ is a Schwartz function: from Remark **??**, we know that the Weyl kernels K_f and K_g of f and g are in $\mathcal{S}(\mathbb{R}^d \times \mathbb{R}^d)$. Hence, we can write the operator product of $\mathrm{Op}(f)$ and $\mathrm{Op}(g)$ in terms of the associated integral kernels,

$$\begin{aligned}\big(\mathrm{Op}(f)\mathrm{Op}(g)\varphi\big)(x) &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}y \int_{\mathbb{R}^d} \mathrm{d}z \, \frac{1}{(2\pi)^{d/2}} K_f(x,z)\, K_g(z,y)\, \varphi(y) \\ &= \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} \mathrm{d}y \, (K_f \diamond K_g)(x,y)\, \varphi(y) \overset{!}{=} \mathrm{Op}\big(f \sharp g\big).\end{aligned}$$

By Lemma 6.1.9, $K_f \diamond K_g \in \mathcal{S}(\mathbb{R}^d \times \mathbb{R}^d)$. We use the inverse of Weyl quantization, Proposition **??**, to conclude

$$f \sharp g = \mathrm{Op}^{-1}\big(\mathrm{Op}(f)\mathrm{Op}(g)\big) = \varepsilon^d \, \mathcal{W}(K_f \diamond K_g) \in \mathcal{S}(T^* \mathbb{R}^d)$$

as $\mathcal{W}$ maps $\mathcal{S}(\mathbb{R}^d \times \mathbb{R}^d)$ bijectively onto $\mathcal{S}(T^*\mathbb{R}^d)$.

To show that the first form of the Weyl product, equation (6.1.2), is equivalent to the second, we have to write out all Fourier tranforms and collect the exponentials properly to obtain

$$\begin{aligned}(f \sharp g)(X) &= \frac{1}{(2\pi)^{4d}} \int_{T^*\mathbb{R}^d} \mathrm{d}Y \int_{T^*\mathbb{R}^d} \mathrm{d}Y' \int_{T^*\mathbb{R}^d} \mathrm{d}Z \int_{T^*\mathbb{R}^d} \mathrm{d}Z' \, \mathrm{e}^{+\mathrm{i}\sigma(X,Y+Z)} \mathrm{e}^{+\mathrm{i}\frac{\varepsilon}{2}\sigma(Y,Z)} \mathrm{e}^{+\mathrm{i}\sigma(Y,Y')} \mathrm{e}^{+\mathrm{i}\sigma(Z,Z')} \cdot \\ &\qquad \cdot f(Y')\, g(Z') \\ &= \frac{1}{(2\pi)^{4d}} \int_{T^*\mathbb{R}^d} \mathrm{d}Y \int_{T^*\mathbb{R}^d} \mathrm{d}Y' \int_{T^*\mathbb{R}^d} \mathrm{d}Z \int_{T^*\mathbb{R}^d} \mathrm{d}Z' \, \mathrm{e}^{+\mathrm{i}\sigma(Y,Y'-X)} \mathrm{e}^{+\mathrm{i}\sigma(Z,Z'-X-\frac{\varepsilon}{2}Y)} \, f(Y')\, g(Z') \\ &= \frac{(2\pi)^{2d}}{(2\pi)^{4d}} \int_{T^*\mathbb{R}^d} \mathrm{d}Y \int_{T^*\mathbb{R}^d} \mathrm{d}Y' \int_{T^*\mathbb{R}^d} \mathrm{d}Z' \, \mathrm{e}^{+\mathrm{i}\sigma(Y,Y'-X)} \, \delta\big(Z' - X - \tfrac{\varepsilon}{2}Y\big)\, f(Y')\, g(Z') \\ &= \frac{1}{(2\pi)^{2d}} \int_{T^*\mathbb{R}^d} \mathrm{d}Y \int_{T^*\mathbb{R}^d} \mathrm{d}Y' \, \mathrm{e}^{+\mathrm{i}\sigma(Y,Y'-X)} \, f(Y')\, g\big(X + \tfrac{\varepsilon}{2}Y\big).\end{aligned}$$

Making one last change of variables, $\tilde{Z} := X + \frac{\varepsilon}{2}Y$, we get the second form of the Weyl product,

$$\begin{aligned}(f \sharp g)(X) &= \frac{1}{(2\pi)^{2d}} \frac{2^{2d}}{\varepsilon^{2d}} \int_{T^*\mathbb{R}^d} \mathrm{d}\tilde{Z} \int_{T^*\mathbb{R}^d} \mathrm{d}Y' \, \mathrm{e}^{+\mathrm{i}\sigma(\frac{2}{\varepsilon}(\tilde{Z}-X),Y'-X)} \, f(Y')\, g(\tilde{Z}) \\ &= \frac{1}{(\pi\varepsilon)^{2d}} \int_{T^*\mathbb{R}^d} \mathrm{d}Y \int_{T^*\mathbb{R}^d} \mathrm{d}Z \, \mathrm{e}^{-\mathrm{i}\frac{2}{\varepsilon}\sigma(X-Y,X-Z)} \, f(Y)\, g(Z).\end{aligned}$$

This concludes the proof. □

Example Let us compute $x \sharp \xi$ which we then Weyl quantize using problem 23 (i). Since neither x nor ξ are Schwartz functions, we will dispense with mathematical rigor for now: in the distributional sense, we see from

$$(\xi_l \sharp x_l)(x,\xi) = \frac{1}{(2\pi)^{2d}} \int_{T^*\mathbb{R}^d} \mathrm{d}Y \int_{T^*\mathbb{R}^d} \mathrm{d}Z \, \mathrm{e}^{+\mathrm{i}\sigma(X,Y+Z)} \mathrm{e}^{+\mathrm{i}\frac{\varepsilon}{2}\sigma(Y,Z)} \, (\mathcal{F}_\sigma \eta_l')(Y)\, (\mathcal{F}_\sigma z_l')(Z)$$

that we need to compute the symplectic Fourier transforms of the factors. As distributions, their Fourier transforms exist and for the first, we get

$$\begin{aligned}(\mathcal{F}_\sigma \eta_l')(Y) &= \frac{1}{(2\pi)^d} \int_{T^*\mathbb{R}^d} \mathrm{d}Y' \, \mathrm{e}^{+\mathrm{i}\sigma(Y,Y')} \, \eta_l' = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathrm{d}y' \int_{\mathbb{R}^d} \mathrm{d}\eta' \, \eta_l' \mathrm{e}^{+\mathrm{i}(\eta \cdot y' - y \cdot \eta')} \\ &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathrm{d}y' \int_{\mathbb{R}^d} \mathrm{d}\eta' \, (+i) \partial_{y_l} \mathrm{e}^{+\mathrm{i}(\eta \cdot y' - y \cdot \eta')} = i \, (2\pi)^d \, \partial_{y_l} \delta(Y).\end{aligned}$$

Similarly, we calculate the second Fourier transform to be $(\mathcal{F}_\sigma z_l')(Z) = -i \, (2\pi)^d \, \partial_{\zeta_l} \delta(Z)$. Plugged into the product formula, we obtain

$$\begin{aligned}(\xi_l \sharp x_l)(x,\xi) &= \frac{1}{(2\pi)^{2d}} \int_{T^*\mathbb{R}^d} \mathrm{d}Y \int_{T^*\mathbb{R}^d} \mathrm{d}Z \, \mathrm{e}^{+\mathrm{i}\sigma(X,Y+Z)} \mathrm{e}^{+\mathrm{i}\frac{\varepsilon}{2}\sigma(Y,Z)} \, (-i^2) \, (2\pi)^{2d} \, \partial_{y_l} \delta(Y) \, \partial_{\zeta_l} \delta(Z) \\ &= - \int_{T^*\mathbb{R}^d} \mathrm{d}Y \int_{T^*\mathbb{R}^d} \mathrm{d}Z \, \partial_{y_l} \Bigl(\mathrm{e}^{+\mathrm{i}\sigma(X,Y+Z)} \mathrm{e}^{+\mathrm{i}\frac{\varepsilon}{2}\sigma(Y,Z)} \Bigr) \, \delta(Y) \, \partial_{\zeta_l} \delta(Z) \\ &= + \int_{T^*\mathbb{R}^d} \mathrm{d}Y \int_{T^*\mathbb{R}^d} \mathrm{d}Z \, \partial_{\zeta_l} \Bigl(\bigl(i\xi_l - i\tfrac{\varepsilon}{2}\zeta_l \bigr) \, \mathrm{e}^{+\mathrm{i}\sigma(X,Y+Z)} \mathrm{e}^{+\mathrm{i}\frac{\varepsilon}{2}\sigma(Y,Z)} \Bigr) \, \delta(Y) \, \delta(Z) \\ &= \int_{T^*\mathbb{R}^d} \mathrm{d}Y \int_{T^*\mathbb{R}^d} \mathrm{d}Z \, \Bigl((-i^2) \bigl(\xi_l - \tfrac{\varepsilon}{2}\zeta_l \bigr) \bigl(x_l + \tfrac{\varepsilon}{2} y_l \bigr) - i\tfrac{\varepsilon}{2} \Bigr) \, \mathrm{e}^{+\mathrm{i}\sigma(X,Y+Z)} \mathrm{e}^{+\mathrm{i}\frac{\varepsilon}{2}\sigma(Y,Z)} \cdot \\ &\qquad \cdot \delta(Y) \, \delta(Z) \\ &= \bigl(\xi_l - \tfrac{\varepsilon}{2} 0 \bigr) \bigl(x_l + \tfrac{\varepsilon}{2} 0 \bigr) - \varepsilon \tfrac{i}{2} = x_l \, \xi_l - \varepsilon \tfrac{i}{2}.\end{aligned}$$

By problem 23 (i) and $\mathrm{Op}(1) = \mathrm{id}_{L^2}$, we can compute the Weyl quantization of this function explicitly,

$$\begin{aligned}\mathrm{Op}\bigl(\xi \sharp x\bigr) &= \mathrm{Op}(x \cdot \xi) - \varepsilon \, d \tfrac{i}{2} \, \mathrm{id}_{L^2} \\ &= Q_\varepsilon \cdot P - \varepsilon \, d \tfrac{i}{2} \, \mathrm{id}_{L^2} - \varepsilon \, d \tfrac{i}{2} \, \mathrm{id}_{L^2} \\ &= Q_\varepsilon \cdot P - \varepsilon \, d i \, \mathrm{id}_{L^2} = P \cdot Q_\varepsilon.\end{aligned}$$

2009.12.22 This is exactly what we have expected.

The Weyl product has the following useful properties which can be easily proven:

Proposition 6.1.10 (Properties of Weyl product) *Let $f, g, h \in \mathcal{S}(T^*\mathbb{R}^d)$ and $\alpha \in \mathbb{C}$. Then the Weyl product has the following properties:*

(i) The Weyl product is bilinear, i. e. $\bigl(f + \alpha g\bigr) \sharp h = f \sharp h + \alpha \bigl(g \sharp h\bigr)$ and $f \sharp \bigl(g + \alpha h\bigr) = f \sharp g + \alpha \bigl(f \sharp h\bigr)$ hold.

(ii) The Weyl product is associative, i. e. $\bigl(f \sharp g\bigr) \sharp h = f \sharp \bigl(g \sharp h\bigr)$.

(iii) If $f \equiv f(x)$ and $g \equiv g(x)$ are functions of position, then the Weyl product reduces to the pointwise product, $f \sharp g = f \, g$.

(iv) If $f \equiv f(\xi)$ and $g \equiv g(\xi)$ are functions of momentum, then the Weyl product reduces to the pointwise product, $f \sharp g = f\, g$.

The proofs are left as an exercise.

6.2 Extending Weyl calculus by duality

6.3 Studying the Weyl product with oscillatory integral techniques

- $\sharp : S^{m_1}_{\rho,\delta} \times S^{m_2}_{\rho,\delta} \longrightarrow S^{m_1+m_2}_{\rho,\delta}$
- Asymptotic expansion

6.3.1 The Weyl product of two Hörmander symbols

6.3.2 Asymptotic expansion

The most important reason why *physicists* should care about Weyl calculus is that there is an *asymptotic expansion* of the product, i. e. we can write

$$f \sharp g \asymp \sum_{n=0}^{\infty} \varepsilon^n (f \sharp g)_{(n)}$$

where all the terms $(f \sharp g)_{(n)}$ are known explicitly. In general, this expansion does not converge to any function, but instead for any $N \in \mathbb{N}$, we can write

$$f \sharp g = \sum_{n=0}^{N} \varepsilon^n (f \sharp g)_{(n)} + \varepsilon^{N+1} R_N = \sum_{n=0}^{N} \varepsilon^n (f \sharp g)_{(n)} + \mathcal{O}(\varepsilon^{N+1}).$$

As $\varepsilon \to 0$, the remainder converges to 0 at least as fast as ε^{N+1}. Expansions that do not converge *are the rule* rather than the exception, e. g. when one sums up Feynman diagrams of different processes, they are often sorted by powers of some small coupling constant, $\alpha \simeq 1/137$, for instance. One has *no reason to believe* that summing over *all* terms, one gets something finite! This is where optimal truncation comes into play: instead of summing *all* terms, one stops at some $N(\varepsilon)$ where the error is minimal. Before we state the main theorem of this section, we will introduce the Landau symbols ("little and big O notation").

Definition 6.3.1 (Landau symbols) *Let $f, g : \mathbb{R}^d \longrightarrow \mathbb{C}$. We say $f(x) = \mathcal{O}\big(g(x)\big)$ as $x \to x_0$ if and only if*

$$\limsup_{x \to x_0} \frac{|f(x)|}{|g(x)|} < \infty$$

or quivalently if there exist $M \geq 0$ and $\delta > 0$ such that $|f(x)| \leq M\,|g(x)|$ for all $|x - x_0| < \delta$.

We say $f(x) = o\big(g(x)\big)$ as $x \to x_0$ if and only if

$$\lim_{x \to x_0} \frac{f(x)}{g(x)} = 0.$$

Example We have $x^2 = \mathcal{O}(x)$ and $x = \mathcal{O}(x)$ as $x \to 0$. However $x \neq o(x)$, but $x^2 = o(x)$ as $x \to 0$ since

$$\lim_{x \to 0} \frac{x^2}{x} = \lim_{x \to 0} x = 0.$$

Let us now formulate the main theorem of this section:

Theorem 6.3.2 *Let $f, g \in \mathcal{S}(T^*\mathbb{R}^d)$ and $\varepsilon < 1$. Then the Weyl product can be expanded asymptotically in ε, i. e. for any $N \in \mathbb{N}_0$ we have*

$$f \sharp g = \sum_{n=0}^{N} \varepsilon^n (f \sharp g)_{(n)} + \varepsilon^{N+1} R_N = \sum_{n=0}^{N} \varepsilon^n (f \sharp g)_{(n)} + \mathcal{O}(\varepsilon^{N+1}). \tag{6.3.1}$$

All the terms of the expansion are Schwartz functions, $(f \sharp g)_{(n)} \in \mathcal{S}(T^*\mathbb{R}^d)$ for all $n \in \mathbb{N}_0$, and are known explicitly,

$$(f \sharp g)_{(n)}(X) = \frac{1}{n!} \frac{i^n}{2^n} \big(\nabla_y \cdot \nabla_\zeta - \nabla_\eta \cdot \nabla_z\big)^n f(Y)\, g(Z) \Big|_{Y = X = Z}. \tag{6.3.2}$$

The remainder $R_N \in \mathcal{S}(T^*\mathbb{R}^d)$ as given by equation (6.3.4) *is also a rapidly decreasing function. The first two terms are given by*

$$f \sharp g = f\, g - \varepsilon \tfrac{i}{2} \big\{f, g\big\} + \mathcal{O}(\varepsilon^2) \tag{6.3.3}$$

where $\big\{f, g\big\} = \sum_{j=1}^d \big(\partial_{\xi_j} f\, \partial_{x_j} g - \partial_{x_j} f\, \partial_{\xi_j} g\big)$ is the Poisson bracket.

Proof Step 1: Expanding the twister. Let $N \in \mathbb{N}_0$ be arbitrary, but fixed. Then we can expand the twisting factor in they Weyl product,

$$e^{i \frac{\varepsilon}{2} \sigma(Y,Z)} = \sum_{n=0}^{N} \varepsilon^n \frac{1}{n!} \frac{i^n}{2^n} \big(\sigma(Y,Z)\big)^n + \tilde{R}_N(Y,Z),$$

where

$$\begin{aligned}\tilde{R}_N(Y,Z) &= \frac{1}{N!}\bigl(i\tfrac{\varepsilon}{2}\sigma(Y,Z)\bigr)^{N+1} \int_0^1 \mathrm{d}s\,(1-s)^N\,\mathrm{e}^{+\mathrm{i}s\frac{\varepsilon}{2}\sigma(Y,Z)} \\ &= \varepsilon^{N+1}\frac{1}{N!}\frac{i^{N+1}}{2^{N+1}}\bigl(\sigma(Y,Z)\bigr)^{N+1} \int_0^1 \mathrm{d}s\,(1-s)^N\,\mathrm{e}^{+\mathrm{i}s\frac{\varepsilon}{2}\sigma(Y,Z)}.\end{aligned}$$

is the remainder of the Taylor expansion. If we plug this into the product formula,

$$\begin{aligned}(f\sharp g)(X) &= \frac{1}{(2\pi)^{2d}}\int_{T^*\mathbb{R}^d}\mathrm{d}Y\int_{T^*\mathbb{R}^d}\mathrm{d}Z\,\mathrm{e}^{+\mathrm{i}\sigma(X,Y+Z)}\,\Bigl(\textstyle\sum_{n=0}^N\varepsilon^n\frac{1}{n!}\frac{i^n}{2^n}\bigl(\sigma(Y,Z)\bigr)^n + \tilde{R}_N(Y,Z)\Bigr)\cdot \\ &\qquad\qquad\cdot(\mathcal{F}_\sigma f)(Y)\,(\mathcal{F}_\sigma g)(Z) \\ &=: \sum_{n=0}^N \varepsilon^n (f\sharp g)_{(n)}(X) + \varepsilon^{N+1}R_N(X),\end{aligned} \tag{6.3.4}$$

we can define the nth order terms and the remainder.

Step 2: Treating $(f\sharp g)_{(n)}$. First of all, we note that

$$\bigl(\sigma(Y,Z)\bigr)^n = \bigl(\eta\cdot z - y\cdot\zeta\bigr)^n = \sum_{|a|+|b|=n}\frac{(-1)^{|b|}}{a!b!}\eta^a y^b z^a \zeta^b$$

is just a polynomial in y, η, z and ζ. Since $f\in\mathcal{S}(T^*\mathbb{R}^d)$, by Theorem **??**

$$y^b\eta^a(\mathcal{F}_\sigma f) = \mathcal{F}_\sigma\bigl((-i\partial_{\eta'})^b(+i\partial_{y'})^a f\bigr)\in\mathcal{S}(T^*\mathbb{R}^d)$$

holds for all multiindices $a,b\in\mathbb{N}_0^d$ and similarly for g. This means, the integral expression for $(f\sharp g)_{(n)}$ exists for all $X\in T^*\mathbb{R}^d$ and reduces to

$$\begin{aligned}(f\sharp g)_{(n)}(X) &= \frac{1}{n!}\frac{i^n}{2^n}\sum_{|a|+|b|=n}\frac{(-1)^{|b|}}{a!b!}\frac{1}{(2\pi)^{2d}}\int_{T^*\mathbb{R}^d}\mathrm{d}Y\int_{T^*\mathbb{R}^d}\mathrm{d}Z\,\mathrm{e}^{+\mathrm{i}\sigma(X,Y+Z)}\cdot \\ &\qquad\qquad\cdot\eta^a y^b(\mathcal{F}_\sigma f)(Y)\,z^a\zeta^b(\mathcal{F}_\sigma g)(Z) \\ &= \frac{1}{n!}\frac{i^n}{2^n}\sum_{|a|+|b|=n}\frac{(-1)^{|b|}}{a!b!}\frac{1}{(2\pi)^{2d}}\int_{T^*\mathbb{R}^d}\mathrm{d}Y\int_{T^*\mathbb{R}^d}\mathrm{d}Z\,\mathrm{e}^{+\mathrm{i}\sigma(X,Y+Z)}\cdot \\ &\qquad\cdot\Bigl(\mathcal{F}_\sigma\bigl((-i\partial_{\eta'})^b(+i\partial_{y'})^a f\bigr)\Bigr)(Y)\Bigl(\mathcal{F}_\sigma\bigl((-i\partial_{\zeta'})^a(+i\partial_{z'})^b g\bigr)\Bigr)(Z).\end{aligned}$$

The remaining integral is nothing but the symplectic Fourier transform in Y and Z.

The symplectic Fourier transform is its own inverse, hence we get expression (6.3.2),

$$\begin{aligned}
\ldots &= \frac{1}{n!} \frac{i^n}{2^n} \sum_{|a|+|b|=n} \frac{(-1)^{|b|}}{a!b!} \Big(\mathcal{F}_\sigma^2 \big((-i\partial_{\eta'})^b (+i\partial_{y'})^a f \big) \Big)(X) \, \Big(\mathcal{F}_\sigma^2 \big((-i\partial_{\zeta'})^a (+i\partial_{z'})^b g \big) \Big)(X) \\
&= \frac{1}{n!} \frac{i^n}{2^n} \sum_{|a|+|b|=n} \frac{(-1)^{|b|}}{a!b!} \big((-i\partial_\xi)^b (+i\partial_x)^a f \big)(X) \, \big((-i\partial_\xi)^a (+i\partial_x)^b g \big)(X) \\
&= \frac{1}{n!} \frac{i^n}{2^n} \big(\nabla_y \cdot \nabla_\zeta - \nabla_\eta \cdot \nabla_z \big)^n f(Y) \, g(Z) \Big|_{Y=X=Z} .
\end{aligned}$$

Since each of the factors consists of derivatives of Schwartz functions, $(f \sharp g)_{(n)} \in \mathcal{S}(T^*\mathbb{R}^d)$ is also a Schwartz function.

Step 3: Treating the remainder R_N. If we combine Theorem 6.1.8, $f \sharp g \in \mathcal{S}(T^*\mathbb{R}^d)$ with Step 2, $(f \sharp g)_{(n)} \in \mathcal{S}(T^*\mathbb{R}^d)$, we conclude that the remainder is also a Schwartz function,

$$R_N = f \sharp g - \sum_{n=0}^{N} \varepsilon^n (f \sharp g)_{(n)} \in \mathcal{S}(T^*\mathbb{R}^d).$$

From the explicit expression for the remainder, equation (6.3.4), we also see that it is indeed of order $\mathcal{O}(\varepsilon^{N+1})$. This concludes the proof. □

Example In the previous section, we have calculated $\xi \sharp x$. If we use the asymptotic expansion – which in this case is exact, we can obtain this result with much less work. Pluggin in equation (6.3.3), we get

$$\begin{aligned}
\big(\xi \sharp x \big)(x, \xi) &= \xi \cdot x - \varepsilon \, \frac{i}{2} \sum_{l=1}^{d} \big\{ \xi_l , x_l \big\} + \mathcal{O}(\varepsilon^2) \\
&= x \cdot \xi - \varepsilon \, \frac{i}{2} \sum_{l,j=1}^{d} \big(\partial_{\xi_j} \xi_l \, \partial_{x_j} x_l - \partial_{x_j} \xi_l \, \partial_{\xi_j} x_l \big) + \mathcal{O}(\varepsilon^2) \\
&= x \cdot \xi - \varepsilon \, \frac{i}{2} \sum_{l,j=1}^{d} \delta_{lj} + \mathcal{O}(\varepsilon^2) = x \cdot \xi - \varepsilon \, d \tfrac{i}{2} + \mathcal{O}(\varepsilon^2).
\end{aligned}$$

The remainder, however, is exactly 0: the factor $\big(\sigma(Y, Z) \big)^2$ is a polynomial of second order in $y \cdot \zeta$ and $\eta \cdot z$ and leads to *two* derivatives with respect to position and momentum. But x and ξ are linear so their second and higher-order derivatives vanish identically. Hence, we have shown

$$\xi \sharp x = \xi \cdot x - \varepsilon \, d \tfrac{i}{2}.$$

Generally, if one of the factors is a polynomial of kth degree, the asymptotic expansion terminates after finitely many terms and is exact.

In Chapter **??**, we have emphasized the importance that the the Moyal commutator

$$\big[f, g\big]_\sharp := f \sharp g - g \sharp f$$

vanishes as $\varepsilon \to 0$. Using equation (6.3.3), we immediately get

$$\big[f, g\big]_\sharp = -i\varepsilon \big\{f, g\big\} + \mathcal{O}(\varepsilon^3).$$

The error is of *third* order as all contributions from even powers vanish. This can be traced back to

$$\big(\sigma(Y, Z)\big)^{2k} = \big(-\sigma(Z, Y)\big)^{2k} = \big(\sigma(Z, Y)\big)^{2k} \qquad \qquad \forall k \in \mathbb{N}_0$$

and the commutativity of multiplication in $\mathbb{C}$. We mention the latter explicitly as we will quantize matrix-valued functions in the next section. There, it is not at all clear that even the 0th order vanishes.

- 2 lectures max
- Keep it short and sweet

Chapter 7
The Stationary Phase Method: Oscillatory Integrals of Another Kind

- Asymptotic formulas for oscillatory integrals $\rightsquigarrow$ evaluation of oscillatory, high-dimensional integrals (which are way too difficult to do numerically)
- Reference: non-linear Maxwell equations (Babin & Figotin?!)
- Stein, Chapter VIII
- Taylor expand phase $\rightsquigarrow$ which order is the first non-degenerate order?
- Outline applications

There is another type of integral that is commonly referred to as an oscillatory integral, namely those of the form

$$I(\varepsilon) := \int \mathrm{d}x \, \mathrm{e}^{\frac{\mathrm{i}}{\varepsilon} \vartheta(x)} \, \psi(x) \simeq \sum_k \varepsilon^{n(k)} \, a_k$$

where $\varepsilon \ll 1$ is small, but non-zero. Unlike the oscillatory integrals we have gotten to know in Chapter 5, the function ψ can be integrable. Although it is certainly possible to combine the ideas from Chapter 5 with those developed now.

Add example from non-linear condensed matter physics/non-linear electromatism

7.1 Extracting the essential features from a simple one-dimensional example

To avoid having to deal with unnecessary technical complications, suppose we are in one dimension and want to integrate over some compact interval $[a, b]$. Both, the phase $\vartheta \in \mathcal{C}^\infty([a, b])$ and the function $\psi \in \mathcal{C}^\infty([a, b])$ are smooth, which automatically makes the latter and all of its derivatives integrable. Moreover, we suppose that the support

$$\operatorname{supp} \psi := \big\{ x \in [a, b] \;\big|\; \psi(x) \neq 0 \big\} \subsetneq (a, b)$$

is a proper subset of (a, b), which frees us from having to deal with boundary terms.

As in the case of oscillatory integrals from Chapter 5, the idea is that we will only get appreciable contributions to the integral in regions where $\vartheta(x) \approx 0$, everything else will average out to 0. More precisely, we in addition need to require that ϑ has no critical points. Indeed, this is correct in our simple setting.

Proposition 7.1.1 *Let $\vartheta, \psi \in \mathcal{C}^\infty([a, b])$ be smooth functions on a compact interval, and assume that* $\operatorname{supp} \psi \subsetneq (a, b)$ *is contained in the interior of the interval. Then if $\vartheta'(x) \neq 0$ for all $x \in [a, b]$, the oscillatory integral*

$$I(\varepsilon) = \int_a^b \mathrm{d}x \, \mathrm{e}^{\frac{i}{\varepsilon} \vartheta(x)} \, \psi(x) = \mathcal{O}(\varepsilon^\infty)$$

vanishes to infinite order. That is, for each $k \in \mathbb{N}$ we have $I(\varepsilon) = \mathcal{O}(\varepsilon^k)$.

Proof The proof uses the L operator

$$L = -\varepsilon \, \frac{\mathrm{i}}{\vartheta'(x)} \, \partial_x$$

for the phase $\mathrm{e}^{+\frac{\mathrm{i}}{\varepsilon} \vartheta(x)} = L \mathrm{e}^{+\frac{\mathrm{i}}{\varepsilon} \vartheta(x)}$. Clearly, the existence of the L operator hinges on our assumption that the phase $\vartheta'(x) \neq 0$ vanishes nowhere in the interval.

Given that $\operatorname{supp} \psi$ is contained in the interior of the interval, ψ and all its derivatives vanishe at the boundary. Therefore, the boundary terms after repeated partial integration are all zero, and we conclude that the integral is $\mathcal{O}(\varepsilon^k)$,

$$\begin{aligned} I(\varepsilon) &= \int_a^b \mathrm{d}x \, L^k \mathrm{e}^{\frac{i}{\varepsilon} \vartheta(x)} \, \psi(x) \\ &= \int_a^b \mathrm{d}x \, \mathrm{e}^{\frac{i}{\varepsilon} \vartheta(x)} \, (L')^k \psi(x) \\ &= \mathcal{O}(\varepsilon^k). \end{aligned}$$

That is because each application of the transpose operator

$$(L'\psi)(x) = \varepsilon \, \frac{\partial}{\partial x} \left(\frac{\mathrm{i} \, \psi(x)}{\vartheta'(x)} \right) = \mathcal{O}(\varepsilon)$$

adds a factor of ε. Note that by assumption $\vartheta'(x) \neq 0$ is bounded away from 0 and therefore $\psi/\vartheta' \in \mathcal{C}^\infty([a,b])$ is again a smooth function.

As $k \in \mathbb{N}$ can be chosen arbitrarily large, this proves $I(\varepsilon) = \mathcal{O}(\varepsilon^\infty)$. □

This proof illustrates that even though this notion of oscillatory integral is different from that in Chapter 5, some of the methods are very similar.

Let us proceed in our discussion. Suppose ϑ' has a single isolated point $x_0 \in (a,b)$ where $\vartheta'(x_0) = 0$; this is a critical or *stationary point* — hence the name, stationary phase method. Then we can find a smooth resolution of the identity

$$1 = \chi(x) + \bigl(1 - \chi(x)\bigr),$$

which we will use to isolate the contribution by the stationary point. More precisely, the function $\chi : [a,b] \longrightarrow [0,1]$ is smooth, its support is a neighborhood of x_0 and $\chi(x) = 1$ in some smaller neighborhood of x_0, say, a small open interval $(x_0 - \mu, x_0 + \mu)$ around x_0.

Inserting the resolution of the identity into the integral gives us two integrals,

$$\begin{aligned} I(\varepsilon) &= \int_a^b \mathrm{d}x \, \mathrm{e}^{+\mathrm{i}\vartheta(x)} \, \chi(x) \, \psi(x) + \int_a^b \mathrm{d}x \, \mathrm{e}^{+\mathrm{i}\vartheta(x)} \, \bigl(1 - \chi(x)\bigr) \, \psi(x) \\ &= \int_a^b \mathrm{d}x \, \mathrm{e}^{+\mathrm{i}\vartheta(x)} \, \chi(x) \, \psi(x) + \mathcal{O}(\varepsilon^\infty) \end{aligned}$$

one of which satisfies the assumptions of Proposition 7.1.1. Consequently, the integral away from x_0 can be neglected and only the region near the critical point matters.

Clearly, we can generalize this argument to deal with finitely many stationary points.

Note that $\chi(x) = 1$ in a neighborhood of the stationary point, so derivatives $\partial_x^n \chi(x) = 0$ on $(x_0 - \mu, x_0 + \mu)$ all vanish and only the behavior of the function ψ will matter — up to $\mathcal{O}(\varepsilon^\infty)$, of course. And the question becomes: what is the contribution by the stationary point x_0. Seeing as the phase ϑ is a smooth function, we may Taylor expand it around the point x_0 to any order n,

$$\vartheta(x) = \sum_{k=0}^{n} \frac{1}{n!} \partial_x^k \vartheta(x_0) \, (x - x_0)^k + R_n(x).$$

The stationarity assumption $\vartheta(x_0) = 0$ means that the first term has to vanish. Moreover, Proposition 7.1.1 shows us that as long as $\vartheta'(x_0) \neq 0$ (along with it being non-zero everywhere else), the oscillatory integral vanishes to any order. In the context of the example of the introduction, the first condition is the frequency matching condition, the second one the group velocity matching condition.

So the first potentially non-zero term is quadratic,

$$\vartheta(x) = \tfrac{1}{2}\partial_x^2\vartheta(x_0)\,(x-x_0)^2 + \mathcal{O}\big((x-x_0)^3\big),$$

although it could be even higher-order.

The intuition tells us that the higher the order of the first non-zero term, the stronger the contribution will be. That is because $\vartheta(x) \approx 0$ will remain small for longer near x_0. Indeed, this is correct.

Let us assume that ϑ only vanishes to *finite* order at x_0, i. e. there exists k so that $\partial_x^k\vartheta(x_0) \neq 0$.

Chapter 8

Syllabus

Outline

(1) Introduction

- PDEs $\rightsquigarrow$ differential operators (e. g. $H = -\Delta_x + V(x)$)
- Generalization: simple case $\sqrt{m^2 - \Delta_z}$ via Fourier transform
- More general operators: pseudodifferential operators
- Give formula
- Question: how to interpret this formula? $\rightsquigarrow \int$ not absolutely convergent
- Notion of oscillatory integral: fast oscillation of integrand $\Rightarrow \int \approx 0$
- Give strategy for simple example (phase = $\mathrm{e}^{-\mathrm{i}\xi \cdot x}$): L operators
  $\Longrightarrow$ Make sense of $\mathcal{F}f$ when $f \in \mathcal{C}^\infty_{\mathrm{u,pol}}(\mathbb{R}^d)$, but not in $L^1(\mathbb{R}^d)$
- Extract math problems: $\mathcal{S}$, $\mathcal{S}'$, L^p-spaces, symbol spaces

(2) Ordinary integrals

- contrast and compare with oscillatory integrals
- $L^1(\mathbb{R}^d)$
- $L^p(\mathbb{R}^d) \rightsquigarrow$ Banach spaces
- $L^2(\mathbb{R}^d) \rightsquigarrow$ Hilbert spaces
- Dominated and Monotone Convergence
- Fourier transform
- Riemann-Lebesgue Lemma

- Be quick!

(3) Oscillatory integrals
- Schwartz functions
- Tempered distributions
- Rigorous definition of oscillatory integrals
- Frechét topology

(4) Primer on operator theory
- Bounded operators
- Continuous operators (in case we consider Frechét topologies)
- Zoology of bounded operators
- Unbounded operators
- Be quick!

(5) Pseudodifferential operators
- Definition of Op
- Extension to Hörmander classes
- Follow Mantoiu & Purice
- Define product
- Define Wigner transform

Bibliography

[Bui88] J. Buitrago. *A comment on the deduction of the Dirac equation*. European Journal of Physics **9**. 149–150, 1988.

[Cor83] H. O. Cordes. *A pseudodifferential Foldy-Wouthuysen transform*. Communications in Partial Differential Equations **8**. 1475–1485, 1983.

[Cor04] H. O. Cordes. *A precise pseudodifferential Foldy-Wouthuysen transform for the Dirac equation*. Journal of Evolution Equations **4**. 128–138, 2004.

[FW50] L. L. Foldy and S. A. Wouthuysen. *On the Dirac Theory of Spin 1/2 Particles and Its Non-Relativistic Limit*. Phys. Rev. **78**. 29–36, 1950.

[Fol89] G. B. Folland. *Harmonic Analysis on Phase Space*. Vol. 122. Annals of Mathematics Studies. Princeton University Press, 1989.

[Fol09] G. B. Folland. *Fourier Analysis and Its Applications*. Vol. 4. Pure and Applied Undergraduate Texts. American Mathematical Society, 2009.

[FL13] M. Fürst and M. Lein. *Semi- and Non-relativistic Limit of the Dirac Dynamics with External Fields*. Annales Henri Poincaré **14**. 1305–1347, 2013.

[Hör71] L. Hörmander. *Fourier Integral Operators I*. Acta Mathematica **127**. 79–183, 1971.

[Hör79] L. Hörmander. *The Weyl Calculus of Pseudo-Differential Operators*. Communications on Pure and Applied Mathematics **XXXII**. 359–443, 1979.

[Lei20] M. Lein. *A Mathematical Journey Through Differential Equations of Physics*. World Scientific, 2020.

[Ler96] L. Lerner. *Derivation of the Dirac equation from a relativistic representation of spin*. European Journal of Physics **17**. 172–175, 1996.

[LL01] E. Lieb and M. Loss. *Analysis*. American Mathematical Society, 2001.

[RS72] M. Reed and B. Simon. *Methods of Mathematical Physics I: Functional Analysis*. Vol. 1. Methods of Modern Mathematical Physics. Academic Press, 1972.

[Rob87] D. Robert. *Autour de l'Approximation Semi-Classique*. Birkhäuser, 1987.

[SS03] E. M. Stein and R. Shakarchi. *Fourier Analysis: An Introduction*. Vol. 1. Princeton Lectures in Analysis. Princeton University Press, 2003.

[Tes19] G. Teschl. *Topics in Real and Functional Analysis*. to appear in Graduate Studies in Mathematics. American Mathematical Society, 2019.

[Teu03] S. Teufel. *Adiabatic Perturbation Theory in Quantum Dynamics*. Vol. 1821. Lecture Notes in Mathematics. Springer-Verlag, 2003.

[Tha92] B. Thaller. *The Dirac Equation*. Springer-Verlag, 1992.

[Wig39] E. P. Wigner. *On Unitary Representations of the Inhomogeneous Lorentz Group*. Annals of Mathematics **40**. 149–204, 1939.

[Ynd96] F. Yndurain. *Relativistic Quantum Mechanics and Introduction to Field Theory*. Springer-Verlag, 1996.