Topology and Fourier Transform

Michael Dreher
Institute of Mathematics
University of Rostock

Spring 2019 — Almost Completed !

Contents

Chapter 1

Topology

1.1 (Set–Theoretic) Topology in General

This part follows [2], [3], [4] and [6].

1.1.1 What is Topology All About ?

Consider some sets X and Y, and they could be $\Omega \subset \mathbb{R}^3$, or $L^2(\Omega)$, or $C_0^\infty(\Omega)$. We wish to understand

- what are continuous maps from X to Y,
- what are open, closed, compact subsets of X,

and the list of questions does not stop here. We would like to define these concepts without measuring distances in X or Y (because there is no metric in $C_0^\infty(\Omega)$ that describes what we want from that space). Recall that in $\mathbb{R}^n$ all these properties can be defined by means of norms, but in topological spaces this is no longer possible. Instead, neighbourhoods and open sets will become the key concepts from which everything else will be defined. One price we have to pay is that we should no longer work with converging sequences because they are not the most relevant idea. As an example: we will see that every continuous map from a topological space X to a topological space Y will map every converging sequence to a converging sequence (and the limit is mapped to the limit), but the converse does *not* hold. This differs from the metric situation. On the other hand, we will learn that we can achieve a lot without ever using converging sequences.

One more remark is in order. When we speak about some *space*, we mean a set with a certain structure. Examples are

linear spaces (aka vector spaces): these are sets with a linear structure that comes from the two vector space operations (vector plus vector gives vector; number times vector gives vector),

metric spaces: these are sets with some metric (maybe a certain metric space is also a linear space, maybe it is not)

normed spaces: these are linear spaces that have a norm structure on top, and both structures are compatible. Each normed space can be interpreted as a metric space (the converse is wrong).

topological spaces: these are sets with some topological structure (maybe a certain topological space is also a linear space, maybe it is not). Each metric space can be interpreted as a topological space (the converse is wrong).

topological vector spaces: these are vector spaces which also have a topological structure on top, and both structures are compatible.

1.1.2 Basic Concepts: Neighbourhoods Bases and Open Sets

Definition 1.1 (**Neighbourhood basis**). *Let X be a non-empty set. We assume that to each $a \in X$, there belongs some subset $\mathcal{B}(a)$ of the power set $\mathcal{P}(X)$ with the following properties:*

$$\forall a \in X\colon \quad \mathcal{B}(a) \neq \emptyset; \quad \text{and } \forall B \in \mathcal{B}(a)\colon \quad a \in B \tag{B1}$$

$$\forall a \in X\colon \quad \forall B_1, B_2 \in \mathcal{B}(a)\colon \quad \exists B_3 \in \mathcal{B}(a)\colon \quad B_3 \subset B_1 \cap B_2 \tag{B2}$$

$$\forall a \in X\colon \quad \forall B \in \mathcal{B}(a)\colon \quad \forall b \in B\colon \quad \exists C \in \mathcal{B}(b)\colon \quad b \in C \subset B \tag{B3}$$

Each such $B \in \mathcal{B}(a)$ is called basic neighbourhood of a.

Example 1.2. *Take $X = \mathbb{R}^n$, equipped with the usual Euclidean norm $|x| = (x_1^2 + \ldots + x_n^2)^{1/2}$, and then put $\mathcal{B}(a) = \{U_\varepsilon(a)\colon 0 < \varepsilon < 1\}$, with $U_\varepsilon(a) = \{x \in \mathbb{R}^n\colon |x - a| < \varepsilon\}$ being the usual open balls of radius ε centred at $a \in \mathbb{R}^n$. Visualise yourself what (B1), (B2), (B3) mean in this example.*

Definition 1.3 (**Open sets, Topology, Topological space**). *We assume the situation of the previous definition. A set $\Omega \subset X$ is called* open set *provided that*

$$\forall a \in \Omega\colon \quad \exists B \in \mathcal{B}(a)\colon \quad a \in B \subset \Omega.$$

A set $\tau \subset \mathcal{P}(X)$ is called topology of X *if it comprises all the open sets in X. Then (X, τ) is called* topological space.

Lemma 1.4. *Every basic neighbourhood $B \in \cup_{x \in X} \mathcal{B}(x)$ in a topological space (X, τ) is an open set.*

Proof. Let $B \in \mathcal{B}(x)$ for some $x \in X$. Take any $b \in B$. Due to (B3), some $C \in \mathcal{B}(b)$ exists with $b \in C \subset B$. This means that B is open. □

Proposition 1.5 (**Properties of open sets**). *Let (X, τ) be some topological space. Then we have:*

$$\emptyset \quad \text{and} \quad X \quad \text{are open sets} \tag{O1}$$

$$\text{If} \quad \Omega_1 \quad \text{and} \quad \Omega_2 \quad \text{are open sets, then also} \quad \Omega_1 \cap \Omega_2 \tag{O2}$$

$$\text{If} \quad \Omega_i \quad \text{are open sets, with } i \text{ from some non-empty index set } I\text{, then also} \quad \bigcup_{i \in I} \Omega_i \tag{O3}$$

Proof. The principle *ex falso quodlibet* implies $\emptyset \in \tau$, and $X \in \tau$ is obvious as well, which gives (O1). Let Ω_1 and Ω_2 be open, and $a \in \Omega_1 \cap \Omega_2$. Since Ω_j are open, there are $B_j \in \mathcal{B}(a)$ with $a \in B_j \subset \Omega_j$, for $j = 1, 2$. Owing to (B2), there is some $B_3 \in \mathcal{B}(a)$ with $a \in B_3 \subset B_1 \cap B_2 \subset \Omega_1 \cap \Omega_2$, and therefore also $\Omega_1 \cap \Omega_2$ is open. And to show (O3), let us be given open sets Ω_i for i from some (possibly uncountable) index set I. Assume $a \in \cup_{i \in I} \Omega_i$. Then $a \in \Omega_{i_0}$ for some $i_0 \in I$, and Ω_{i_0} is open, hence some $B \in \mathcal{B}(a)$ exists with $a \in B \subset \Omega_{i_0} \subset \cup_{i \in I} \Omega_i$, proving (O3). □

Definition 1.6 (**Neighbourhoods and Neighbourhood filters**). *Let (X, τ) be a topological space. We say that some set $U \subset X$ is a* neighbourhood of $a \in X$ *if some $B \in \mathcal{B}(a)$ exists with $a \in B \subset U$. All the neighbourhoods U of $a \in X$ form the neighbourhood filter $\mathcal{N}(a)$:*

$$\mathcal{N}(a) := \{U \subset X\colon \quad U \text{ is a neighbourhood of } a\}.$$

Contrary to the expectation, a neighbourhood need not be open.

Remark 1.7. *In other books the following approach can be found: $\tau \subset \mathcal{P}(X)$ is called a topology on X if its elements Ω satisfy (O1), (O2), and (O3), and those Ω are then called opens sets in X. Subsequently, $U \subset X$ will be called a neighbourhood of $a \in X$ if some open Ω exists with $a \in \Omega \subset U$. Then the basic neighbourhoods of a point are defined as the* open *neighbourhoods of that point. Then one has to prove that these basic neighbourhoods satisfy (B1), (B2), (B3). However, both approaches are equivalent and will lead to the same theory and same results. We have chosen the approach here with the hope that we will arrive faster at the applications.*

Definition 1.8 (**Interior of a set**). *Let A be a set in a topological space (X, τ). We say that a point $a \in A$ is an* interior point of A *(also called* inner point*) if some basic neighbourhood $B \in \mathcal{B}(a)$ exists with $a \in B \subset A$. All interior points of A form the* interior of A,

$$\text{interior}(A) := \{a \in A\colon a \quad \text{is inner point of} \quad A\}.$$

Proposition 1.9. *For every set A in a topological space,* $\mathrm{interior}(A)$ *is an open set.*

Proof. Pick $a \in \mathrm{interior}(A)$. Definition 1.8 says that then some $B \in \mathcal{B}(a)$ exists with $a \in B \subset A$. From (B3) we learn that to each $b \in B$, there is some $C \in \mathcal{B}(b)$ with $b \in C \subset B \subset A$. This means that each $b \in B$ is an interior point of A, hence $B \subset \mathrm{interior}(A)$.

Expressed differently: for each $a \in \mathrm{interior}(A)$, there is some $B \in \mathcal{B}(a)$ with $a \in B \subset \mathrm{interior}(A)$. That is the definition of $\mathrm{interior}(A)$ being open. □

Remark 1.10. *We mention without proof:* $\mathrm{interior}(A)$ *is the union of all open subsets of A.*

1.1.3 Examples of Neighbourhood Bases and Topological Spaces

The Banach space $\mathbb{R}^n$: consider the vector space $\mathbb{R}^n$ with open sets in the usual sense, and pick $a \in \mathbb{R}^n$. Then we define a neighbourhood base

$$\mathcal{B}(a) = \Big\{ \{x \in \mathbb{R}^n : \ |x-a| < \varepsilon\} : \ 0 < \varepsilon < 1 \Big\}.$$

Any metric space: Let (X, d) be a metric space (not necessarily complete), and put (for each $a \in X$)

$$\mathcal{B}(a) = \left\{ \left\{ x \in X : \ d(a,x) < \frac{1}{2^k} \right\} : \ k \in \mathbb{N} \right\}.$$

By Definition 1.3, this choice of $\mathcal{B}$ completely determines a topology τ in X. We call it the *metric topology*. This is the standard method of interpreting a metric space as a topological space.

The extended real line: Put $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty, -\infty\}$ and define neighbourhood bases as follows:

$$\begin{aligned} \mathcal{B}(a) &:= \Big\{ (a-\varepsilon, a+\varepsilon) : \ 0 < \varepsilon < 1 \Big\}, \qquad a \in \mathbb{R}, \\ \mathcal{B}(+\infty) &:= \Big\{ (k, \infty) \cup \{+\infty\} : \ k \in \mathbb{N} \Big\}, \\ \mathcal{B}(-\infty) &:= \Big\{ (-\infty, -k) \cup \{-\infty\} : \ k \in \mathbb{N} \Big\}. \end{aligned}$$

The extended natural numbers: Put $\overline{\mathbb{N}} := \mathbb{N}_0 \cup \{+\infty\}$ and define neighbourhood bases like this:

$$\begin{aligned} \mathcal{B}(a) &:= \Big\{ \{a\} \Big\}, \qquad a \in \mathbb{N}_0, \\ \mathcal{B}(+\infty) &:= \Big\{ \{n \in \mathbb{N} : n \ge k\} \cup \{+\infty\} : \ k \in \mathbb{N} \Big\}. \end{aligned}$$

The vector space of all sequences in $\mathbb{R}$: Consider the vector space

$$X = \Big\{ (x_1, x_2, x_3, \dots) : \ x_j \in \mathbb{R} \ \ \forall j \Big\}$$

and define a metric

$$d_X(x,y) := \sum_{k=1}^{\infty} \frac{1}{2^k} \frac{|x_k - y_k|}{1 + |x_k - y_k|}, \qquad \forall x, y \in X.$$

[As an exercise: prove the triangle inequality for this d_X !] Then (X, d_X) becomes a metric space which we can topologise in the usual way.

The vector space of all sequences in a normed space: Choose some normed space Z with norm $z \mapsto \|z\|_Z$. Then define X as the vector space of all sequences in Z,

$$X := \Big\{ (x_1, x_2, x_3, \dots) : \ x_j \in Z \ \ \forall j \Big\}$$

and define a metric

$$d_X(x,y) := \sum_{k=1}^{\infty} \frac{1}{2^k} \frac{\|x_k - y_k\|_Z}{1 + \|x_k - y_k\|_Z}, \qquad \forall x, y \in X. \tag{1.1}$$

The space $C(\mathbb{R})$ of continuous functions on $\mathbb{R}$: observe that $f \mapsto \sup_{t\in\mathbb{R}} |f(t)|$ is *not* a norm on $X := C(\mathbb{R})$ since $f = \exp$ belongs to X but the mentioned supremum is infinite. On the other hand, we can define a metric on $C(\mathbb{R})$ as

$$d(f, g) = \sum_{k=1}^{\infty} \frac{1}{2^k} \frac{\|f - g\|_{L^\infty(-k,k)}}{1 + \|f - g\|_{L^\infty(-k,k)}}. \tag{1.2}$$

The discrete topology: let X be any non-empty set and define a topology as $\tau = \mathcal{P}(X)$, called the *discrete* topology.

The null topology: let X be any non-empty set and define a topology as $\tau = \{\emptyset, X\}$, called the *null* topology, also called the *chaotic topology.*

This is a good moment to present a method of comparing topologies.

Definition 1.11 (**Finer topologies and coarser topologies**). *Let X be a non-empty set with two topologies τ_1 and τ_2. If $\tau_1 \subset \tau_2$ (which means that every τ_1–open set is also τ_2–open), then we say that*

- *τ_2 is* finer *than τ_1,*
- *τ_1 is* coarser *than τ_2.*

We also say that τ_2 is stronger *than τ_1, and τ_1 is* weaker *than τ_2.*

In particular, the discrete topology is the finest topology of all topologies, and the null topology is the coarsest topology of all topologies. They have a feature—we know that these topologies *always* exist; like the tax office.

Lemma 1.12 (**Hausdorff criterion**[1]). *Let τ_1 and τ_2 be topologies on X. Then are equivalent:*

- *τ_2 is finer than τ_1*
- $\forall a \in X\colon \quad \forall B_1 \in \mathcal{B}_1(a)\colon \quad \exists B_2 \in \mathcal{B}_2(a)\colon \quad B_2 \subset B_1.$

1.1.4 Closed Sets

Definition 1.13 (**Closed set**). *Let (X, τ) be a topological space. A set $A \subset X$ is called* closed set *if its complement $X \setminus A$ is an open subset of X.*

Proposition 1.14 (**Properties of closed sets**). *Let (X, τ) be some topological space. Then we have:*

$$\emptyset \quad \text{and} \quad X \quad \text{are closed sets} \tag{C1}$$

$$\text{If} \quad C_1 \quad \text{and} \quad C_2 \quad \text{are closed sets, then also} \quad C_1 \cup C_2 \tag{C2}$$

$$\text{If} \quad C_i \quad \text{are closed sets, with } i \text{ from some index set } I\text{, then also} \quad \bigcap_{i\in I} C_i \tag{C3}$$

Proof. Nice exercise. Using (O1), (O2), (O3) is recommended here. □

Example 1.15. *Let $X = \mathbb{R}^3$ and $\tau = \mathcal{P}(X)$. Then every subset of $\mathbb{R}^3$ is open and closed at the same time. We admit that this topology is a weirdo.*

Remark 1.16. *One could also have chosen (C1), (C2), (C3) as the foundation of defining a topology instead of (B1), (B2), (B3), leading to the same results in the end.*

Definition 1.17 (**Cluster point of a set, Closure of a set**). *Let A be a set in a topological space (X, τ). We say that a point $a \in X$ is a* cluster point of A *if each basic neighbourhood $B \in \mathcal{B}(a)$ contains some $a' \in A$ with $a' \neq a$. All cluster points of A form the* cluster point set of A,

$$\text{c.p.}(A) := \{a \in X : a \quad \text{is cluster point of} \quad A\}.$$

We define the closure of A *as*

$$\overline{A} := A \cup \text{c.p.}(A).$$

[1] Felix Hausdorff, 1868–1942

Proposition 1.18. *For every set A in a topological space, $\overline{A}$ is a closed set.*

Proof. We will show that $X \setminus \overline{A}$ is an open set. Pick some

$$a_0 \in X \setminus \overline{A} = (X \setminus A) \cap (X \setminus \mathrm{c.p.}(A)).$$

Then $a_0 \notin A$ and $a_0 \notin \mathrm{c.p.}(A)$.

Now some basic neighbourhood $B_1 \in \mathcal{B}(a_0)$ exists with $B_1 \subset (X \setminus A)$, because: otherwise every $B_1 \in \mathcal{B}(a_0)$ would contain some element of A (which differs from a_0 due to $a_0 \notin A$), hence a_0 would be a cluster point of A.

And also some basic neighbourhood $B_2 \in \mathcal{B}(a_0)$ exists with $B_2 \subset (X \setminus \mathrm{c.p.}(A))$, because: otherwise every $B_2 \in \mathcal{B}(a_0)$ would contain some $b_2 \in \mathrm{c.p.}(A)$, and then (B3) gives us some $C_2 \in \mathcal{B}(b_2)$ with $b_2 \in C_2 \subset B_2$. Now b_2 being a cluster point of A means that there is some $b_2' \in C_2 \cap A$ with $b_2' \neq b_2$. And also $b_2' \neq a_0$ because $a_0 \notin A$. Hence we had shown that for each $B_2 \in \mathcal{B}(a_0)$ some $b_2' \in B_2 \cap A$ exists with $b_2' \neq a_0$, hence a_0 would be a cluster point of A, contrary to the choice of a_0.

Now we apply (B2) and find some $B_3 \in \mathcal{B}(a_0)$ with

$$a_0 \in B_3 \subset B_1 \cap B_2 \subset (X \setminus A) \cap (X \setminus \mathrm{c.p.}(A)) = X \setminus \overline{A},$$

and this is the definition of $X \setminus \overline{A}$ being open. □

Proposition 1.19. *A set A in a topological space is closed if and only if* $\mathrm{c.p.}(A) \subset A$.

Proof. Assume $\mathrm{c.p.}(A) \subset A$. Then $\overline{A} = A$, and $\overline{A}$ is always closed, hence A is closed.

Now let A be a closed set. Then $X \setminus A$ is open. Choose any $b \in X \setminus A$. Then there is some neighbourhood $B \in \mathcal{B}(b)$ with $b \in B \subset X \setminus A$, in particular $B \cap A = \emptyset$. Hence b can not be a cluster point of A, and therefore each cluster point of A is contained in A. □

Remark 1.20. *We mention without proof: $\overline{A}$ is the intersection of all closed sets that contain A.*

We come to an application that goes back to Euclid.

Theorem 1.21. *The set of prime numbers is infinite.*

Proof. We introduce a topology on $\mathbb{Z}$ by defining a neighbourhood base $\mathcal{B}(a)$ for every $a \in \mathbb{Z}$:

$$\mathcal{B}(a) := \{N_{a,b} \colon\ b \in \mathbb{N}_+\}, \qquad N_{a,b} := \{x = a + nb \colon\ n \in \mathbb{Z}\}.$$

Then $a \in N_{a,b}$ for every $a \in \mathbb{Z}$ and every $b \in \mathbb{N}_+$, hence (B1) holds. And given $N_{a,b} \in \mathcal{B}(a)$, $N_{a,c} \in \mathcal{B}(a)$, we quickly check $N_{a,bc} \subset N_{a,b} \cap N_{a,c}$, which yields (B2). To check (B3), we take $a \in \mathbb{Z}$ and some $N_{a,c} \in \mathcal{B}(a)$ with arbitrary $c \in \mathbb{N}_+$. Then we take an arbitrary $b \in N_{a,c}$, hence $b = a + n_0 c$ for some $n_0 \in \mathbb{Z}$. We need to construct $C \in \mathcal{B}(b)$ with $C \subset N_{a,c}$. We can write $C = N_{b,d}$ with some $d \in \mathbb{N}_+$ and then have to make sure that $N_{b,d} \subset N_{a,c}$ which means $b + nd \in N_{a,c}$ for every $n \in \mathbb{Z}$. Just take $d = c$.

Having defined this collection of neighbourhood bases $\mathcal{B}(a)$, we then build open sets as in Definition 1.3, and it follows that every open set is either empty or it contains an infinite number of elements. Moreover, each $N_{a,b}$ is an open set. But it is also a closed set, because it is the complement of an open set:

$$N_{a,b} = \mathbb{Z} \setminus \bigcup_{i=1}^{b-1} N_{a+i,b}.$$

Now every integer different from ± 1 has at least one prime divisor (call it p), and then this integer belongs to $N_{0,p}$. It follows that

$$\mathbb{Z} \setminus \{-1, +1\} = \bigcup_{p \text{ prime}} N_{0,p}. \tag{1.3}$$

Assume that there are only finitely many primes. Then the right-hand side of (1.3) is closed (as a union of finitely many closed sets), and then $\{-1, +1\}$ must be open (as the complement of a closed set). Contradiction. □

1.1.5 Induced Topologies

Definition 1.22 (**Induced topology, Relative topology**). *Consider some topological space* (X, τ) *and some non-empty subset* X_0 *of* X*. To each* $a_0 \in X_0$*, we define a collection* $\mathcal{B}_0(a_0)$ *of basic neighbourhoods on* X_0 *by*

$$\mathcal{B}_0(a_0) := \Big\{ X_0 \cap B \colon\; B \in \mathcal{B}(a_0) \Big\},$$

and then we construct—from all the neighbourhood bases $\mathcal{B}_0(a_0)$ *as* a_0 *runs through* X_0*—a topology* τ_0 *of* X_0*. This topology is called the* induced topology on X_0*, also called the* relative topology on X_0 *or the* trace topology on X_0*.*

Obviously, it needs to be shown that the collections $\mathcal{B}_0$ as constructed above actually satisfy (B1), (B2), (B3). This is a nice exercise.

We say that a set Ω_0 is *open in* X_0 if it is contained in X_0 and open in the induced topology τ_0. For instance, take $X = \mathbb{R}$ and equip it with the metric topology. Put $X_0 = [0, 1]$, equipped with the relative topology. Then $(0.7, 1]$ is *open in* X_0, but not open in X.

1.1.6 Converging Sequences, and the Hausdorff Property

Definition 1.23 (**Converging sequence**). *Let* (X, τ) *be a topological space. We say that a sequence* $(x_1, x_2, x_3, \dots) \subset X$ converges to $x^* \in X$ *(and write this as* $\lim_{n\to\infty} x_n = x^*$*) if*

$$\forall B \in \mathcal{B}(x^*)\colon \quad \exists N_0 \in \mathbb{N}\colon \quad \forall n \geq N_0\colon \quad x_n \in B.$$

Example 1.24. *Take some set* X *and equip it with the null topology. Then every sequence in* X *converges simultaneously to each element of* X*.*

This is not what one usually wants, and the obstacle is that we do not have enough neighbourhoods.

Definition 1.25 (**Hausdorff property**). *Let* (X, τ) *be a topological space. We say that* τ *has the* Hausdorff property *and call* (X, τ) *a* Hausdorff space *if the following holds:*

$$\forall x_1, x_2 \in X\colon \quad \text{if } x_1 \neq x_2 \quad \text{then} \quad \exists B_1 \in \mathcal{B}(x_1), \quad \exists B_2 \in \mathcal{B}(x_2)\colon \quad B_1 \cap B_2 = \emptyset.$$

Example 1.26. *The only topology that turns* $\mathbb{N}$ *into a Hausdorff space is the discrete topology.*

Lemma 1.27. *Every metric topology has the Hausdorff property.*

Lemma 1.28. *In every Hausdorff space, every set* $\{p\}$ *that contains only one point is a closed set.*

We discuss some examples of converging sequences.

- Take the vector space X of sequences that take values in a normed space Z, and equip it with the topology that comes from the metric (1.1). Then a sequence of elements of X converges (topologically) if and only if it converges component-wise.
- Take the vector space $C(\mathbb{R})$ of continuous functions on $\mathbb{R}$, equipped with the topology that comes from the metric (1.2). Then this convergence is equivalent to the uniform convergence on every compact sub-interval of $\mathbb{R}$.

Definition 1.29 (**Dense subset**). *Let* (X, τ) *be a topological space, and* $R \subset S \subset X$*. Then we say that* R is dense in S *if*

$$\forall s \in S\colon \quad \forall B \in \mathcal{B}(s)\colon \quad \exists r \in R \cap B.$$

An equivalent statement is: $\overline{R} \supset S$.

Warning 1.30. *We have to be careful with sequences. All the following statements are* ***wrong****, even in Hausdorff spaces:*

- *The subset A of B is dense in B if and only if, for each $b \in B$, a sequence in A can be found that converges to b.*
- *A point $a \in X$ is a cluster point of A if and only if a sequence in A can be found that converges to a.*
- *A map between two topological spaces is continuous (to be defined in the next section) if and only if it is sequentially continuous (defined below).*
- *A set A is compact (defined by means of covers by open sets in the usual way) if and only if it is sequentially compact (every sequence contains a subsequence that has a limit in A).*

The obstacle is always that there need not be a countable neighbourhood base. On the other hand, those statements are correct in metric spaces.

The general strategy is to define all relevant concepts using neighbourhoods alone, and to stop thinking in terms of converging sequences.

1.1.7 Continuous Maps

We recall some rules for any map $f\colon X \to Y$ between two sets X and Y. Here $I \neq \emptyset$ is any index set:

$$f\left(\bigcup_{i\in I} A_i\right) = \bigcup_{i\in I} f(A_i), \qquad \forall A_i \subset X,$$
$$f(A_1 \cap A_2) \subset f(A_1) \cap f(A_2), \qquad \forall A_1, A_2 \subset X,$$
$$f(A_1 \setminus A_2) \supset f(A_1) \setminus f(A_2), \qquad \forall A_1, A_2 \subset X,$$

and equality in the last two lines (for all $A_i \subset X$) will hold only if and only if f is injective.
Additionally, if $A \subset X$, then

$$A \subset f^{-1}(f(A)),$$

with equality (for all A) if and only if f is injective. Similarly, for all $B \subset Y$,

$$f(f^{-1}(B)) \subset B,$$

with equality (for all B) if and only if f is surjective.
The good news is that f^{-1} alone behaves perfectly:

$$f^{-1}\left(\bigcup_{i\in I} B_i\right) = \bigcup_{i\in I} f^{-1}(B_i), \qquad \forall B_i \subset Y,$$
$$f^{-1}\left(\bigcap_{i\in I} B_i\right) = \bigcap_{i\in I} f^{-1}(B_i), \qquad \forall B_i \subset Y,$$
$$f^{-1}(B_1 \setminus B_2) = f^{-1}(B_1) \setminus f^{-1}(B_2), \qquad \forall B_1, B_2 \subset Y.$$

Definition 1.31 (Continuous maps). *Let (X, τ) and (Y, σ) be topological spaces. A map $f\colon X \to Y$ is called* continuous at $x^* \in X$ *if*

$$\forall V \in \mathcal{B}(f(x^*)): \quad \exists U \in \mathcal{B}(x^*): \quad f(U) \subset V.$$

We say that this f is continuous *if it is continuous at every $x^* \in X$.*

Lemma 1.32. *For a map $f\colon X \to Y$ with X and Y as above, the following statements are equivalent:*

- *f is continuous,*
- *for every open set $\Omega \subset Y$, its pre-image $f^{-1}(\Omega)$ is open in X,*
- *for every closed set $B \subset Y$, its pre-image $f^{-1}(B)$ is closed in X.*

From the penultimate • we quickly deduce that the composition $g \circ f$ of two continuous maps g and f is again continuous.

For sake of completeness, we clarify what is meant by *sequential continuity at* x^*: for every sequence $(x_1, x_2, \dots) \subset X$ that converges to x^*, the sequence $(f(x_1), f(x_2), \dots) \subset Y$ converges to $f(x^*)$.

Lemma 1.33 (**Continuity implies sequential continuity**). *Let the map $f: X \to Y$ between the topological spaces (X, τ) and (Y, σ) be continuous. Then it is sequentially continuous.*

Proof. Pick $x^* \in X$ and some sequence $(x_1, x_2, \dots) \subset X$ that converges to x^*. We need to show that $\lim_{k\to\infty} f(x_k) = f(x^*)$. To this end, we pick some basic neighbourhood V of $f(x^*)$. Owing to the continuity of f, there is some basic neighbourhood U of x^* with $f(U) \subset V$. And by Definition 1.23, we find some $N_0 \in \mathbb{N}$ with $x_n \in U$ for all $n \geq N_0$. But then $f(x_n) \subset V$ for such n. □

Lemma 1.34. *Let (X, τ) be a topological space, and $f: X \to \mathbb{C}$ be continuous, where we have endowed $\mathbb{C}$ with the usual metric topology. Then $\ker f = \{x \in X : f(x) = 0\}$ is a closed subset of X.*

1.1.8 How to Construct Topologies

Proposition 1.35 (**Intersection of topologies**). *Let us be given a non-empty set X and a non-empty index set I. If τ_i with $i \in I$ are topologies on X, then so is*

$$\tau_* := \bigcap_{i\in I} \tau_i.$$

Of all topologies which are coarser than every τ_i, τ_* is the finest.

Sketch of proof. Verify (O1), (O2), (O3). □

Remark 1.36.
- *The intersection τ_* need not be Hausdorff even if each τ_i is Hausdorff because the intersection might consist of $\emptyset$ and X only.*
- *Each identity map*

$$\mathrm{id}_X : (X, \tau_*) \to (X, \tau_i), \qquad x \mapsto x$$

  is continuous.
- *Recall the definition of the infimum of some set $A \subset \mathbb{R}$: of all lower bounds of A, you take the largest. The intersection of topologies is defined similarly.*
- *A neighbourhood base of $a \in X$ in τ_* is given by all B with $a \in B \subset X$ which are τ_i–open for each $i \in I$.*

Proposition 1.37 (**Topology generated by a subset**). *Let us be given a non-empty set X. If $\mathcal{U} \subset \mathcal{P}(X)$ is some non-empty set, then*

$$\tau(\mathcal{U}) := \bigcap \Big\{\sigma : \mathcal{U} \subset \sigma, \quad \sigma \text{ is topology on } X\Big\} \tag{1.4}$$

(which means the intersection of all topologies σ that contain $\mathcal{U}$) is a topology on X, and it is the coarsest topology that contains $\mathcal{U}$. A neighbourhood base for $a \in X$ then is given by all finite intersections of elements of the set (of sub-sets of X)

$$\{X\} \cup \{U \in \mathcal{U} : \quad a \in U\}. \tag{1.5}$$

Proof. The discrete topology participates in the intersection in (1.4), hence the index set is non-empty. Now apply Proposition 1.35.

For $a \in X$, define $\mathcal{B}(a)$ as the set of all finite intersections of elements of the set (1.5). Next we verify (B1), (B2), (B3). Then $X \in \mathcal{B}(a)$, hence $\mathcal{B}(a) \neq \emptyset$, then (B1) quickly follows. To show (B2), it suffices to choose $B_3 = B_1 \cap B_2$. And (B3) is verified as follows. Consider any $B \subset \mathcal{B}(a)$. If $B = X$ then the claim is obvious. If $B \neq X$ then $B = \cap_{j=1}^n U_j$ for some $U_j \in \mathcal{U}$ with $a \in U_j$ for all j. For every $b \in B$, choose

$C = B$. Hence (B3) holds. This shows that all the $\mathcal{B}(a)$ (as a runs through X) are a neighbourhood base for some topology ϱ on X.

Next we show that $\varrho = \tau(\mathcal{U})$. Each $U \in \mathcal{U}$ is contained in all topologies σ that participate in the intersection of (1.4), consequently $\mathcal{B}(a) \subset \tau(\mathcal{U})$ for all $a \in X$, hence $\varrho \subset \tau(\mathcal{U})$, which means ϱ is coarser than $\tau(\mathcal{U})$. But ϱ contains $\mathcal{U}$, and therefore $\varrho \supset \tau(\mathcal{U})$. □

Proposition 1.38 (**"Union" of topologies**). *For some non-empty index set I, let us be given topologies τ_i (with $i \in I$) on X. Define*

$$\tau^* := \bigcap \{\sigma\colon\ \tau_i \subset \sigma \ \ (\forall i), \quad \sigma \text{ is topology on } X\}\,.$$

Then τ^* is again a topology on X, and

- *of all topologies which are finer than every τ_i, it is the coarsest;*
- *each identity map*

  $$\mathrm{id}_X\colon (X, \tau_i) \to (X, \tau^*), \qquad x \mapsto x$$

  is continuous.
- *A neighbourhood base for $a \in X$ is given by all finite intersections of elements of*

  $$\{X\} \cup \left\{ U \in \bigcup_{i \in I} \tau_i\colon\ \ a \in U \right\}.$$

Proof. Nice exercise. □

Remark 1.39.

- *The topology τ^* is* not *the set–theoretical union of the sets τ_i (similarly to the direct sum of sub-vector spaces, which is not equal to the union set of these sub-vector spaces either).*
- *But $\tau^* = \tau(\bigcup_{i\in I} \tau_i)$ in the sense of Proposition 1.37.*

Next we show how topologies can be defined using maps between sets of which some are already topological spaces. Suppose you have sets X and Y, with Y already being equipped with a topology. And you consider maps from X to Y. The goal is to construct a topology on X. If you provide too few open sets on X, then almost none map from X to Y will be continuous, which is a situation without sense. On the other hand, if X has too many open sets, then almost every map from X to Y will be continuous, and again the concept "continuity" becomes meaningless. It seems advisable to endow X with "just the right amount of open sets" as a topology. We call it the *initial topology*.

Proposition 1.40 (**Initial topology**). *Let us be given a non-empty set X. Assume that (Y_i, σ_i) with $i \in I$ are topological spaces, and we have maps $f_i\colon X \to Y_i$ for each i. Then*

$$\bigcap \Big\{ \tau \subset \mathcal{P}(X)\colon\ \ \tau \text{ is topology on } X \text{ and all } f_i\colon (X, \tau) \to (Y_i, \sigma_i) \text{ are continuous } (i \in I) \Big\}$$

is a topology on X, and it is the coarsest topology on X for which all f_i become continuous. We call it the initial topology on X generated by the maps $f_i\colon X \to Y_i$. *A neighbourhood base of $a \in X$ is given by the finite intersections of elements of*

$$\{X\} \cup \Big\{ U \subset X\colon\ \exists i \in I, \quad \exists V \in \mathcal{B}_i(f_i(a))\colon \quad U = f_i^{-1}(V) \Big\} \tag{1.6}$$

with $\mathcal{B}_i(f_i(a))$ as a neighbourhood base consisting of (σ_i–)open neighbourhoods of $f_i(a) \in Y_i$ in σ_i.

Proof. Observe that $\tau = \mathcal{P}(X)$ participates in the intersection, hence the index set is non-empty. Now apply Proposition 1.35. A neighbourhood base for the topology of the point $f_i(a)$ is given by all open sets $V \subset Y_i$ that contain $f_i(a)$. Then $f_i^{-1}(V)$ contains $a \in X$ and is required to be open in τ, and therefore $\tau \subset \mathcal{P}(X)$ must contain all the elements of the set (1.6), as a runs through X. Now τ is (by its construction) the coarsest topology that contains (for all $a \in X$) the set (1.6) as a subset. It suffices to apply Proposition 1.38. □

Example 1.41. *Take $Y = [-1, 1]$ with the metric topology, $X = \mathbb{R}$ and $f: X \to Y$ with $y = \sin(x)$. Determine the initial topology on X. Explain why it does not have the Hausdorff property.*

Example 1.42 (The induced topology, reloaded). *Consider some topological space (X, τ) and choose some non-empty subset $X_0 \subset X$. Then we have the inclusion map $f: X_0 \to X$ that sends every $x \in X_0$ to itself. The initial topology on X_0 generated by the map f is exactly the* induced topology *on X_0.*

Here comes one more example of the initial topology.

Definition 1.43 (Product space topology). *Let (X_i, τ_i) be topological spaces, with $i = 1, \dots, n$. Put $X := X_1 \times \dots \times X_n$, and define the projections*

$$\pi_i : X \to X_i, \qquad \pi_i(x) := x_i,$$

where $i = 1, \dots, n$ and $x = (x_1, \dots, x_n)$. Then the product space topology *on X is defined as the initial topology on X generated by the maps π_i.*

Lemma 1.44. *In the situation of Definition 1.43, a neighbourhood base of $a = (a_1, \dots, a_n) \in X$ is*

$$\mathcal{B}(a) = \prod_{i=1}^{n} \mathcal{B}_i(a_i) := \mathcal{B}_1(a_1) \times \dots \times \mathcal{B}_n(a_n),$$

where $\mathcal{B}_i(a_i)$ denotes a neighbourhood base of a_i in the topological space (X_i, τ_i).

Proof. Let $B_2 \in \mathcal{B}_2(\pi_2(a))$ be a basic neighbourhood of $\pi_2(a) = a_2 \in X_2$. By Proposition 1.40, the pre-image $\pi_2^{-1}(B_2) = X_1 \times B_2 \times X_3 \times \dots \times X_n$ is a member of the neighbourhood base of $a \in X$, and then $\bigcap_{i=1}^n \pi_i^{-1}(B_i)$ also is a basic neighbourhood of $a \in X$. But $\bigcap_{i=1}^n \pi_i^{-1}(B_i) = B_1 \times B_2 \times \dots \times B_n \in \prod_{i=1}^n \mathcal{B}_i(a_i)$. □

Having understood how to transfer a topology from Y to X, we now consider the opposite direction.

Proposition 1.45 (Final topology). *For $i \in I \neq \emptyset$, let us be given topological spaces (X_i, τ_i), and additionally some set Y. Consider maps $f_i : X_i \to Y$ for all $i \in I$. Then the following holds:*

- *there is some topology σ on Y, given by*

$$\sigma = \left\{ \Omega \subset Y : \; f_i^{-1}(\Omega) \text{ is open subset of } X_i, \qquad \forall i \in I \right\},$$

  called the final topology on Y generated by $\{f_i : i \in I\}$,

- *this topology on Y is the* finest *topology on Y that makes all the maps $f_i : X_i \to Y$ continuous,*
- *we have the intersection*

$$\sigma = \bigcap \{ \sigma_i \subset \mathcal{P}(Y) : \; \sigma_i \text{ is final topology generated by } f_i \text{ alone} \}.$$

Describing a neighbourhood base for the final topology is not easy, see [2, 0.2.16] for some ideas.

Example 1.46. *Take $X = \mathbb{R}$ with the metric topology $\tau_{|\cdot|}$, $Y = \mathbb{R}$ and $f: X \to Y$ with $y = \sin(x)$. Determine the final topology on Y.*

Example 1.47 (Quotient topology). *Let (X, τ) be a topological space with some equivalence relation $\sim$. Define $X/\sim$ as the set of equivalence classes generated by the relation $\sim$. Then the map*

$$\pi : X \to X/\sim, \qquad x \mapsto [x]$$

that maps x to its equivalence class $[x]$ is called canonical surjection of X onto $X/\sim$. *The* quotient topology on $X/\sim$ *is defined as the final topology on $X/\sim$ generated by π.*

1.2 Topological Vector Spaces

1.2.1 Definition and Immediate Consequences

Definition 1.48 (Topological vector space). *A vector space X over $\mathbb{K} \in \{\mathbb{C}, \mathbb{R}\}$ is called* topological vector space (tvs) *if X is equipped with a topology τ, and the usual vector space operations*

$$+: X \times X \to X,$$
$$\cdot: \mathbb{K} \times X \to X$$

are continuous, where $X \times X$ and $\mathbb{K} \times X$ have the product topologies, and $\mathbb{K}$ has the norm topology.

Lemma 1.49. *Let $(X, \mathbb{K}, \tau)$ be a tvs. Then the vector space operations are separately continuous, which means: for each $x_0 \in X$ and each $\lambda_0 \in \mathbb{K}$, the maps*

$$f: X \to X, \qquad x \mapsto x + x_0,$$
$$g: X \to X, \qquad x \mapsto \lambda_0 x$$

are continuous.

Proof. Nice exercise. Using Lemma 1.44 is recommended. □

Corollary 1.50. *Let $(X, \mathbb{K}, \tau)$ be a tvs, and choose some $a \in X$. Then a neighbourhood base $\mathcal{B}(a)$ is given by $a + \mathcal{B}(0)$ (which means by translating a neighbourhood base of the origin by the vector a).*

In particular: the topology on a tvs is uniquely determined after a neighbourhood base of the origin has been specified.

Proposition 1.51. *Let X and Y be tvs over $\mathbb{K}$.*

- *A linear map from X to Y is continuous if and only if it is continuous at the origin.*
- *Sums, scalar multiples and compositions of linear maps are again linear maps. In particular, $\mathcal{L}(X, Y)$ is a vector space, and $\mathcal{L}(X, X)$ is an algebra.*
- *For every $A \in \mathcal{L}(X, Y)$, $\ker(A)$ is closed if and only if Y is a Hausdorff space.*

Lemma 1.52. *Let $(X, \mathbb{K}, \tau)$ be a tvs, and U be a basic neighbourhood of 0. Then the following holds:*

- *There is some $U' \in \mathcal{B}(0)$ with $U' + U' \subset U$.*
- *There is some $U' \in \mathcal{B}(0)$ with $\alpha U' \subset U$ for all α with $|\alpha| \le 1$.*
- *For each $x_0 \in X$, there is some positive δ such that $x_0 \in \lambda U$ for all $\lambda \in \mathbb{K}$ with $|\lambda| \ge \frac{1}{\delta}$.*

Proof. We make repeated use of Definition 1.31 and Lemma 1.44, without mentioning it. The first claim is shown like this: Addition is continuous at $(0, 0) \in X \times X$, hence, for all $U \subset \mathcal{B}(0)$, there are $U_1 \in \mathcal{B}(0)$ and $U_2 \in \mathcal{B}(0)$ with $U_1 + U_2 \subset U$. Now choose $U' = U_1 \cap U_2$.

We come to the penultimate claim. The multiplication by scalars is continuous at $(0, 0) \in \mathbb{K} \times X$. Hence there is some $W \in \mathcal{B}(0)$ and some $\varepsilon > 0$ such that $\alpha x \subset U$ for all $\alpha \in \mathbb{K}$ with $|\alpha| < \varepsilon$ and all $x \in W$. Now choose $U' = \frac{\varepsilon}{2} W$.

Concerning the last claim: the multiplication by scalars is continuous at $(0, x_0) \in \mathbb{K} \times X$. As a special case, for each $U \in \mathcal{B}(0)$ some positive δ exists such that for all $\alpha \in \mathbb{K}$ with $|\alpha| < \delta$, we have $\alpha x_0 \in U$. It suffices to substitute $\lambda = \frac{1}{\alpha}$. □

We say that all $U \in \mathcal{B}(0)$ are *absorbent sets*, by which we mean that if you blow up U sufficiently (multiply it by growing λ) then you eventually arrive at any $x_0 \in X$ and afterwards never lose it when λ keeps growing.

Definition 1.53 (Bounded subset of a tvs). *A subset A of a tvs X is called* bounded *if for each neighbourhood U of the origin, there is a positive λ_0 with $A \subset \lambda U$ for each $\lambda \ge \lambda_0$.*

Intuitively, this means that the set A will be swallowed by each neighbourhood of the origin if you pump up the neighbourhood big enough.

1.2.2 Seminorms

Sometimes we prefer to work with seminorms on a tvs instead of neighbourhoods of the origin because seminorms allow for easier calculations.

Definition 1.54 (Minkowski–functional). *Let $(X, \mathbb{K}, \tau)$ be a tvs and $U \in \mathcal{B}(0)$ be a basic neighbourhood of the origin. Define a function $p_U : X \to \mathbb{R}_{\geq 0}$ by*

$$p_U(x) := \inf \{\lambda \geq 0 \colon x \in \lambda U\},$$

called Minkowski–functional of U.

Definition 1.55. *A set M in a vector space $(X, \mathbb{K})$ is called* absolutely convex *if*

$$x, y \in M, \qquad |\alpha| + |\beta| \leq 1 \quad \textit{imply} \quad \alpha x + \beta y \in M.$$

Proposition 1.56. *If $U \in \mathcal{B}(0)$ is an* absolutely convex *basic neighbourhood of the origin, then p_U is a seminorm on X, which means*

$$\begin{aligned} p_U(x) &\geq 0 \qquad \forall x \in X, \\ p_U(\lambda x) &= |\lambda| p_U(x), \qquad \forall x \in X, \quad \forall \lambda \in \mathbb{K}, \\ p_U(x+y) &\leq p_U(x) + p_U(y), \qquad \forall x, y \in X. \end{aligned}$$

Proof. Nice exercise. □

Remark 1.57. *Now we have a method of constructing a collection of seminorms from a base of* convex *neighbourhoods of the origin:*

- *Let $\mathcal{B}_c(0)$ be a neighbourhood base of the origin, consisting of absorbent convex sets.*
- *Then (by Lemma 1.52) there is also a neighbourhood base $\mathcal{B}_{ac}(0)$ consisting of absolutely convex absorbent neighbourhoods of the origin (and $\mathcal{B}_{ac}(0)$ generates the same topology as $\mathcal{B}_c(0)$).*
- *Each $U \in \mathcal{B}_{ac}(0)$ gives rise to a seminorm p_U on X. This seminorm p_U is the Minkowski functional associated to U.*
- *The topology of X has the Hausdorff property if (for each $0 \neq x \in X$) some p_U (with U depending on x) exists with $p_U(x) \neq 0$.*

This procedure can be reversed:

Proposition 1.58 (Building a TVS from a Seminorm Family). *Let X be a vector space over $\mathbb{K}$, and let $\{p_i \colon i \in I\}$ be a family of seminorms on X. To each $i \in I$ and each $\varepsilon > 0$, we define*

$$U_{i,\varepsilon} := \{x \in X \colon \; p_i(x) < \varepsilon\},$$

and then we put, for every finite subset J of I,

$$U_{J,\varepsilon} := \bigcap_{j \in J} U_{j,\varepsilon}.$$

- *Then*

  $$\mathcal{B}(0) := \{U_{J,\varepsilon} \colon \; \varepsilon > 0, \quad J \textit{ finite subset of } I\}, \qquad \mathcal{B}(a) := \mathcal{B}(0) + a$$

  form a neighbourhood base of a topology on X.
- *If X had already a topology before, and the seminorms p_i have been constructed by the procedure in Remark 1.57, then the topology constructed from these seminorms is equivalent to the original topology.*
- *The topology has the Hausdorff property if for each $x \in X$ some $i \in I$ exists with $p_i(x) > 0$.*

- *A set $A \subset X$ is bounded if and only if*

$$\forall i \in I\colon \quad \exists C_i > 0\colon \quad \forall x \in A\colon \quad p_i(x) < C_i.$$

- *A sequence $(x_1, x_2, \dots) \subset X$ converges to a limit $x^* \in X$ if and only if*

$$\forall i \in I\colon \quad \lim_{m\to\infty} p_i(x_m - x^*) = 0.$$

Such a vector space X, equipped with this topology, is called *locally convex (topological vector) space.*

Proposition 1.59. *Let X and Y be locally convex topological vector spaces, with $\{p_i\colon i \in I\}$ being the collection of seminorms on X. A linear map f of X into Y is continuous if for each seminorm q on Y, there is a finite set $I' \subset I$ and a constant C_q such that*

$$\forall x \in X\colon \quad q(f(x)) \le C_q \sum_{i\in I'} p_i(x).$$

The situation becomes nicer if the index set I is countable (WLOG $I = \mathbb{N}_+$), in which case we can define a function $d\colon X \times X \to \mathbb{R}$ like this:

$$d(x_1, x_2) = \sum_{j=1}^{\infty} 2^{-j} \frac{p_j(x_1 - x_2)}{1 + p_j(x_1 - x_2)}, \qquad x_1, x_2 \in X.$$

We quickly show that d has all the properties of a metric. We say that the tvs X is *metrisable.*

If we now suppose additionally that X is a complete space (which means that each metric Cauchy sequence has a limit in X), then X is called a Fréchet [2] space.[3] The advantage is now that continuity is again equivalent to sequential continuity, density equivalent to sequential density, compactness (defined by coverings using open sets) equivalent to sequential compactness.

Lemma 1.60. *Let X be a Fréchet space, and let Y be any locally convex tvs. Then a linear map $f\colon X \to Y$ is continuous if and only if it is sequentially continuous.*

Proof. We need to show that f is continuous at the origin (see Proposition 1.51). Suppose that f were not continuous at the origin. Then there is a neighbourhood V of $0 \in Y$ with $f^{-1}(V)$ not containing any neighbourhood of $0 \in X$. But the topology of X can also be generated by a metric d, hence for every $\varepsilon > 0$, the ε-ball $\{x \in X\colon d(0, x) < \varepsilon\}$ is not fully contained in $f^{-1}(V)$. This means that for any $\varepsilon > 0$, there is some $x_\varepsilon \in X$ with $d(0, x_\varepsilon) < \varepsilon$ but $f(x_\varepsilon) \notin V$. This contradicts the sequential continuity of f. □

We need to work with quotient spaces, too:

Definition 1.61 (Topology of a quotient space of a TVS). *Let X be a tvs and $Y \subset X$ be a sub-vector space. On X, we define an equivalence relation $\sim$ as $x_1 \sim x_2$ if and only if $x_1 - x_2 \in Y$. The quotient space X/Y comprises all the equivalence classes, and the canonical projection $\pi\colon X \to X/Y$ maps each element x to its equivalence class $[x] =\colon \pi x$.*

We say that a set $A \subset X/Y$ is open if $\pi^{-1}(A)$ is open in X.

Then X/Y actually is again a tvs.

Lemma 1.62. *Let X be a tvs, and Y be a sub-vector space of X.*

- *The quotient space X/Y is a Hausdorff space if and only if Y is a closed subspace of X.*
- *Let X be a normed space and Y is a closed sub-vector space of X. Then X/Y is actually a normed space with norm*

$$\|\pi x\|_{X/Y} := \inf_{y\in Y} \|x - y\|_X \,.$$

Proof.

- The quotient space X/Y is a Hausdorff space

  $\iff$ if and only if the set $\{[0]\}$ that contains only the zero vector $[0]$ of X/Y is a closed subset of X/Y,

  $\iff$ if and only if the complement $C := (X/Y) \setminus \{[0]\}$ is an open subset of X/Y,

  $\iff$ if and only if the pre-image $\pi^{-1}(C)$ is an open subset of X.

  Now it suffices to recognise that $\pi^{-1}(C)$ is the complement of Y in X.

[2] Maurice René Fréchet, 1878–1973

[3] The actual definition of Fréchet spaces includes the invariance: $d(x_1 + y, x_2 + y) = d(x_1, x_2)$, which obviously holds.

- By direct calculation, one checks that $\|(\pi x_1)+(\pi x_2)\|_{X/Y} \le \|\pi x_1\|_{X/Y} + \|\pi x_2\|_{X/Y}$ and $\|\lambda\cdot(\pi x)\|_{X/Y} = |\lambda|\,\|\pi x\|_{X/Y}$, and also $\|\pi x\|_{X/Y}=0$ if and only if $x\in Y$. Therefore, the expression $\pi x\mapsto \|\pi x\|_{X/Y}$ actually is a norm. It remains to have a look at Example 1.47 to verify that this quotient norm actually generates the correct topology.

□

Example 1.63. *Let $X=\mathbb{K}^M$ (for some $M\in\mathbb{N}$) be a vector space of column vectors, equipped with the natural norm topology. Then the vector space of linear maps from X into $\mathbb{K}$ is isomorphic to the vector space of row vectors with M entries, and these maps (from X to $\mathbb{K}$) are automatically continuous. We write X' for the set of linear and continuous maps from $\mathbb{K}^M$ into $\mathbb{K}$. Its elements are written as x', rows with M entries. A dash ($'$) on a vector informs us that it is a row.*

Next consider $Y=\mathbb{K}^N$ for some $N\in\mathbb{N}$. Then every linear and continuous map $\mathcal{A}\colon X\to Y$ is being generated by a matrix $A\in\mathbb{K}^{N\times M}$, in the sense of $y=\mathcal{A}x=A\cdot x$ as a product of matrix times column vector. We are being pedantic on purpose and insist on distinguishing $\mathcal{A}x$ (which is the map $\mathcal{A}$ applied to the vector x) from $A\cdot x$ (which is the matrix-times-vector product of A and x). Applying $y'\in Y'$ to $y\in Y$ gives a number $y'\cdot y\in\mathbb{K}$.

Moreover, the transposed matrix $A^\top$ then maps from Y to X, in the sense of $x=A^\top\cdot y$. If you transpose this equation, you get $x^\top=(A^\top\cdot y)^\top=y^\top\cdot A$. Now we can write (with $x'\in X'$ defined as $x':=y'\cdot A$)

$$\underbrace{y'}_{\in Y'}\cdot\underbrace{y}_{\in Y}=y'\cdot(A\cdot x)=(y'\cdot A)\cdot x=\underbrace{x'}_{\in X'}\cdot\underbrace{x}_{\in X},$$

and this could be abbreviated as

$$y'\cdot(\mathcal{A}x)=(\mathcal{A}^t y')\cdot x.$$

We may say that $A^\top$ belongs to a map $\mathcal{A}^t\colon Y'\to X'$.

We lift the ideas from this example to a more abstract level.

Definition 1.64 (Topological dual space). *Let X be a locally convex tvs over the field $\mathbb{K}$. The set of all linear and continuous maps from X into $\mathbb{K}$ is denoted by X', and it is called the* topological dual space[4].

Each $x\in X$ generates a seminorm p_x on X' via

$$T\mapsto p_x(T):=\left|\langle T\mid x\rangle_{X'\times X}\right|,\qquad T\in X'.$$

The topology of X', generated by these seminorms, is called the weak–∗–topology.

We make no attempt at describing a neighbourhood base for the weak–∗–topology, because seminorms are so much nicer.

Definition 1.65 (Transposed operator). *Let X and Y be locally convex spaces, and let $\mathcal{A}\colon X\to Y$ be linear and continuous. Then $\mathcal{A}$ generates a linear operator $\mathcal{A}^t\colon Y'\to X'$ via*

$$\left\langle \mathcal{A}^t y'\mid x\right\rangle_{X'\times X}:=\langle y'\mid \mathcal{A}x\rangle_{Y'\times Y},\qquad y'\in Y',\quad x\in X,$$

which we call transposed operator.

Proposition 1.66. *With the above notations, $\mathcal{A}^t$ is continuous.*

Proof. We start from Proposition 1.59. For each seminorm p of X', we need a finite collection of seminorms $q_1,\dots,q_k$ of Y' and a constant C, such that

$$\forall y'\in Y':\quad p(\mathcal{A}^t y')\le C\sum_{j=1}^k q_j(y').$$

Only those p are interesting that are generated by an $x\in X$ via

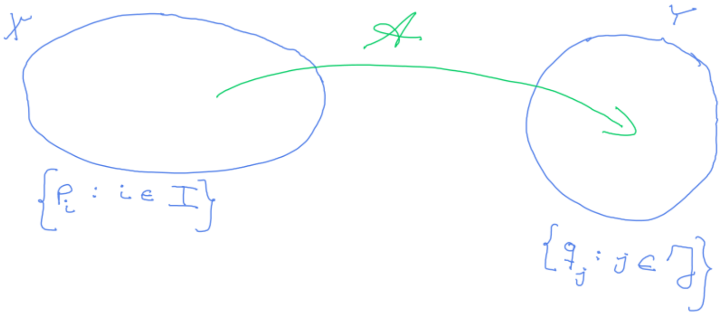
$$p(x')=\left|\langle x'\mid x\rangle_{X'\times X}\right|,\qquad \forall x'\in X',$$

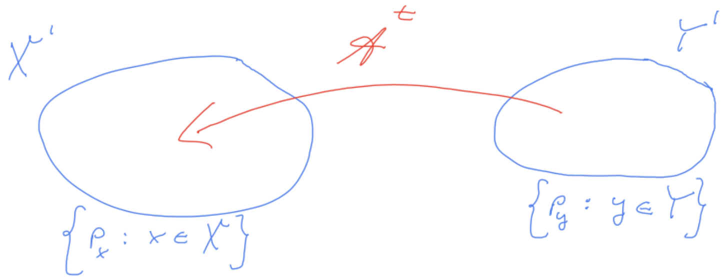
[4] We quickly check that X' is an algebraic vector space over the field $\mathbb{K}$

hence we have

$$p(\mathcal{A}^t y') = \left|\left\langle \mathcal{A}^t y' \mid x\right\rangle_{X'\times X}\right| = \left|\langle y' \mid \mathcal{A}x\rangle_{Y'\times Y}\right| =: q(y'),$$

where q is a seminorm on Y' that is being generated by $\mathcal{A}x \in Y$. □





You may ask "what are these transposed operators good for", and a part of the answer can be given when we look at matrices $A \in \mathbb{K}^{N\times M}$ that map from $\mathbb{K}^M$ into $\mathbb{K}^N$. It is easy to check that $\ker(A) \perp \operatorname{img}(A^\top)$ and also $\ker(A^\top) \perp \operatorname{img}(A)$. If you utilise the dimension formula for sub-vector spaces and the rank–nullity theorem[5], then you can show that

$$\mathbb{K}^M = \ker(A) \oplus \operatorname{img}(A^\top), \qquad \mathbb{K}^N = \ker(A^\top) \oplus \operatorname{img}(A),$$

as orthogonal direct sum of sub-vector spaces. A conclusion then is: A generates a surjective map if and only if $A^\top$ generates an injective map. We think that this is a nice result.

The next proposition is an attempt at generalising this principle.

Proposition 1.67. *Let X and Y be locally convex topological vector spaces with the Hausdorff property. If the operator $\mathcal{A}\colon X \to Y$ is continuous and has dense range, then $\mathcal{A}^t$ is injective.*

Conversely: if $\mathcal{A}\colon X \to Y$ is continuous and $\mathcal{A}^t$ is injective, then $\mathcal{A}$ has dense range.

Proof. **Suppose that $\operatorname{img}(\mathcal{A})$ is a dense subset of Y,** which means $\overline{\operatorname{img}(\mathcal{A})} = Y$. We wish to show that $\mathcal{A}^t y' = 0$ implies $y' = 0$. Let $\mathcal{A}^t y' = 0 \in X'$, which means

$$\forall x \in X: \quad \langle \mathcal{A}^t y' \mid x\rangle_{X'\times X} = 0, \quad \text{hence} \quad \langle y' \mid \mathcal{A}x\rangle_{Y'\times Y} = 0.$$

Hence $\operatorname{img}(\mathcal{A})$ is contained in the pre-image $(y')^{-1}(\{0\})$. But the one-member set $\{0\}$ is a closed subset of $\mathbb{K}$, and y' is continuous, and therefore $(y')^{-1}(\{0\})$ is a closed subset of Y. Finally we observe

$$Y \supset (y')^{-1}(\{0\}) \supset \overline{\operatorname{img}(\mathcal{A})} = Y,$$

with the implication $(y')^{-1}(\{0\}) = Y$, or $y' = 0 \in Y'$. For the second $\supset$, we have used that the closure $\overline{\operatorname{img}(\mathcal{A})}$ is the intersection of all closed sets that contain $\operatorname{img}(\mathcal{A})$, and $(y')^{-1}(\{0\})$ is one of them.

Conversely, let $\mathcal{A}^t$ be injective, and suppose $\operatorname{img}(\mathcal{A})$ to not be dense in Y. We will now construct some non-zero $y' \in Y'$ with $\mathcal{A}^t y' = 0$, which will contradict the injectivity of $\mathcal{A}^t$. There is some $y_0 \in Y$ with $y_0 \notin \overline{\operatorname{img}(\mathcal{A})}$.

We build the quotient space $Z = Y/\overline{\operatorname{img}(\mathcal{A})}$, and we let $\pi\colon Y \to Z$ be its canonical projection. Define $z_0 := \pi y_0$. Then $z_0 \neq 0$ because $y_0 \notin \overline{\operatorname{img}(\mathcal{A})}$. By Lemma 1.62, Z is a Hausdorff space (in particular a normed space, and $\|z_0\|_Z \neq 0$).

[5] Dimensionsformel für lineare Abbildungen

Now we apply the Hahn–Banach theorem[6] with $Z_0 = \operatorname{span}(z_0)$, $p(z) := \frac{\|z\|_Z}{\|z_0\|_Z}$, and $\lambda\colon Z_0 \to \mathbb{K}$ is given by $\lambda(tz_0) = t$ for $t \in \mathbb{K}$. Then the Hahn–Banach theorem gives us some linear map $\Lambda\colon Z \to \mathbb{K}$ with $\Lambda(tz_0) = t$ for all $t \in \mathbb{C}$ and $|\Lambda(z)| \le p(z)$ on Z. In particular, Λ is continuous as a map from Z to $\mathbb{K}$, hence Λ is an element of Z'. Put $z' := \Lambda$. Then this $z' \in Z'$ enjoys $\langle z' \mid z_0\rangle_{Z'\times Z} \neq 0$. We define the desired y' as $y' := z' \circ \pi$, and we observe

$$\langle y' \mid y_0\rangle_{Y'\times Y} = \langle z' \circ \pi \mid y_0\rangle_{Y'\times Y} = \langle z' \mid \pi y_0\rangle_{Z'\times Z} \neq 0.$$

This proves $y' \neq 0 \in Y'$. On the other hand, for all $x \in X$ we have

$$\langle \mathcal{A}^t y' \mid x\rangle_{X'\times X} = \langle y' \mid \mathcal{A}x\rangle_{Y'\times Y} = \langle z' \circ \pi \mid \mathcal{A}x\rangle_{Y'\times Y} = \langle z' \mid \pi\mathcal{A}x\rangle_{Z'\times Z} = \langle z' \mid 0\rangle_{Z'\times Z} = 0 = \langle 0 \mid x\rangle_{X'\times X},$$

and therefore $\mathcal{A}^t y' = 0$ as identity in X', contradicting the injectivity of $\mathcal{A}^t$.

□

Proposition 1.68. *We keep the assumptions of the previous proposition. If $\mathcal{A}\colon X \to Y$ is a topological isomorphism (which means that $\mathcal{A}\colon X \to Y$ is bijective, and $\mathcal{A}$, $\mathcal{A}^{-1}$ are both continuous), then also $\mathcal{A}^t$ is a topological isomorphism, and we have $(\mathcal{A}^t)^{-1} = (\mathcal{A}^{-1})^t$.*

Proof. The proof is split into several steps.

$(\mathcal{A}^{-1})^t$ exists and is continuous: Since $\mathcal{A}^{-1}$ is continuous, there is a continuous transposed operator $(\mathcal{A}^{-1})^t\colon X' \to Y'$ defined by

$$\left\langle (\mathcal{A}^{-1})^t x' \mid y\right\rangle_{Y'\times Y} := \left\langle x' \mid \mathcal{A}^{-1}y\right\rangle_{X'\times X}.$$

$\mathcal{A}^t$ is injective: this holds since $\mathcal{A}$ is surjective, hence its range is dense in Y, and then Proposition 1.67 yields the injectivity of $\mathcal{A}^t$.

$\mathcal{A}^t$ is surjective: to show this, let us be given $x' \in X'$, and we wish to find $y' \in Y'$ with $\mathcal{A}^t y' = x'$. We claim that $y' := (\mathcal{A}^{-1})^t x'$ does the trick: For all $x \in X$ we then have

$$\begin{aligned}\left\langle \mathcal{A}^t y' \mid x\right\rangle_{X'\times X} &= \langle y' \mid \mathcal{A}x\rangle_{Y'\times Y} = \left\langle (\mathcal{A}^{-1})^t x' \mid \mathcal{A}x\right\rangle_{Y'\times Y} = \left\langle x' \mid \mathcal{A}^{-1}\mathcal{A}x\right\rangle_{X'\times X} \\ &= \langle x' \mid x\rangle_{X'\times X},\end{aligned}$$

and this means $\mathcal{A}^t y' = x'$.

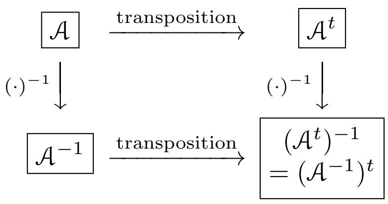
$(\mathcal{A}^t)^{-1} = (\mathcal{A}^{-1})^t$: we need to check that $(\mathcal{A}^{-1})^t\mathcal{A}^t = \mathrm{id}_{Y'}$ and $\mathcal{A}^t(\mathcal{A}^{-1})^t = \mathrm{id}_{X'}$. To this end, we calculate

$$\begin{aligned}\langle y' \mid y\rangle_{Y'\times Y} &= \left\langle y' \mid \mathcal{A}\mathcal{A}^{-1}y\right\rangle_{Y'\times Y} = \left\langle \mathcal{A}^t y' \mid \mathcal{A}^{-1}y\right\rangle_{X'\times X} = \left\langle (\mathcal{A}^{-1})^t\mathcal{A}^t y' \mid y\right\rangle_{Y'\times Y}, \\ \langle x' \mid x\rangle_{X'\times X} &= \left\langle x' \mid \mathcal{A}^{-1}\mathcal{A}x\right\rangle_{X'\times X} = \left\langle (\mathcal{A}^{-1})^t x' \mid \mathcal{A}x\right\rangle_{Y'\times Y} = \left\langle \mathcal{A}^t(\mathcal{A}^{-1})^t x' \mid x\right\rangle_{X'\times X},\end{aligned}$$

from which both identities follow.

□

The identity $(\mathcal{A}^t)^{-1} = (\mathcal{A}^{-1})^t$ can be visualised as a commutative diagram, which tells us that it does not matter which path you take from the top-left corner to the bottom-right corner:



[6] The Hahn–Banach theorem is one of the holiest theorems of Functional Analysis, compare 36.1 in [3]. Its statement is this: Let Z be a vector space over $\mathbb{K}$, p be a seminorm on Z, and Z_0 be a sub-vector space of Z. Suppose a linear map $\lambda\colon Z_0 \to \mathbb{K}$ satisfies $|\lambda(z)| \le p(z)$ for all $z \in Z_0$. Then there is some linear map $\Lambda\colon Z \to \mathbb{K}$ which coincides with λ on Z_0, and $|\Lambda(z)| \le p(z)$ for all $z \in Z$.

Chapter 2

The Fourier Transform

This chapter is following [5]. Another nice presentation can be found in [10], and a further approach (which is quite nice from the functional analytic aspect) is in [7]. Compare also [8] and [9].

2.1 The Fourier Transform Applied to Functions

2.1.1 The Schwartz Function Space $\mathcal{S}(\mathbb{R}^n)$

Definition 2.1 (Schwartz function space $\mathcal{S}(\mathbb{R}^n)$). *The* SCHWARTZ[1] space $\mathcal{S}(\mathbb{R}^n)$ *consists of all those functions* $f \in C^\infty(\mathbb{R}^n)$ *with*

$$p_k(f) := \sum_{|\alpha|+|\beta|\le k} \sup_{x\in\mathbb{R}^n} \left| x^\beta \partial_x^\alpha f(x) \right| < \infty, \qquad \forall k \in \mathbb{N}_0. \tag{2.1}$$

Here we have set, for α, $\beta \in \mathbb{N}_0^n$,

$$x^\beta := x_1^{\beta_1} \cdot x_2^{\beta_2} \cdot \ldots \cdot x_n^{\alpha_n}, \qquad \partial_x^\alpha := \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \frac{\partial^{\alpha_2}}{\partial x_2^{\alpha_2}} \cdots \frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}}.$$

We equip the vector space $\mathcal{S}(\mathbb{R}^n)$ *with the locally convex topology that comes from this countable collection of seminorms* p_k*, and then* $\mathcal{S}(\mathbb{R}^n)$ *turns into a Fréchet space.*

These Schwartz functions are infinitely smooth, they decay at infinity faster than all powers of $|x|^{-1}$, and all their derivatives decay at infinity faster than all powers of $|x|^{-1}$, too.

Lemma 2.2 (Convergence in $\mathcal{S}(\mathbb{R}^n)$). *A sequence* $(\varphi_1, \varphi_2, \dots) \subset \mathcal{S}(\mathbb{R}^n)$ *converges to* $\varphi \in \mathcal{S}$ in the topology of $\mathcal{S}(\mathbb{R}^n)$ *if* $\lim_{j\to\infty} p_k(\varphi_j - \varphi) = 0$ *for all* k*. We write this convergence as*

$$\varphi_j \xrightarrow[j\to\infty]{\mathcal{S}} \varphi.$$

This means (at least) that the sequence $(\varphi_1, \varphi_2, \dots)$ converges to φ uniformly, and all the sequences of derivatives enjoy uniform convergence, too. The convergence in the topology of $\mathcal{S}$ is exceptionally powerful.

It will turn out to be helpful to introduce some notations:

$$D := \frac{1}{\mathrm{i}}\nabla, \qquad đ\xi := \frac{\mathrm{d}\xi}{(2\pi)^n}.$$

Definition 2.3 (Multiplication operator). *For* $\beta \in \mathbb{N}_0^n$*, we define a multiplication operator* M_β *that maps a function* $f\colon \mathbb{R}^n \to \mathbb{C}$ *to a function* $M_\beta f\colon \mathbb{R}^n \to \mathbb{C}$*, defined as*

$$(M_\beta f)(x) := x^\beta f(x), \qquad x \in \mathbb{R}^n.$$

[1] LAURENT SCHWARTZ, 1915–2002, inventor of the distributions (after Sobolev), Fields medallist 1950

Lemma 2.4. *The multiplication operator M_β and the differential operator D^α are continuous maps of the Fréchet space $\mathcal{S}(\mathbb{R}^n)$ into itself.*

Definition 2.5 (Convolution in $L^1(\mathbb{R}^n)$). *For functions f and g from $L^1(\mathbb{R}^n)$, we define their convolution $f * g = (f * g)(x)$ as*

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

Lemma 2.6 (Properties of the Convolution). *If $f, g \in L^1(\mathbb{R}^n)$, then also $f * g \in L^1(\mathbb{R}^n)$, and*

$$\|f * g\|_{L^1(\mathbb{R}^n)} \leq \|f\|_{L^1(\mathbb{R}^n)} \|g\|_{L^1(\mathbb{R}^n)} .$$

Moreover, the convolution product is a bilinear and continuous map of $\mathcal{S}(\mathbb{R}^n) \times \mathcal{S}(\mathbb{R}^n)$ into $\mathcal{S}(\mathbb{R}^n)$.

Proof. The inequality is an exercise in applying Fubini's theorem. Concerning the second claim, we estimate $p_k(f*g)$. Differentiating under the integral sign with respect to x is allowed, hence $f*g \in C^\infty(\mathbb{R}^n)$, and then we wish to understand

$$\left|x^\beta\right| \int_{\mathbb{R}^n_y} |\partial_x^\alpha f(x-y)| \cdot |g(y)| \, \mathrm{d}y$$

for $|\alpha| + |\beta| \leq k$. Now we have

$$\begin{aligned} \left|x^\beta\right| &\leq |x|^{|\beta|} \leq (|x-y| + |y|)^{|\beta|} \leq \Big((1+|x-y|) \cdot (1+|y|)\Big)^{|\beta|} = (1+|x-y|)^{|\beta|} \cdot (1+|y|)^{|\beta|} \\ &\leq C_k \left(\sum_{i=0}^k |x-y|^i\right) \left(\sum_{j=0}^k |y|^j\right), \end{aligned}$$

and therefore (with certainly a new C_k in every line)

$$\begin{aligned} p_k(f*g) &\leq C_k \sum_{|\alpha| \leq k} \sum_{i,j=0}^k \sup_{x \in \mathbb{R}^n} \int_{\mathbb{R}^n_y} \left||x-y|^i \partial_x^\alpha f(x-y)\right| \cdot \left||y|^j g(y)\right| \, \mathrm{d}y \\ &\leq C_k \sum_{|\alpha| \leq k} \sum_{i,j=0}^k \left(\sup_{z \in \mathbb{R}^n} \left||z|^i \partial_z^\alpha f(z)\right|\right) \int_{\mathbb{R}^n_y} \left||y|^i g(y)\right| \, \mathrm{d}y \\ &\leq \dots \\ &\leq C_k p_{2k}(f) \cdot p_{k+n+1}(g), \end{aligned}$$

because of $|z| \leq 1 + |z_1| + \ldots + |z_n|$ and

$$\|u\|_{L^1(\mathbb{R}^n)} = \int_{\mathbb{R}^n_y} \frac{1}{1+|y|^{n+1}} \cdot (1+|y|^{n+1})|u(y)| \, \mathrm{d}y \leq C_n \left(\int_{\mathbb{R}^n_y} \frac{\mathrm{d}y}{1+|y|^{n+1}}\right) \cdot p_{n+1}(u),$$

to be applied to the function $u(y) = |y|^i g(y)$. □

2.1.2 The Fourier Transform on $\mathcal{S}(\mathbb{R}^n)$

Definition 2.7 (Fourier[2] transform on $\mathcal{S}(\mathbb{R}^n)$). *For $f \in \mathcal{S}(\mathbb{R}^n)$, we define $\mathcal{F}f = \hat{f}$ as*

$$(\mathcal{F}f)(\xi) := \int_{\mathbb{R}^n} e^{-\mathrm{i}x\cdot\xi} f(x) \, \mathrm{d}x, \qquad \xi \in \mathbb{R}^n, \quad x \cdot \xi := x_1\xi_1 + \cdots + x_n\xi_n,$$

called the Fourier transform of f.

We quickly check that $\hat{f} \in C^\infty(\mathbb{R}^n)$ for $f \in \mathcal{S}(\mathbb{R}^n)$.

[2] Jean Baptiste Joseph Fourier, 1768–1830

Lemma 2.8. *If $f \in \mathcal{S}(\mathbb{R}^n)$ and $\alpha \in \mathbb{N}_0^n$, then*

$$(\mathcal{F}(D_x^\alpha f))(\xi) = \xi^\alpha \hat{f}(\xi) = (M_\alpha(\mathcal{F}f))(\xi), \tag{2.2}$$

$$D_\xi^\alpha \hat{f}(\xi) = (-1)^{|\alpha|}\widehat{x^\alpha f(x)}(\xi) = (-1)^{|\alpha|}\mathcal{F}(M_\alpha f)(\xi). \tag{2.3}$$

Sketch of proof. The first identity follows from integration by parts, the second from differentiation with respect to a parameter. □

The following commutative diagrams should be self-explanatory:

Combining both equations reveals (observe the switched roles of α and β)

$$\left|\Big(\mathcal{F}(x^\beta D_x^\alpha f(x))\Big)(\xi)\right| = \left|D_\xi^\beta\Big(\xi^\alpha \hat{f}(\xi)\Big)\right|.$$

Proposition 2.9. *The Fourier transform is a continuous map from $\mathcal{S}(\mathbb{R}_x^n)$ into $\mathcal{S}(\mathbb{R}_\xi^n)$ in the sense that for each k there is a constant C_k with*

$$p_k(\hat{f}) \le C_k p_{k+n+1}(f),$$

for each $f \in \mathcal{S}$. Moreover, the Fourier transform preserves the $\mathcal{S}$ convergence,

$$f_j \xrightarrow[j\to\infty]{\mathcal{S}} f \quad\Longrightarrow\quad \hat{f}_j \xrightarrow[j\to\infty]{\mathcal{S}(\mathbb{R}_\xi^n)} \hat{f}.$$

Proof. Fix some k and some α, β with $|\alpha| + |\beta| \le k$. Then

$$\begin{aligned}
\sup_\xi \left|\xi^\alpha D_\xi^\beta \hat{f}(\xi)\right| &\le C_k' \sup_\xi \sum_{|\gamma|+|\delta|\le k} \left|D_\xi^\gamma\left(\xi^\delta \hat{f}(\xi)\right)\right| \\
&= C_k' \sup_\xi \sum_{|\gamma|+|\delta|\le k} \left|\Big(\mathcal{F}(x^\gamma D_x^\delta f(x))\Big)(\xi)\right| \le C_k' \sum_{|\gamma|+|\delta|\le k} \left\|x^\gamma D_x^\delta f(x)\right\|_{L^1(\mathbb{R}_x^n)} \\
&\le C_k p_{k+n+1}(f),
\end{aligned}$$

where we have used once again that $\|u\|_{L^1(\mathbb{R}^n)} \le Cp_{n+1}(u)$ for every $u \in \mathcal{S}(\mathbb{R}^n)$, and also $\|\hat{v}\|_{L^\infty} \le \|v\|_{L^1}$. □

Proposition 2.10. *The Fourier transform has the following property for f and g from $\mathcal{S}(\mathbb{R}^n)$ and all $\xi \in \mathbb{R}^n$:*

$$(f * g)\hat{}\,(\xi) = \hat{f}(\xi)\cdot\hat{g}(\xi).$$

Furthermore, we have for f, $g \in \mathcal{S}(\mathbb{R}^n)$ and all $x \in \mathbb{R}^n$ the identity

$$\int_{\mathbb{R}_\xi^n} e^{\mathrm{i}x\cdot\xi}\hat{f}(\xi)g(\xi)\,\mathrm{d}\xi = \int_{\mathbb{R}_y^n} f(x+y)\hat{g}(y)\,\mathrm{d}y. \tag{2.4}$$

Proof. Note that $f * g \in \mathcal{S}(\mathbb{R}^n)$ by Lemma 2.4, hence $(f*g)\hat{}$ exists, and the first statement is an insightful exercise in using Fubini's theorem. And finally we have

$$\begin{aligned}
\int_{\mathbb{R}_\xi^n} e^{\mathrm{i}x\cdot\xi}\hat{f}(\xi)g(\xi)\,\mathrm{d}\xi &= \int_{\mathbb{R}_\xi^n} e^{\mathrm{i}x\cdot\xi}\left(\int_{\mathbb{R}_y^n} e^{-\mathrm{i}y\cdot\xi}f(y)\,\mathrm{d}y\right)g(\xi)\,\mathrm{d}\xi = \int_{\mathbb{R}_y^n}\left(\int_{\mathbb{R}_\xi^n} e^{\mathrm{i}(x-y)\cdot\xi}g(\xi)\,\mathrm{d}\xi\right)f(y)\,\mathrm{d}y \\
&= \int_{\mathbb{R}_y^n}\hat{g}(y-x)f(y)\,\mathrm{d}y = \int_{\mathbb{R}_y^n}\hat{g}(y)f(x+y)\,\mathrm{d}y.
\end{aligned}$$

□

Again we have a commutative diagram:

$$\begin{array}{ccc} \boxed{(f,g)} & \xrightarrow{\;*\;} & \boxed{f*g} \\ \mathcal{F}\big\downarrow & & \mathcal{F}\big\downarrow \\ \boxed{(\hat f,\hat g)} & \xrightarrow{\;\cdot\;} & \boxed{\begin{array}{l}(f*g)\hat{}\\ =\hat f\cdot\hat g\end{array}} \end{array}$$

Corollary 2.11. *From (2.4) we immediately obtain*

$$\int_{\mathbb{R}^n} \hat f g \,\mathrm{d}x = \int_{\mathbb{R}^n} f\hat g \,\mathrm{d}x, \qquad f,g\in \mathcal{S}(\mathbb{R}^n), \tag{2.5}$$

by setting $x=0$ and renaming variables.

The next example prepares the inversion formula.

Example 2.12. *Take $f=f(x)=\exp(-ax^2)$ on $\mathbb{R}^1$, with $\Re a>0$. Then*

$$\begin{aligned}\partial_\xi \hat f(\xi) &= \partial_\xi \int_{\mathbb{R}_x} e^{-\mathrm{i}x\cdot\xi} f(x)\,\mathrm{d}x = \int_{\mathbb{R}_x} (-\mathrm{i}x)e^{-\mathrm{i}x\cdot\xi} f(x)\,\mathrm{d}x = \frac{-\mathrm{i}}{-2a}\int_{\mathbb{R}_x} e^{-\mathrm{i}x\cdot\xi}\partial_x f(x)\,\mathrm{d}x \\ &= \frac{-1}{2a}\int_{\mathbb{R}_x} e^{-\mathrm{i}x\cdot\xi}(D_x f)(x)\,\mathrm{d}x = -\frac{1}{2a}(D_x f)\hat{}(\xi) = -\frac{\xi}{2a}\hat f(\xi),\end{aligned}$$

and this ordinary differential equation (ODE) has the solution

$$\hat f(\xi) = \exp\left(-\frac{\xi^2}{4a}\right)\hat f(0)$$

with an unknown constant $\hat f(0) =: I$ which can be found by

$$I^2 = \iint_{\mathbb{R}^2} \exp(-a(x^2+y^2))\,\mathrm{d}x\,\mathrm{d}y = \int_{\phi=0}^{2\pi}\int_{r=0}^{\infty}\exp(-ar^2)r\,\mathrm{d}r\,\mathrm{d}\phi = 2\pi\int_{r=0}^{\infty} e^{-ar^2}r\,\mathrm{d}r = \frac{\pi}{a}.$$

The final answer is (with the square root defined in $\mathbb{C}$ using a cut along $(-\infty,0)$)

$$\hat f(\xi) = \frac{\sqrt{\pi}}{\sqrt{a}}\exp\left(-\frac{\xi^2}{4a}\right).$$

Next we take $f=f(x)=\exp(-a|x|^2)$ on $\mathbb{R}^n$ with $\Re a>0$. Then

$$\hat f(\xi) = \int_{\mathbb{R}^n_x} e^{-\mathrm{i}(x_1\xi_1+\cdots+x_n\xi_n)} f(x)\,\mathrm{d}x = \prod_{k=1}^{n}\left(\int_{\mathbb{R}^1_t} e^{-\mathrm{i}t\xi_k}e^{-at^2}\,\mathrm{d}t\right) = \left(\frac{\sqrt{\pi}}{\sqrt{a}}\right)^n \exp\left(-\frac{|\xi|^2}{4a}\right).$$

Proposition 2.13 (Inverse Fourier transform on $\mathcal{S}(\mathbb{R}^n)$). *The Fourier transform is an isomorphism from $\mathcal{S}(\mathbb{R}^n_x)$ onto $\mathcal{S}(\mathbb{R}^n_\xi)$, and the inverse Fourier transform is given by*

$$\varphi(x) = \int_{\mathbb{R}^n_\xi} e^{+\mathrm{i}x\cdot\xi}\hat\varphi(\xi)\,\mathrm{đ}\xi, \qquad x\in\mathbb{R}^n, \quad \hat\varphi\in\mathcal{S}(\mathbb{R}^n). \tag{2.6}$$

Proof. We wish to show that

$$(2\pi)^n\varphi(x) = \int_{\mathbb{R}^n_\xi}\left(\int_{\mathbb{R}^n_y} e^{\mathrm{i}(x-y)\cdot\xi}\varphi(y)\,\mathrm{d}y\right)\mathrm{d}\xi,$$

for all $\varphi\in\mathcal{S}(\mathbb{R}^n)$ and all $x\in\mathbb{R}^n$. However, the integral on the RHS does not converge absolutely, so we cannot swap the integrals.

Pick some $\psi\in\mathcal{S}(\mathbb{R}^n)$. Then we know already from (2.4) that

$$\int_{\mathbb{R}^n_\xi} e^{\mathrm{i}x\cdot\xi}\hat\varphi(\xi)\psi(\xi)\,\mathrm{d}\xi = \int_{\mathbb{R}^n_y}\varphi(x+y)\hat\psi(y)\,\mathrm{d}y.$$

In this equation, we choose ψ as $\psi(\xi) = \exp(-\varepsilon^2|\xi|^2)$, and then we have

$$\hat{\psi}(y) = \frac{\pi^{n/2}}{\varepsilon^n} \exp\left(-\frac{|y|^2}{4\varepsilon^2}\right).$$

As next step, we substitute $z := \varepsilon^{-1}y$ and find

$$\begin{aligned}\int_{\mathbb{R}^n_\xi} e^{\mathrm{i}x\cdot\xi}\hat{\varphi}(\xi)\exp(-\varepsilon^2|\xi|^2)\,\mathrm{d}\xi &= \int_{\mathbb{R}^n_y} \varphi(x+y)\frac{\pi^{n/2}}{\varepsilon^n}\exp\left(-\frac{|y|^2}{4\varepsilon^2}\right)\mathrm{d}y \\ &= \pi^{n/2}\int_{\mathbb{R}^n_z}\varphi(x+\varepsilon z)\exp\left(-\frac{|z|^2}{4}\right)\mathrm{d}z.\end{aligned}$$

We send ε to $+0$ and apply the Convergence theorem of Lebesgue, which yields

$$\int_{\mathbb{R}^n_\xi} e^{\mathrm{i}x\cdot\xi}\hat{\varphi}(\xi)\,\mathrm{d}\xi = \pi^{n/2}\varphi(x)\int_{\mathbb{R}^n_y}\exp\left(-\frac{|z|^2}{4}\right)\mathrm{d}z = \pi^{n/2}\varphi(x)\,(4\pi)^{n/2} = (2\pi)^n\varphi(x).$$

This was our goal. □

We have only shown that the Fourier transform is an isomorphism between algebraic vector spaces. However, Proposition 2.9 enables us to quickly prove that the Fourier transform is also a topological isomorphism (exercise).

We draw some conclusions:

Remark 2.14. *For any Schwartz function w, we have*

$$\begin{aligned}\widehat{w(-x)}(\xi) &= \int_{\mathbb{R}^n_x} e^{-\mathrm{i}x\cdot\xi}w(-x)\,\mathrm{d}x = \int_{\mathbb{R}^n_y} e^{\mathrm{i}y\cdot\xi}w(y)\,\mathrm{d}y = \hat{w}(-\xi),\\ (\mathcal{F}w)(\xi) &= (2\pi)^n(\mathcal{F}^{-1}w)(-\xi),\\ (\mathcal{F}\mathcal{F}w)(x) &= (2\pi)^n\mathcal{F}\Big((\mathcal{F}^{-1}w)(-\xi)\Big)(x) = (2\pi)^n(\mathcal{F}\mathcal{F}^{-1}w)(-x) = (2\pi)^n w(-x).\end{aligned}$$

Now choose Schwartz functions f and g. Then Schwartz functions u and v exist with $f = \hat{u}$ and $g = \hat{v}$, and we also have

$$\hat{f}(\xi) = (\mathcal{F}\mathcal{F}u)(\xi) = (2\pi)^n u(-\xi), \qquad \hat{g}(\xi) = (\mathcal{F}\mathcal{F}v)(\xi) = (2\pi)^n v(-\xi).$$

Then we conclude, from Proposition 2.10, that

$$\begin{aligned}\widehat{fg}(\xi) &= (\mathcal{F}(fg))\,(\xi) = \Big(\mathcal{F}(\hat{u}\hat{v})\Big)(\xi) = \Big(\mathcal{F}\Big(\mathcal{F}(u*v)\Big)\Big)(\xi) = (2\pi)^n(u*v)(-\xi)\\ &= (2\pi)^n\int_{\mathbb{R}^n_\eta} u(-\xi-\eta)v(\eta)\,\mathrm{d}\eta = (2\pi)^n\int_{\mathbb{R}^n_\zeta} u(-\xi+\zeta)v(-\zeta)\,\mathrm{d}\zeta\\ &= (2\pi)^{-n}\int_{\mathbb{R}^n_\zeta}(2\pi)^n u(-(\xi-\zeta))\cdot(2\pi)^n v(-\zeta)\,\mathrm{d}\zeta\\ &= (2\pi)^{-n}\int_{\mathbb{R}^n_\zeta}\hat{f}(\xi-\zeta)\cdot\hat{g}(\zeta)\,\mathrm{d}\zeta\\ &= \Big(\hat{f}*\hat{g}\Big)(\xi),\end{aligned}$$

with $*$ as the convolution operator in the ξ–world (note the $đ\eta$ instead of $\mathrm{d}\eta$):

$$(\hat{f}*\hat{g})(\xi) := \int_{\mathbb{R}^n_\eta}\hat{f}(\eta)\cdot\hat{g}(\xi-\eta)\,đ\eta. \tag{2.7}$$

We draw an intermediate summary: the difference between the Fourier transform $\mathcal{F}$ and the inverse transform $\mathcal{F}^{-1}$ are the exchange of $\exp(+\mathrm{i}x\cdot\xi)$ against $\exp(-\mathrm{i}x\cdot\xi)$, and an additional factor $(2\pi)^{-n}$.

2.1.3 The Fourier Transform on $L^2(\mathbb{R}^n)$

Proposition 2.15. *The Fourier transform preserves the L^2 scalar product (up to a factor $(2\pi)^n$) in the sense of*

$$\forall f, g \in \mathcal{S}(\mathbb{R}^n): \qquad \int_{\mathbb{R}^n_x} f(x)\overline{g(x)}\,\mathrm{d}x = \int_{\mathbb{R}^n_\xi} \hat{f}(\xi)\overline{\hat{g}(\xi)}\,\text{đ}\xi,$$

Proof. We begin with

$$\int_{\mathbb{R}^n_\xi} \hat{f}(\xi)\overline{\hat{g}(\xi)}\,\text{đ}\xi = (2\pi)^{-n}\int_{\mathbb{R}^n_\xi} \hat{f}(\xi)\overline{\hat{g}(\xi)}\,\mathrm{d}\xi = (2\pi)^{-n}\int_{\mathbb{R}^n_\xi} f(x)\widehat{\overline{\hat{g}}}(x)\,\mathrm{d}x,$$

from (2.5). Next we calculate

$$\overline{\hat{g}(\xi)} = \int_{\mathbb{R}^n} e^{+\mathrm{i}x\cdot\xi}\overline{g(x)}\,\mathrm{d}x = (2\pi)^n \mathcal{F}^{-1}_{x\to\xi}\Big(\overline{g(x)}\Big)(\xi),$$

and therefore $\widehat{\overline{\hat{g}}} = (2\pi)^n\overline{g}$. □

This property yields the PARSEVAL[3] *identity*:

$$\|f\|_{L^2(\mathbb{R}^n_x)} = \left\|\hat{f}\right\|_{L^2(\mathbb{R}^n_\xi)} := \left(\int_{\mathbb{R}^n_\xi} |\hat{f}(\xi)|^2\,\text{đ}\xi\right)^{1/2}, \qquad \forall\, f \in \mathcal{S}(\mathbb{R}^n).$$

The space $L^2(\mathbb{R}^n)$ is of great physical importance because it has a scalar product. The definition of $\mathcal{F}$ on $L^2(\mathbb{R}^n)$ will be made possible by $\mathcal{S}(\mathbb{R}^n)$ being *sequentially dense* in $L^2(\mathbb{R}^n)$: for each $f \in L^2(\mathbb{R}^n)$, there is a sequence $(f_1, f_2, \dots) \subset \mathcal{S}(\mathbb{R}^n)$ with $\lim_{j\to\infty}\|f_j - f\|_{L^2(\mathbb{R}^n)} = 0$. And because $\mathcal{S}(\mathbb{R}^n)$ is a Fréchet space (and $L^2(\mathbb{R}^n)$ is a metric space), we are happy because in this setting *sequentially dense* means the same as *topologically dense* (the topological closure of the smaller set is the bigger set).

Definition 2.16 (Fourier transform on $L^2(\mathbb{R}^n)$). *For $f \in L^2(\mathbb{R}^n)$, let $(f_1, f_2, \dots) \subset \mathcal{S}(\mathbb{R}^n)$ be a sequence approximating f. Then we define*

$$\hat{f}(\xi) := \lim_{j\to\infty} \hat{f}_j(\xi).$$

This limit is independent of the choice of the sequence $(f_1, f_2, \dots)$. The convergence of the sequence $(\hat{f}_1, \hat{f}_2, \dots)$ in the norm of $L^2(\mathbb{R}^n_\xi)$ follows from the Parseval formula.

2.2 The Fourier Transform Applied to Temperate Distributions

2.2.1 What are Distributions All About ?

We wish to define generalised functions (called *temperate distributions*) with the following properties:

- every "reasonable function" can be understood as a distribution, where "reasonable" could mean measurability, absence of strong poles, and at most polynomial growth for $|x| \to \infty$;
- every temperate distribution can be differentiated as often as we want, and the result is again a temperate distribution;
- we have a meaning of a temperate distribution being the limit of a sequence of temperate distributions;
- we have various operations that we can let act upon temperate distributions (adding them, multiplying them by numbers and by smooth functions, computing their Fourier transforms, determining their convolutions);

[3] MARC–ANTOINE PARSEVAL DES CHÊNES, 1755–1836

- these operations are continuous in a sense of appropriately chosen topological vector spaces, and they coincide with the previously defined operations if the temperate distributions coincide with smooth functions of moderate growth.

Unfortunately, we have a price to pay: it will not be possible to define products of distributions. In that sense, the distribution theory will always be a linear theory.

2.2.2 The Space $\mathcal{S}'(\mathbb{R}^n)$ of Temperate Distributions

Definition 2.17 (Temperate distribution). *A map* $T\colon \mathcal{S}(\mathbb{R}^n) \to \mathbb{C}$ *is called* temperate distribution[4] *if* T *is linear and continuous. The set of all temperate distributions is called* $\mathcal{S}'(\mathbb{R}^n)$.

Example 2.18 (Regular distribution). *A temperate distribution is provided by a function* $f = f(x)$ *with the property that*

$$x \mapsto \frac{f(x)}{1+|x|^N} \in L^1(\mathbb{R}^n)$$

for some $N \in \mathbb{N}$, *which will generate a distribution* T_f *via*

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

Such a distribution T_f *is called* regular distribution.

Every function with moderate growth can be seen as a temperate distribution.

Example 2.19 (Dirac's Delta distribution[5]). *Let* $x_0 \in \mathbb{R}^n$. *We define a map* T *via* $T(\varphi) = \langle T \mid \varphi\rangle_{\mathcal{S}'\times\mathcal{S}} := \varphi(x_0)$.

Finally we mention that a function with fast growth for $|x| \to \infty$ can be a temperate distribution as well:

Lemma 2.20. *The function* $f = f(x) = \exp(x)\cdot\cos(\exp(x)) = \partial_x(\sin(\exp(x)))$ *generates a distribution* $T_f \in \mathcal{S}'(\mathbb{R}^1)$ *in the sense of*

$$\langle T_f \mid \varphi\rangle_{\mathcal{S}'\times\mathcal{S}} := \lim_{R\to\infty} \int_{|x|\le R} f(x)\varphi(x)\,\mathrm{d}x.$$

Proof. This follows from

$$\int_{|x|\le R} f(x)\varphi(x)\,\mathrm{d}x = \sin(\exp(x))\cdot\varphi(x)\Big|_{x=-R}^{x=R} - \int_{|x|\le R} \sin(\exp(x))\cdot\varphi'(x)\,\mathrm{d}x,$$

hence

$$\langle T_f \mid \varphi\rangle_{\mathcal{S}'\times\mathcal{S}} = -\int_{\mathbb{R}^1} \sin(\exp(x))\cdot\varphi'(x)\,\mathrm{d}x.$$

□

The reason is that there are two types of integrals in $\mathbb{R}^n$ that have to be distinguished. One is the usual Lebesgue integral as in the definition of $L^1(\mathbb{R}^n)$, the other is $\lim_{R\to\infty}\int_{|x|<R}\dots\,\mathrm{d}x$. These are not the same, as can be seen from the function $g(x) = \frac{\sin(x)}{x}$ which is integrable in the second sense, but not in the first (recall that if g were a member of $L^1(\mathbb{R}^n)$, then so would $|g|$).

[4] Perhaps you will read sometimes the expression *tempered* distribution, even written by a native speaker of English. Such authors have no taste.

[5] Paul Adrien Maurice Dirac, 1902–1984

We lift our investigations to a more abstract level.

Definition 2.21 (Canonical map of $\mathcal{S}(\mathbb{R}^n)$ into $\mathcal{S}'(\mathbb{R}^n)$). *We have a map*[6] $\iota\colon \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}'(\mathbb{R}^n)$ *that interprets a Schwartz function* ψ *as a temperate distribution* $\iota\psi$*, defined as*

$$\langle \iota\psi \mid \varphi\rangle_{\mathcal{S}'\times\mathcal{S}} := \int_{\mathbb{R}^n} \psi(x)\varphi(x)\,\mathrm{d}x, \qquad \forall\varphi \in \mathcal{S}(\mathbb{R}^n).$$

From now on, we often write ιf instead of T_f. Both notations mean the same thing: the Schwartz function f is being construed as a temperate distribution.

Lemma 2.22. *This map* ι *is continuous and injective.*

Proof. ι **is continuous:** We appeal to Proposition 1.59. Let p_χ be any seminorm on $\mathcal{S}'(\mathbb{R}^n)$. Then there is some $\chi \in \mathcal{S}(\mathbb{R}^n)$ with

$$p_\chi(\iota\psi) = \left|\langle \iota\psi \mid \chi\rangle_{\mathcal{S}'\times\mathcal{S}}\right| \le \int_{\mathbb{R}^n} |\psi(x)||\chi(x)|\,\mathrm{d}x \le \|\psi\|_{L^\infty}\|\chi\|_{L^1} = \|\chi\|_{L^1}\, p_0(\psi),$$

with p_0 being the seminorm on $\mathcal{S}(\mathbb{R}^n)$ as defined in (2.1).

ι **is injective:** Suppose $\iota\psi = 0 \in \mathcal{S}'$, which means $\langle \iota\psi \mid \varphi\rangle = 0$ for all $\varphi \in \mathcal{S}(\mathbb{R}^n)$. Now choose $\varphi(x) = \overline{\psi(x)}$, the complex conjugate. This gives $\|\psi\|_{L^2}^2 = 0$, hence $\psi = 0 \in \mathcal{S}(\mathbb{R}^n)$.

□

We have the following operators:

$\partial^\alpha\colon \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$**:** continuous, by Lemma 2.4,

$(\partial^\alpha)^t\colon \mathcal{S}'(\mathbb{R}^n) \to \mathcal{S}'(\mathbb{R}^n)$**:** defined in Definition 1.65. It is continuous, by Proposition 1.66

$\iota\colon \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}'(\mathbb{R}^n)$ continuous and injective by Lemma 2.22, with $\mathrm{img}(\iota)$ dense in $\mathcal{S}'(\mathbb{R}^n)$ (by Lemma 2.37)

2.2.3 An Operation in $\mathcal{S}'(\mathbb{R}^n)$: Taking Derivatives

For sake of clarity, we begin with the **pedestrian approach**.

We observe: if $\varphi \in \mathcal{S}(\mathbb{R}^n)$ and $f \in C_b^\infty(\mathbb{R}^n)$ (which shall mean that the smooth function f and all its derivatives are bounded on $\mathbb{R}^n$) then

$$\int_{\mathbb{R}^n} (\partial_x^\alpha f(x))\varphi(x)\,\mathrm{d}x = (-1)^{|\alpha|}\int_{\mathbb{R}^n} f(x)(\partial_x^\alpha\varphi(x))\,\mathrm{d}x,$$

which can be shown by repeated integration by parts. Using the concept of regular distributions, we write this identity as

$$\langle T_{\partial_x^\alpha f} \mid \varphi\rangle_{\mathcal{S}'\times\mathcal{S}} = (-1)^{|\alpha|}\langle T_f \mid \partial_x^\alpha\varphi\rangle_{\mathcal{S}'\times\mathcal{S}}, \qquad \forall\varphi\in\mathcal{S}(\mathbb{R}^n).$$

Our goal is now to define a derivative $\partial_x^\alpha T$ for every temperate distribution $T \in \mathcal{S}'(\mathbb{R}^n)$. The requirement is that in case T is a regular distribution (hence being generated by a smooth function f in the sense of $T = T_f$), then

$$\partial_x^\alpha T_f \overset{!}{=} T_{\partial_x^\alpha f}. \tag{2.8}$$

Definition 2.23 (Derivative of a distribution). *If* $T \in \mathcal{S}'(\mathbb{R}^n)$ *and* $\alpha \in \mathbb{N}_0^n$ *then* $\partial_x^\alpha T$ *is defined as*

$$\langle \partial_x^\alpha T \mid \varphi\rangle_{\mathcal{S}'\times\mathcal{S}} := (-1)^{|\alpha|}\langle T \mid \partial_x^\alpha\varphi\rangle_{\mathcal{S}'\times\mathcal{S}}, \qquad \forall\varphi\in\mathcal{S}(\mathbb{R}^n).$$

Lemma 2.24. *This* $\partial_x^\alpha T$ *is a member of* $\mathcal{S}'(\mathbb{R}^n)$.

Proof. Exercise. □

[6] pronounced *iota*

If we identify a function $f \in L^1(\mathbb{R}^n)$ and its associated regular distribution $T_f \in \mathcal{S}'(\mathbb{R}^n)$, then we see that each function $f \in L^1(\mathbb{R}^n)$ will possess a derivative $\partial_x^\alpha f$, which then usually will no longer be a function from $L^1(\mathbb{R}^n)$, but a distribution.

Lemma 2.25. *If $T \in \mathcal{S}'(\mathbb{R}^n)$ and $1 \leq j, k \leq n$ then*

$$\frac{\partial}{\partial x_k}\frac{\partial}{\partial x_j}T = \frac{\partial}{\partial x_j}\frac{\partial}{\partial x_k}T.$$

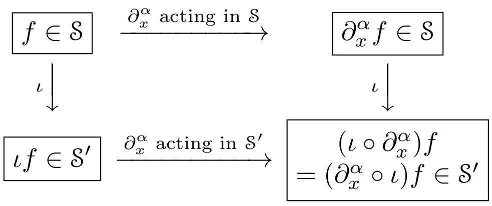
Proof. Exercise. □

After the pedestrian approach, here comes the **academic approach**.

The identity (2.8) can be rewritten as

$$\partial_x^\alpha \circ \iota \stackrel{!}{=} \iota \circ \partial_x^\alpha. \tag{2.9}$$

Beware that we have abused notation here: on the LHS, the ∂_x^α operates on a distribution from $\mathcal{S}'$, but on the RHS, the operator ∂_x^α operates on a function from $\mathcal{S}$. It has been a physicist's idea to use the same notation for two different operations. Next we wish to understand how ∂_x^α (when applied to distributions) relates to the transposed operator $(\partial_x^\alpha)^t$ (which is always being applied to distributions).



Take some $f \in \mathcal{S}(\mathbb{R}^n)$. Then we have, for all Schwartz functions φ,

$$\begin{aligned}
\langle \partial_x^\alpha(\iota f) \mid \varphi\rangle_{\mathcal{S}'\times\mathcal{S}} &= \langle \iota\partial_x^\alpha f \mid \varphi\rangle_{\mathcal{S}'\times\mathcal{S}} && \text{by (2.9)}\\
&= \int_{\mathbb{R}^n}\Big(\partial_x^\alpha f(x)\Big)\varphi(x)\,\mathrm{d}x && \text{Definition of } \iota\\
&= (-1)^{|\alpha|}\int_{\mathbb{R}^n} f(x)\Big(\partial_x^\alpha\varphi(x)\Big)\,\mathrm{d}x && \text{integration by parts}\\
&= (-1)^{|\alpha|}\,\langle \iota f \mid \partial_x^\alpha\varphi\rangle_{\mathcal{S}'\times\mathcal{S}} && \text{definition of } \iota\\
&= (-1)^{|\alpha|}\,\big\langle (\partial_x^\alpha)^t(\iota f) \mid \varphi\big\rangle_{\mathcal{S}'\times\mathcal{S}} && \text{Definition 1.65 of } (\partial_x^\alpha)^t,
\end{aligned}$$

and then the conclusion is

$$\forall T \in \operatorname{img}(\iota): \qquad \partial_x^\alpha T = (-1)^{|\alpha|}(\partial_x^\alpha)^t T.$$

What is this observation good for ? Well, $(\partial_x^\alpha)^t$ is continuous (by Proposition 1.66), and since we also wish ∂_x^α to be continuous as a map from the distribution space $\mathcal{S}'$ into itself, we *have to* define the distributional derivative ∂_x^α as in Definition 2.23, since $\operatorname{img}(\iota)$ is dense in $\mathcal{S}'$.

2.2.4 An Operation in $\mathcal{S}'(\mathbb{R}^n)$: Multiplying by Smooth Temperate Functions

Again, we present the **pedestrian approach** first.

Definition 2.26 (**Multiplying T by a smooth function**)**.** *Let $T \in \mathcal{S}'(\mathbb{R}^n)$ and $a \in C^\infty(\mathbb{R}^n)$ with the property that $a(x)$ and all its derivatives $\partial_x^\alpha a(x)$ grow at most polynomially for $|x| \to \infty$. Then we define aT via*

$$\langle aT \mid \varphi\rangle_{\mathcal{S}'\times\mathcal{S}} := \langle T \mid a\varphi\rangle_{\mathcal{S}'\times\mathcal{S}}\,, \qquad \forall \varphi \in \mathcal{S}(\mathbb{R}^n).$$

Lemma 2.27. *This expression aT is indeed a member of $\mathcal{S}'(\mathbb{R}^n)$.*

Proof. Exercise. □

Lemma 2.28. *If $a = a(x)$ is as above, and $T \in \mathcal{S}'(\mathbb{R}^n)$, then*

$$\frac{\partial}{\partial x_j}(aT) = \frac{\partial a}{\partial x_j} T + a \frac{\partial T}{\partial x_j}$$

as an identity in the space $\mathcal{S}'(\mathbb{R}^n)$.

Proof. Exercise. □

And again, we also show the **academic approach**.

We have an operator M_a that multiplies a function $\varphi \in \mathcal{S}(\mathbb{R}^n)$ by a smooth function $a = a(x)$ which has the above mentioned properties. Then M_a is a continuous map of $\mathcal{S}$ into itself. Definition 2.26 then defines an operator M_a on the distribution space $\mathcal{S}'$, and it can be shown that this operator equals $(M_a)^t$ as defined in Definition 1.65.

2.2.5 An Operation in $\mathcal{S}'(\mathbb{R}^n)$: Taking Fourier Transforms

We bring (2.5) in the form

$$\langle T_{\hat{\varphi}} \mid \psi \rangle_{\mathcal{S}' \times \mathcal{S}} = \left\langle T_\varphi \mid \hat{\psi} \right\rangle_{\mathcal{S}' \times \mathcal{S}}, \qquad \forall \varphi \in \mathcal{S}, \quad \forall \psi \in \mathcal{S},$$

which inspires us to define the Fourier transform $\hat{T}$ for some $T \in \mathcal{S}'(\mathbb{R}^n)$ like this:

Definition 2.29 (Fourier transform in $\mathcal{S}'$). *For $T \in \mathcal{S}'(\mathbb{R}^n)$, we define $\hat{T} \in \mathcal{S}'(\mathbb{R}^n)$ by the identity*

$$\left\langle \hat{T} \mid \varphi \right\rangle_{\mathcal{S}' \times \mathcal{S}} := \langle T \mid \hat{\varphi} \rangle_{\mathcal{S}' \times \mathcal{S}}, \qquad \forall \varphi \in \mathcal{S}(\mathbb{R}^n).$$

Lemma 2.30. *This map*

$$T \mapsto \hat{T}, \qquad \mathcal{S}' \to \mathcal{S}',$$

is the transposed map (as defined in Definition 1.65) to the Fourier transform

$$\mathcal{F} \colon \varphi \mapsto \hat{\varphi}, \qquad \mathcal{S} \to \mathcal{S}.$$

Proof. This is immediate from

$$\langle \mathcal{F}^t T \mid \varphi \rangle_{\mathcal{S}' \times \mathcal{S}} = \langle T \mid \mathcal{F}\varphi \rangle_{\mathcal{S}' \times \mathcal{S}} = \langle T \mid \hat{\varphi} \rangle_{\mathcal{S}' \times \mathcal{S}} = \left\langle \hat{T} \mid \varphi \right\rangle_{\mathcal{S}' \times \mathcal{S}},$$

valid for all $\varphi \in \mathcal{S}(\mathbb{R}^n)$. □

Proposition 1.68 is waiting for an application, and here it goes:

Corollary 2.31. *The Fourier transform map $\mathcal{F}^t \colon \mathcal{S}' \to \mathcal{S}'$ is continuous, and it is an isomorphism.*

Example 2.32. *What is $\hat{\delta}$?*

$$\left\langle \hat{\delta} \mid \varphi \right\rangle_{\mathcal{S}' \times \mathcal{S}} := \langle \delta \mid \hat{\varphi} \rangle_{\mathcal{S}' \times \mathcal{S}} = \hat{\varphi}(0) = \int_{\mathbb{R}^n} e^{-\mathrm{i}0 \cdot x} \cdot \varphi(x)\, \mathrm{d}x = \langle 1 \mid \varphi \rangle_{\mathcal{S}' \times \mathcal{S}}.$$

Answer: the Fourier transform of Dirac's delta is that function which is one everywhere.

Lemma 2.33. *The following identities are valid for every $T \in \mathcal{S}'$:*

$$\mathcal{F}^t(D^\alpha T) = M_\alpha(\mathcal{F}^t T),$$
$$\mathcal{F}^t(M_\alpha T) = (-1)^{|\alpha|} D^\alpha(\mathcal{F}^t T),$$

with the multiplication operators M_α as in Definition 2.3.

Proof. The first identity is shown like this:

$\left\langle \mathcal{F}^t(D^\alpha T) \mid \varphi \right\rangle_{\mathcal{S}'\times\mathcal{S}} = \langle D^\alpha T \mid \mathcal{F}\varphi \rangle_{\mathcal{S}'\times\mathcal{S}}$	Definition 2.29
$= (-1)^{\lvert\alpha\rvert} \langle T \mid D^\alpha(\mathcal{F}\varphi)\rangle_{\mathcal{S}'\times\mathcal{S}}$	Definition 2.23
$= (-1)^{\lvert\alpha\rvert}(-1)^{\lvert\alpha\rvert} \langle T \mid \mathcal{F}(M_\alpha\varphi)\rangle_{\mathcal{S}'\times\mathcal{S}}$	Lemma 2.8
$= \left\langle \mathcal{F}^t T \mid M_\alpha\varphi \right\rangle_{\mathcal{S}'\times\mathcal{S}}$	Definition 2.29
$= \left\langle M_\alpha(\mathcal{F}^t T) \mid \varphi \right\rangle_{\mathcal{S}'\times\mathcal{S}}$	Definition 2.26.

And for the second identity, the reasoning is similar:

$$\begin{aligned}\left\langle \mathcal{F}^t(M_\alpha T) \mid \varphi \right\rangle_{\mathcal{S}'\times\mathcal{S}} &= \langle M_\alpha T \mid \mathcal{F}\varphi \rangle_{\mathcal{S}'\times\mathcal{S}} = \langle T \mid M_\alpha(\mathcal{F}\varphi)\rangle_{\mathcal{S}'\times\mathcal{S}} = \langle T \mid \mathcal{F}(D^\alpha\varphi)\rangle_{\mathcal{S}'\times\mathcal{S}} \\ &= \left\langle \mathcal{F}^t T \mid D^\alpha\varphi \right\rangle_{\mathcal{S}'\times\mathcal{S}} = (-1)^{\lvert\alpha\rvert} \left\langle D^\alpha(\mathcal{F}^t T) \mid \varphi \right\rangle_{\mathcal{S}'\times\mathcal{S}}.\end{aligned}$$

Now remember that $\varphi \in \mathcal{S}$ can be chosen arbitrarily. □

The identities here are very similar to those of Lemma 2.8, the only difference is that $\mathcal{F}$ has become $\mathcal{F}^t$, and now it we have equations in $\mathcal{S}'$ instead of $\mathcal{S}$ before.

Example: *Let $T = 1$. What is $\mathcal{F}^t T$?*

We calculate like this:

$$\begin{aligned}\left\langle \mathcal{F}^t 1 \mid \varphi \right\rangle_{\mathcal{S}'\times\mathcal{S}} &= \langle 1 \mid \hat{\varphi} \rangle_{\mathcal{S}'\times\mathcal{S}} = \int_{\mathbb{R}^n_\xi} \hat{\varphi}(\xi)\,\mathrm{d}\xi = (2\pi)^n \int_{\mathbb{R}^n_\xi} e^{\mathrm{i}0\xi}\hat{\varphi}(\xi)\,\text{đ}\xi = (2\pi)^n \left(\mathcal{F}^{-1}\{\varphi\}\right)(x=0) \\ &= (2\pi)^n\varphi(0) = (2\pi)^n \langle \delta \mid \varphi \rangle_{\mathcal{S}'\times\mathcal{S}},\end{aligned}$$

and therefore $\mathcal{F}^t 1 = (2\pi)^n\delta$.

Example: *In $\mathbb{R}^1$, what is $\mathcal{F}^t\{x^2\}$?*

Observe that $x^2 = M_{x^2}1$, hence the previous lemma yields

$$\mathcal{F}^t(M_{x^2}1) = (-1)^2 D_x^2(\mathcal{F}^t 1) = +(-\mathrm{i}\partial_x)^2(\mathcal{F}^t 1) = -\partial_x^2((2\pi)^1\delta) = -2\pi\delta''.$$

2.2.6 An Operation in $\mathcal{S}'(\mathbb{R}^n)$: Substituting the Variable

What is the reason why we have $\delta(ax) = \frac{1}{a}\delta(x)$ for $x \in \mathbb{R}^1$ and $a > 0$?

We recall that we have a sequence $(f_1, f_2, \dots) \subset \mathcal{S}(\mathbb{R}^1)$ that approximates δ in the sense of

$$T_{f_j} \xrightarrow[j\to\infty]{\mathcal{S}'} \delta.$$

These functions f_j satisfy $f_j(x) \geq 0$ for all j and x, and

$$\lim_{j\to\infty} f_j(x) = \begin{cases} 0 & : x \neq 0, \\ +\infty & : x = 0, \end{cases} \qquad \int_{-\infty}^{\infty} f_j(x)\,\mathrm{d}x = 1.$$

The convergence property then means $\lim_{j\to\infty}\int_{-\infty}^{\infty} f_j(x)\varphi(x)\,\mathrm{d}x = \varphi(0)$, for all $\varphi \in \mathcal{S}(\mathbb{R}^1)$.

The calculation

$$\int_{-\infty}^{\infty} f_j(ax)\varphi(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} f_j(z)\varphi\left(\frac{z}{a}\right)\frac{1}{a}\,\mathrm{d}z \xrightarrow[j\to\infty]{\mathbb{C}} \varphi(0)\cdot\frac{1}{a} = \frac{1}{a}\langle \delta \mid \varphi \rangle_{\mathcal{S}'\times\mathcal{S}}$$

then motivates our approach.

Definition 2.34 (Substitution inside a distribution). *Let $A \in \mathbb{R}^{n\times n}$ be an invertible matrix, and $T \in \mathcal{S}'(\mathbb{R}^n)$. Then we define*

$$\langle T(Ax) \mid \varphi(x)\rangle_{\mathcal{S}'(\mathbb{R}^n_x)\times\mathcal{S}(\mathbb{R}^n_x)} := \frac{1}{\lvert\det A\rvert}\left\langle T(x) \mid \varphi(A^{-1}x)\right\rangle_{\mathcal{S}'(\mathbb{R}^n_x)\times\mathcal{S}(\mathbb{R}^n_x)}.$$

Example 2.35. *If the matrix A describes a scaling by the factor λ, then*

$$\langle T(\lambda x) \mid \varphi(x)\rangle_{\mathcal{S}'\times\mathcal{S}} = \frac{1}{|\lambda|^n}\left\langle T(x) \mid \varphi\left(\frac{x}{\lambda}\right)\right\rangle_{\mathcal{S}'\times\mathcal{S}}.$$

Example 2.36. *If A describes a rotation or reflection, then A is an orthogonal matrix, hence*

$$\langle T(Ax) \mid \varphi(x)\rangle_{\mathcal{S}'\times\mathcal{S}} = \langle T(x) \mid \varphi(A^\top x)\rangle_{\mathcal{S}'\times\mathcal{S}}.$$

2.2.7 Surprising Properties of Distributions

Why are mathematical distributions called distributions, by the way ? In common language, to distribute something means to dispense it, to scatter it, to disseminate it. If you have 17 units of some stuff, and you scatter it somehow, then you can describe the density of that stuff by an $L^1(\mathbb{R}^3)$ function $f = f(x)$ with $f(x) \geq 0$ for obvious reasons, and $\int_{\mathbb{R}^3} f(x)\,\mathrm{d}x = 17$. On the other hand, if you decide to concentrate all 17 units of that stuff at the point called x_0, then you describe this situation by the mathematical term $17\delta(x - x_0)$.

Mathematical distributions have been invented to generalise both situations. In particular, each L^1 function can be construed as a distribution.

It comes as a surprise that although functions $f \in L^1(\mathbb{R}^n)$ usually do not possess derivatives in the usual sense of limits of quotients of differences, they do in the distributional sense: to each $f \in L^1(\mathbb{R}^n)$, there is a regular distribution $T_f \in \mathcal{S}'(\mathbb{R}^n)$, and this distribution always possesses any derivative $\partial_x^\alpha T_f$, although f does not.

Another surprise: ∂_x is a continuous operator from $\mathcal{S}'$ into $\mathcal{S}'$. This means that if a sequence $(f_1, f_2, \dots)$ of functions converges in the topology of $\mathcal{S}'$ to some limit f, then also the sequence of derivatives $(f_1', f_2', f_3', \dots)$ has the limit f' in the sense of $\mathcal{S}'$. Why is that surprising ?

Well, take $f_j(x) = \frac{1}{j}\sin(jx)$ for $x \in \mathbb{R}^1$. Then we have

$$f_j(x) \xrightarrow[j\to\infty]{\text{uniformly}} 0,$$

which has the immediate consequence

$$f_j \xrightarrow[j\to\infty]{\mathcal{S}'(\mathbb{R})} 0, \qquad \text{and therefore} \qquad f_j' \xrightarrow[j\to\infty]{\mathcal{S}'(\mathbb{R})} 0,$$

but this sequence $f_j'(x)$ is obviously the sequence

$$\cos(x), \cos(2x), \cos(3x), \cos(4x), \dots$$

and the claim means that this sequence of ever faster oscillating cosine functions converges to the zero function.

It gets even more mysterious when we choose the function $g_j(x) = \frac{1}{j}\sin(j^3 x)$, which again converges (even uniformly) to the zero function. But their derivatives are $g_j'(x) = j^2\cos(j^3 x)$, and now even the amplitude explodes for $j \to \infty$, but the sequence g_j' still approaches the zero function.

The *advantage* of the concept of distributions is that sequences that converge in $\mathcal{S}'$ can be differentiated as often as we wish. The *disadvantage* is that sometimes we do not properly visualise what this limit actually means.

2.2.8 Duality Magic

The purpose of this part is to show that the natural embedding of $\mathcal{S}(\mathbb{R}^n)$ into $\mathcal{S}'(\mathbb{R}^n)$ has a dense image, by means of excessive use of duality arguments.

We recall the map $\iota\colon \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}'(\mathbb{R}^n)$ that interprets a Schwartz function ψ as a distribution $\iota\psi$:

$$\langle \iota\psi \mid \varphi\rangle_{\mathcal{S}'\times\mathcal{S}} := \int_{\mathbb{R}^n} \psi(x)\varphi(x)\,\mathrm{d}x, \qquad \forall\varphi \in \mathcal{S}(\mathbb{R}^n).$$

Proposition 2.37. *This map ι has dense range:* $\operatorname{img}(\iota)$ *is a dense subset of* $\mathcal{S}'(\mathbb{R}^n)$.

The proof is done in four steps.

Step 1 (already done): We have shown in Lemma 2.22 that ι is continuous and injective. Due to this injectivity, the map ι turns into an embedding $\mathcal{S} \hookrightarrow \mathcal{S}'$. And because ι is continuous, the transposed map

$$\iota^t : \mathcal{S}'' \to \mathcal{S}'$$

is continuous as well.

Step 2 (open): $\mathcal{S}'' = \mathcal{S}$, up to an algebraic isomorphism

Step 3 (open): ι^t is injective

Step 4: Conclusion of the proof: By Proposition 1.67, the operator ι then has dense range. This means that the embedding $\mathcal{S} \hookrightarrow \mathcal{S}'$ is dense, which is what we wanted to show.

Lemma 2.38. *There is an algebraic isomorphism $J : \mathcal{S} \to \mathcal{S}''$ that is continuous.*

We note that J^{-1} cannot be continuous, because $\mathcal{S}$ and $\mathcal{S}''$ have different topologies (only one of them is metrisable).

Proof. We define a map $J : \mathcal{S} \to \mathcal{S}''$ as

$$\langle J\psi \mid T\rangle_{\mathcal{S}''\times\mathcal{S}'} := \langle T \mid \psi\rangle_{\mathcal{S}'\times\mathcal{S}}, \qquad \forall T \in \mathcal{S}',$$

and obviously J is linear.

J is injective: if $J\psi = 0 \in \mathcal{S}''$ for some $\psi \in \mathcal{S}$, then we obtain, for all $T \in \mathcal{S}'$,

$$0 = \langle 0 \mid T\rangle_{\mathcal{S}''\times\mathcal{S}'} = \langle J\psi \mid T\rangle_{\mathcal{S}''\times\mathcal{S}'} = \langle T \mid \psi\rangle_{\mathcal{S}'\times\mathcal{S}}.$$

Now it suffices to choose $T = \iota\overline{\psi}$ to deduce that $\|\psi\|_{L^2} = 0$.

J is continuous: according to Definition 1.64 , the seminorms on $\mathcal{S}''$ are given via all $T \in \mathcal{S}'$:

$$p_T : f \mapsto p_T(f) := \left|\langle f \mid T\rangle_{\mathcal{S}''\times\mathcal{S}'}\right|, \qquad \forall f \in \mathcal{S}''.$$

Choose some $T \in \mathcal{S}'$. Then

$$p_T(J\psi) = \left|\langle J\psi \mid T\rangle_{\mathcal{S}''\times\mathcal{S}'}\right| = \left|\langle T \mid \psi\rangle_{\mathcal{S}'\times\mathcal{S}}\right| \le C_T p_m(\psi),$$

with p_m from (2.1) because T is a continuous map from $\mathcal{S}$ into $\mathbb{C}$. Now apply Proposition 1.59 to deduce that J is indeed continuous as a map from $\mathcal{S}$ to $\mathcal{S}''$.

J is surjective: we will prove that

$$\forall f \in \mathcal{S}'' \quad \exists \psi_f \in \mathcal{S} : \quad \langle f \mid T\rangle_{\mathcal{S}''\times\mathcal{S}'} = \langle T \mid \psi_f\rangle_{\mathcal{S}'\times\mathcal{S}}, \quad \forall T \in \mathcal{S}'. \tag{2.10}$$

Let us be given $f \in \mathcal{S}''$. Because f is continuous as a map from $\mathcal{S}'$ into $\mathbb{C}$, Proposition 1.59 gives us finitely many seminorms $p_{\varphi_1}, p_{\varphi_2}, \dots, p_{\varphi_K}$ on $\mathcal{S}'$ with

$$\left|\langle f \mid T\rangle_{\mathcal{S}''\times\mathcal{S}'}\right| \le C\left(p_{\varphi_1}(T) + p_{\varphi_2}(T) + \ldots + p_{\varphi_K}(T)\right)$$

with C depending only on f, hence

$$\left|\langle f \mid T\rangle_{\mathcal{S}''\times\mathcal{S}'}\right| \le C\sum_{j=1}^{K}\left|\langle T \mid \varphi_j\rangle_{\mathcal{S}'\times\mathcal{S}}\right|, \qquad \forall T \in \mathcal{S}'. \tag{2.11}$$

We may assume that the functions φ_1, ..., φ_K are linearly independent, otherwise we will drop some of them and increase C if needed. By a Gram–Schmidt argument, there are functions ψ_1, ψ_2, ..., $\psi_K \in \mathcal{S}$ with

$$\operatorname{span}(\varphi_1, \dots, \varphi_K) = \operatorname{span}(\psi_1, \dots, \psi_K) \quad \text{and} \quad \int_{\mathbb{R}^n} \overline{\psi_j}\psi_\ell \, \mathrm{d}x = \delta_{j\ell},$$

hence $\left\langle \iota\overline{\psi_j} \mid \psi_\ell \right\rangle_{\mathcal{S}'\times\mathcal{S}} = \delta_{j\ell}$. We can express each φ_j as a linear combination of the ψ_ℓ, and then (2.11) turns into

$$\left|\langle f \mid T\rangle_{\mathcal{S}''\times\mathcal{S}'}\right| \leq C \sum_{j=1}^{K} \left|\langle T \mid \psi_j\rangle_{\mathcal{S}'\times\mathcal{S}}\right|, \qquad \forall T \in \mathcal{S}', \tag{2.12}$$

with a new value of C. Remember that C does not depend on T.

For each $T \in \mathcal{S}'$, we build $T_0 := T - \sum_{j=1}^{K} \langle T \mid \psi_j\rangle_{\mathcal{S}'\times\mathcal{S}}\, \iota\overline{\psi_j}$. Now (2.12) must also hold with T_0 instead of T:

$$\left|\langle f \mid T_0\rangle_{\mathcal{S}''\times\mathcal{S}'}\right| \leq C \sum_{\ell=1}^{K} \left|\langle T_0 \mid \psi_\ell\rangle_{\mathcal{S}'\times\mathcal{S}}\right|.$$

Now what is $\langle T_0 \mid \psi_\ell\rangle$? We calculate

$$\langle T_0 \mid \psi_\ell\rangle_{\mathcal{S}'\times\mathcal{S}} = \langle T \mid \psi_\ell\rangle_{\mathcal{S}'\times\mathcal{S}} - \sum_{j=1}^{K} \langle T \mid \psi_j\rangle_{\mathcal{S}'\times\mathcal{S}} \left\langle \iota\overline{\psi_j} \mid \psi_\ell\right\rangle_{\mathcal{S}'\times\mathcal{S}} = \langle T \mid \psi_\ell\rangle_{\mathcal{S}'\times\mathcal{S}} - \langle T \mid \psi_\ell\rangle_{\mathcal{S}'\times\mathcal{S}} = 0,$$

which enforces $\langle f \mid T_0\rangle_{\mathcal{S}''\times\mathcal{S}'} = 0$. But then we conclude that

$$\begin{aligned}\langle f \mid T\rangle_{\mathcal{S}''\times\mathcal{S}'} &= \langle f \mid T_0\rangle_{\mathcal{S}''\times\mathcal{S}'} + \left\langle f \,\Big|\, \sum_{j=1}^{K} \langle T \mid \psi_j\rangle_{\mathcal{S}'\times\mathcal{S}}\, \iota\overline{\psi_j} \right\rangle_{\mathcal{S}''\times\mathcal{S}'} = \sum_{j=1}^{K} \langle T \mid \psi_j\rangle_{\mathcal{S}'\times\mathcal{S}} \left\langle f \mid \iota\overline{\psi_j}\right\rangle_{\mathcal{S}''\times\mathcal{S}'} \\ &= \left\langle T \,\Big|\, \sum_{j=1}^{K} \left\langle f \mid \iota\overline{\psi_j}\right\rangle_{\mathcal{S}''\times\mathcal{S}'} \psi_j \right\rangle_{\mathcal{S}'\times\mathcal{S}},\end{aligned}$$

which suggests the choice $\psi_f := \sum_{j=1}^{K} \left\langle f \mid \iota\overline{\psi_j}\right\rangle_{\mathcal{S}''\times\mathcal{S}'} \psi_j$. Now look at (2.10).

Hence we have shown J to be linear, bijective, and continuous. □

Now that we know how the elements of $\mathcal{S}''$ look like, we have much better prospects of showing that ι^t is injective:

Lemma 2.39. *The transposed map $\iota^t \colon \mathcal{S}'' \to \mathcal{S}'$ is injective.*

Proof. Assume $\iota^t f = 0 \in \mathcal{S}'$ for some $f \in \mathcal{S}''$. Since J is surjective, there is some $\psi_f \in \mathcal{S}$ with $f = J\psi_f$. Then we can calculate as follows, for each $\varphi \in \mathcal{S}$:

$$0 = \langle 0 \mid \varphi\rangle_{\mathcal{S}'\times\mathcal{S}} = \left\langle \iota^t f \mid \varphi\right\rangle_{\mathcal{S}'\times\mathcal{S}} = \langle f \mid \iota\varphi\rangle_{\mathcal{S}''\times\mathcal{S}'} = \langle J\psi_f \mid \iota\varphi\rangle_{\mathcal{S}''\times\mathcal{S}'}$$

We choose $\varphi = \overline{\psi_f}$ and see

$$0 = \left\langle J\psi_f \mid \iota\overline{\psi_f}\right\rangle_{\mathcal{S}''\times\mathcal{S}'} = \left\langle \iota\overline{\psi_f} \mid \psi_f\right\rangle_{\mathcal{S}'\times\mathcal{S}} = \int_{\mathbb{R}^n} |\psi_f(x)|^2 \, \mathrm{d}x,$$

which implies $\psi_f \equiv 0$, hence $f = 0 \in \mathcal{S}''$. □

Chapter 3

Applications of the Fourier Transform

3.1 Signal Theory

Suppose $f = f(t)$ describes an acustic signal, with $t \in \mathbb{R}$ as time variable. Then the Fourier transform $\hat{f} = \hat{f}(\tau)$ describes the strength of the frequency $\tau \in \mathbb{R}$ in the signal.

Similarly in a two-dimensional setting: consider a function $u = u(x, y)$ that describes an image. Then the Fourier transform can be used (with a lot of work not mentioned here) to detect properties of the image such as edges.

3.2 Statistics

The Fourier transform is often used for times series analysis and for the statistics of stochastic processes.

3.3 Partial Differential Equations

Suppose we wish to solve

$$\frac{\partial^2}{\partial x_1^2} u(x_1, x_2) - \frac{\partial^2}{\partial x_1 \partial x_2} u(x_1, x_2) + 13 \frac{\partial^2}{\partial x_2^2} u(x_1, x_2) - u(x_1, x_2) = f(x_1, x_2), \qquad (x_1, x_2) \in \mathbb{R}^2,$$

for u, where $f \in \mathcal{S}(\mathbb{R}^2)$ is given. You apply the Fourier transform and find

$$\Big(-\xi_1^2 + 2\xi_1\xi_2 - 13\xi_2^2 - 1 \Big) \hat{u}(\xi_1, \xi_2) = \hat{f}(\xi_1, \xi_2), \qquad \forall (\xi_1, \xi_2) \in \mathbb{R}^2.$$

The big bracket on the LHS is never zero, hence you can divide,

$$\hat{u}(\xi_1, \xi_2) = \frac{-1}{\xi_1^2 - 2\xi_1\xi_2 + 13\xi_1^2 + 1} \hat{f}(\xi_1, \xi_2),$$

and then you can write down a formula for $u(x_1, x_2)$:

$$u(x_1, x_2) = \frac{1}{(2\pi)^2} \int_{\xi \in \mathbb{R}^2} e^{\mathrm{i}(x_1\xi_1 + x_2\xi_2)} \frac{-1}{\xi_1^2 - 2\xi_1\xi_2 + 13\xi_1^2 + 1} \hat{f}(\xi_1, \xi_2) \, \mathrm{d}\xi.$$

It is a lot of work (left to future courses) to draw conclusions from this formula, and in particular to handle also the important case when the denominator can become zero. What we can say already is this: for each $f \in \mathcal{S}(\mathbb{R}^2)$, there is a unique solution u, and it belongs to $\mathcal{S}(\mathbb{R}^2)$ as well, and it depends continuously on f, measured in the topology of the Schwartz space.

As a final example, we wish to solve

$$\begin{cases} \partial_t u(t,x) - 19 \triangle u(t,x) = 0, & (t,x) \in [0,\infty) \times \mathbb{R}^n, \\ u(0,x) = u_0(x), & x \in \mathbb{R}^n, \end{cases}$$

where $u_0 \in \mathcal{S}(\mathbb{R}^n)$ is given and u is wanted. Fourier transform with respect to x implies

$$\begin{cases} \partial_t \hat{u}(t,\xi) + 19|\xi|^2 \hat{u}(t,\xi) = 0, & (t,\xi) \in [0,\infty) \times \mathbb{R}^n, \\ \hat{u}(0,\xi) = \widehat{u_0}(\xi), & \xi \in \mathbb{R}^n, \end{cases}$$

which has the solution $\hat{u}(t,\xi) = \exp(-19|\xi|^2 t)\widehat{u_0}(\xi)$, and now the inverse Fourier transform can be applied, where Proposition 2.10 comes in handy. Compare also Example 2.12. The details are left to the readers. You will obtain an explicit formula for the solution u that is valid not only for $u_0 \in \mathcal{S}(\mathbb{R}^n)$, but for locally integrable functions u_0 with at most polynomial growth as well.

Bibliography

[1] Klemens Burg, Herbert Haf, and Friedrich Wille. *Höhere Mathematik für Ingenieure. Band 3. Gewöhnliche Differentialgleichungen, Distributionen, Integraltransformationen. 3., durchges. Aufl.* Stuttgart: Teubner, 1993.

[2] Robert E. Edwards. *Functional analysis.* Dover Publications, Inc., New York, 1995. Theory and applications, Corrected reprint of the 1965 original.

[3] Harro Heuser. *Funktionalanalysis. 3. Auflage.* Stuttgart: Teubner, 1992.

[4] Harro Heuser. *Lehrbuch der Analysis. Teil 2. 10. Auflage.* Stuttgart: Teubner, 1998.

[5] Lars Hörmander. *The Analysis of Linear Partial Differential Operators*, volume 1-4. Springer, 1983.

[6] Gottfried Köthe. *Topological vector spaces. I.* Translated from the German by D. J. H. Garling. Die Grundlehren der mathematischen Wissenschaften, Band 159. Springer-Verlag New York Inc., New York, 1983.

[7] Jeffrey Rauch. *Partial Differential Equations.* Springer, 1991.

[8] Hans Triebel. *Höhere Analysis.* Thun, Frankfurt am Main: Verlag Harri Deutsch, 1980.

[9] Vasilij Sergeevich Vladimirov. *Gleichungen der mathematischen Physik.* Deutscher Verlag der Wissenschaften, 1972.

[10] Wolfgang Walter. *Einführung in die Theorie der Distributionen.* B.I. Wissenschaftsverlag, 1994.