ANALYSIS OF FUNCTIONS, LENT 2026

PÉTER P. VARJÚ

Please send comments to: pv270@dpmms.cam.ac.uk.

1. Recap of measure theory

This section is a brief reminder of Lebesgue's theory of integration, which was covered in the course Part II Probability and Measure. The reason why this theory has become successful is that it makes exchanging limits and integrals or two integrals with each other very easy. This is something that an analyst needs to do very often, so it is very useful that we can do it without much thinking.

To appreciate the power of the theory try to prove the following result using only what you learnt in the first two years of your degree. Let $f_n : [0,1] \to [0,1]$ be a sequence of continuous functions with $\lim f_n(x) = 0$ at every $x \in [0,1]$. Prove that $\int f_n dx \to 0$.

We will state the key results after recalling some basic terminology. Let E be a set. A σ-algebra $\mathcal{B}$ on E is a nonempty collection of sets that is closed under countable unions and complements. A measure μ is a set function $\mathcal{B} \to [0,\infty]$ that is σ-additive, that is $\mu(\bigcup A_n) = \sum \mu(A_n)$ for any countable collection of disjoint sets $A_n \in \mathcal{B}$. The tuple $(E,\mathcal{B})$ is called a measurable space, the triple $(E,\mathcal{B},\mu)$ is called a measure space. When E is a topological space, we always consider it with the Borel σ-algebra, that is the σ-algebra generated by open sets. We say that something holds for almost every x or almost surely if the set of x for which the property does not hold is of 0 measure.

Let $(E,\mathcal{B},\mu)$ be a measure space. A function $f : E \to \mathbf{C}$ is measurable if $f^{-1}(A) \in \mathcal{B}$ for all Borel sets $A \subset \mathbf{C}$. If f is measurable and $f(E) \subset [0,\infty]$, its integral $\int f d\mu$ is always defined, possibly ∞. We say that $f : E \to \mathbf{C}$ is integrable if it is measurable and $\int |f| d\mu < \infty$. In this case its integral $\int f d\mu$ can be defined and it is a (finite) complex number. We write $L^1(E,\mathcal{B},\mu)$ for the set of integrable functions $E \to \mathbf{C}$. When clear from the context, we may drop one or more of the arguments of the L^1 notation.

We are now ready to state the two main theorems about exchanging limits and integrals.

Theorem 1 (Lebesgue dominated convergence)**.** *Let $(E,\mathcal{B},\mu)$ be a measure space, and let $g, f, f_1, f_2, f_3, \ldots \in L^1(E,\mathcal{B},\mu)$. Suppose that*

Version of 26 March 2026.

$f(x) = \lim_{n\to\infty} f_n(x)$ for μ-almost every x and $|f_n(x)| \le g(x)$ for all n and μ-almost every x. Then

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

A measure space $(E, \mathcal{B}, \mu)$ is σ-finite if there is a countable collection of sets A_n such that $E = \bigcup A_n$ and $\mu(A_n) < \infty$ for all n.

Theorem 2 (Fubini)**.** *Let $(E, \mathcal{A}, \mu)$ and $(F, \mathcal{B}, \nu)$ be σ-finite measure spaces. Let $f : E \times F \to \mathbf{C}$ be an $\mathcal{A} \otimes \mathcal{B}$-measurable function.*

(1) *If $x \mapsto \int |f(x, y)| d\nu(y) \in L^1(E)$, then $f \in L^1(E \times F)$.*
(2) *If $f \in L^1(E \times F)$, then $f(x, \cdot) \in L^1(F)$ for μ-almost all values of x, and*

$$\int_{E\times F} f d\mu \otimes d\nu = \int_E \int_F f(x, y) d\nu(y) d\mu(x). \tag{1}$$

Note that when $f(x, y)$ is non-negative and measurable, then (1) holds even without any finiteness condition. Indeed, if either side is finite, then the first part of the the theorem implies $f \in L^1(E \times F)$. If both sides are ∞, then we also have the equality. Measurability is rarely an issue. Continuous functions are measurable, the composition of measurable functions is measurable and the limit of measurable functions is also measurable. So we can indeed exchange integrals without thinking!

We denote by dx or dy, etc the Lebesgue measure on $\mathbf{R}^d$ when x or y denotes the variable. We write $|A|$ for the Lebesgue measure of a measurable $A \subset \mathbf{R}^d$.

2. Signed and complex measures

Definition 3. Let $(E, \mathcal{B})$ be a measurable space. A complex measure is a set function $\mu : \mathcal{B} \to \mathbf{C}$ that is σ-additive.

If μ takes values in $\mathbf{R}$ we call it a signed measure.

A word on terminology. If we say measure without any adjectives, we mean one which takes values in $[0, \infty]$, that is, a measure in the sense of the Probability and Measure course. When we want to stress that we mean a measure that takes values in $[0, \infty]$ we will call it a positive measure.

Take note that a complex or signed measure only takes finite values by definition, while positive measures are allowed to take infinite values. This means that not all positive measures are signed or complex measures. Such an example is the Lebesgue measure on $\mathbf{R}$. Some authors allow signed measures to take one of $-\infty$ and ∞ but not both to accommodate all positive measures. There is no sensible way to extend this convention for complex measures.

We have two purposes for talking about complex and signed measures. One of them is that we want to turn the space of measures into a Banach space. Given two complex measures μ_1, μ_2 and $a, b \in \mathbf{C}$, we define

$$(a\mu_1 + b\mu_2)(A) = a\mu_1(A) + b\mu_2(A) \qquad \text{for } A \in \mathcal{B}.$$

This is easily seen to be a complex measure, and that this operation turns the space of complex measures on a measurable space into a vector space over $\mathbf{C}$.

The second purpose is that the Hahn-decomposition, the main structure theorem of signed measures will be used in our proof of the Radon-Nikodym theorem later.

Given a complex measure μ, we can define the set functions

$$\mathrm{Re}(\mu)(A) = \mathrm{Re}(\mu(A)), \qquad \mathrm{Im}(\mu)(A) = \mathrm{Im}(\mu(A)) \qquad \text{for } A \in \mathcal{B},$$

which we call the real and imaginary parts of μ. It is easy to see that $\mathrm{Re}(\mu)$ and $\mathrm{Im}(\mu)$ are signed measures. This construction allows us to reduce most problems about complex measures to signed measures.

Definition 4. Let $(E, \mathcal{B}, \mu)$ be a signed measure space. A set $A \subset \mathcal{B}$ is a positive set for μ, if $\mu(B) \geq 0$ for all $B \subset A$, and it is a negative set for μ if $\mu(B) \leq 0$ for all $B \subset A$.

Theorem 5 (Hahn decomposition of signed measures)**.** *Let $(E, \mathcal{B}, \mu)$ be a signed measure space. Then there is a decomposition $E = P \sqcup N$ such that P is a positive and N is a negative set for μ.*

Using this result, we can write a signed measure as the difference of two positive measures. Indeed, we define

$$\mu^+ = \mu|_P, \qquad \mu^- = -\mu|_N,$$

which are called the positive and negative parts of μ. Here and everywhere, if μ is a complex, signed or positive measure, and A is a measurable set, then $\mu|_A$ is the restriction of μ to A defined by

$$\mu|_A(B) = \mu(A \cap B).$$

It is clear from the definitions that μ^+ and μ^- are positive measures and $\mu = \mu^+ - \mu^-$. These two measures are called the Jordan decomposition of μ.

The decomposition $P \sqcup N$ is not unique, because we may move a measure 0 subset of P into N and a measure 0 subset of N to P. However any choice of the Hahn decomposition $P \sqcup N$ gives rise to the same μ^+ and μ^-.

This theorem allows us to decompose the space E as a disjoint union of two spaces, such that μ restricts to a positive measure on one of them and it restricts to -1 times a positive measure on the other. This allows us to reduce most problems about signed measures to positive measures.

In the proof, we will define P as a positive set of the largest possible measure. That such a set exist will require proof, but even just to get started, we need to show that non-trivial positive sets exist. This is done in the next lemma.

Lemma 6. *For all $A \subset E$, there is a positive set $D \subset A$ such that $\mu(D) \geq \mu(A)$.*

The idea of the proof is the following. If A is a positive set, then we are happy. If not, take some $B \subset A$ with $\mu(B) < 0$ and discard it, that is, replace A by $A \backslash B$. This will only increase the measure of A. To show that this process terminates requires some knowledge of set theory or a trick. We will go for the second option.

Proof. If $\mu(A) \leq 0$, then we can just take $B = \varnothing$ and if A is a positive set, we take $B = A$, so we assume neither of these is the case. Let $B_1 \subset A$ be with $\mu(B_1) \leq 0$ that is as negative as possible in the approximate sense that there is no $B \subset A$ and $k \in \mathbf{Z}_{>0}$ with $\mu(B_1) > -1/k \geq \mu(B)$. We define $A_1 = A \backslash B_1$.

We proceed with this process defining $B_2, B_3, \ldots$ and $A_2, A_2, \ldots$ in such a way that $B_{i+1} \subset A_i$ and $\mu(B_{i+1}) > -1/k$ for some $k \in \mathbf{Z}_{>0}$ only if this is so for all subsets of A_i. Then we take $A_{i+1} = A_i \backslash B_{i+1}$.

We take $D = \bigcap A_i$. Now

$$A = D \sqcup B_1 \sqcup B_2 \sqcup \ldots,$$

so

$$\mu(A) = \mu(D) + \sum_{i=1}^{\infty} \mu(B_i).$$

Since all $\mu(B_i) \leq 0$, we must have $\mu(D) \geq \mu(A)$. Moreover, by the finiteness of $\mu(A)$ and $\mu(D)$ the series must converge, and $\mu(B_i) \to 0$. In particular, for all k, there is some i such that $\mu(B_{i+1}) > -1/k$. Then $\mu(B) > -1/k$ for all $B \subset A_i$, hence for all $B \subset D$. Since this is true for all k, D must be a positive set. □

Proof of Theorem 5. Let s be the supremum of the measures of all positive subsets of E, and for each i, let $P_i \in \mathcal{B}$ be a positive set with $\lim \mu(P_i) = s$. Now $P = \bigcup P_i$ is a positive set since

$$\mu(B) = \sum_i \mu(B \cap P_i \backslash (P_1 \cup \ldots \cup P_{i-1})) \geq 0$$

for all $B \subset P$. Also $\mu(P) \geq \mu(P_i)$ for all i, hence $\mu(P) = s$.

Now suppose to the contrary that $N := E \backslash P$ is not a negative set. Then there is some set $B \subset N$ with $\mu(B) > 0$. By Lemma 6, there is some positive set $D \subset B$ with $\mu(D) \geq \mu(B) > 0$. However, then $P \cup B$ is a positive set of measure $s + \mu(D) > s$, a contradiction. □

Given a signed measure space $(E, \mathcal{B}, \mu)$, we say that $f : E \to \mathbf{C}$ is integrable if it is integrable with respect to both μ^+ and μ^-. In that case, we define

$$\int f(x) d\mu(x) = \int f(x) d\mu^+(x) - \int f(x) d\mu^-(x).$$

If μ is a complex measure, then f is integrable with respect to μ if it is integrable with respect to both $\mathrm{Re}(\mu)$ and $\mathrm{Im}(\mu)$, and if this is the case, we define

$$\int f(x) d\mu(x) = \int f(x) d \operatorname{Re} \mu(x) + i \int f(x) d \operatorname{Im} \mu(x).$$

The basic properties of integration including the dominated convergence theorem and Fubini's theorem can be extended to integration with respect to complex measures.

If μ is a signed measure, we define its total variation measure as $|\mu| = \mu^+ + \mu^-$ and its total variation norm by $\|\mu\| = |\mu|(E) = \mu^+(E) + \mu^-(E)$. These notions can be extended to complex measures, see the first example sheet. It can be shown that the space of complex measures on a measureable space $(E, \mathcal{B})$ forms a Banach space with respect to the total variation norm.

3. Radon-Nikodym theorem

Given a random variable X, there is a probability measure μ on $\mathbf{R}$ (that is a positive measure with total mass 1), such that $\mathbf{P}(X \in A) = \mu(A)$ for all Borel sets $A \subset \mathbf{R}$. We call this measure μ the distribution of X. All probability measures on $\mathbf{R}$ arise in this way. In Part II Probability and Measure, you have seen examples such that $\mu(A) = \int_A f dx$ for some density function $f \in L^1(\mathbf{R})$ or such that $\mu(A) = \sum_{x \in A} p_x$, where p_x are some non-negative numbers and the summation runs through a certain countable set. Are there probability distributions which do not fall in either category? How can we decide if a random variable has a density? We are going to answer these questions in this lecture.

Let μ and ν be positive measures on a measurable space $(E, \mathcal{B})$. We say that ν is absolutely continuous with respect to μ and write $\nu \ll \mu$ if $\mu(A) = 0$ implies $\nu(A) = 0$ for all $A \in \mathcal{B}$. We say that ν is singular with respect to μ and write $\nu \perp \mu$ if there is a decomposition $A \cup B = E$ such that $\mu(A) = 0$ and $\nu(B) = 0$. Observe that singularity is symmetric, that is, $\nu \perp \mu$ if and only if $\mu \perp \nu$. We say that μ concentrates on a set $A \in \mathcal{B}$ if $\mu(E \backslash B) = 0$. Therefore, ν and μ are singular (with respect to each other) if and only if there are disjoint sets $A, B \in \mathcal{B}$ such that μ is concentrated on A and ν is concentrated on B.

Theorem 7 (Radon-Nikodym)**.** *Let μ and ν be finite measures on a measurable space $(E, \mathcal{B})$. Assume $\nu \ll \mu$. Then there is a function*

$f \in L^1(\mu)$ such that

$$\nu(A) = \int_A f(x) d\mu(x)$$

for all $A \in \mathcal{B}$.

If g is another function that satisfies the conclusion of this theorem, then $f = g$ holds μ-almost everywhere.

The function $f(x)$ in the theorem is called the Radon-Nikodym derivative and it is denoted by

$$\frac{d\nu}{d\mu}(x).$$

Theorem 8 (Lebesgue decomposition)**.** *Let μ and ν be finite measures on a measurable space $(E, \mathcal{B})$. Then there are unique measures ν_a and ν_s such that $\nu = \nu_a + \nu_s$, $\nu_a \ll \mu$ and $\nu_s \perp \mu$.*

These theorems can be extended in an appropriate way to complex measures and to infinite measures under suitable conditions. See for example [4, Chapter 6].

We prove the two theorems together. For each $t \in \mathbf{R}_{\geq 0}$, let $P_t \sqcup N_t = E$ be a Hahn decomposition of $\nu - t\mu$. By definition, this means that for all measurable $A \subset P_t$, we have $(\nu - t\mu)(A) \geq 0$, hence $\nu(A) \geq t\mu(A)$, while a similar consideration gives $\nu(A) \leq t\mu(A)$ for $A \subset N_t$. When $t_1 < t_2$ and $A \subset P_{t_1} \cap N_{t_2}$ then

$$t_1\mu(A) \leq \nu(A) \leq t_2\mu(A),$$

so it is reasonable to expect that the Radon-Nikodym derivative will be between t_1 and t_2 on $P_{t_1} \cap N_{t_2}$.

With this intuition in mind, we make the following definition. For each $n \in \mathbf{Z}_{\geq 0}$, let

$$f_n(x) = \sup(t \in 2^{-n}\mathbf{Z}_{\geq 0} : x \in P_t).$$

here $2^{-n}\mathbf{Z}_{\geq 0}$ stands for the set of numbers of the form $2^{-n}a$, where $a \in \mathbf{Z}_{\geq 0}$. To avoid issues with thinking about the supremum of the empty set, we assume that $P_0 = E$, which is a legitimate choice for the Hahn decomposition. Observe that the set over which we take the supremum is increasing with n, so the sequence of functions f_n is monotone increasing at each point x. Therefore, the limit $f(x) := \lim_{n\to\infty} f_n(x)$ exists for every x, but may be infinite.

Lemma 9. *There is $A \in \mathcal{B}$ with $\mu(E \backslash A) = 0$ such that*

$$\nu(B) = \int_B f(x) d\mu(x)$$

for all $B \subset A$.

Proof. Fix some $n \in \mathbf{Z}_{\geq 0}$, and let $A_n = \{x : f_n(x) < \infty\}$. We prove $\mu(E \backslash A_n) = 0$. Fix an arbitrary $t \in 2^{-n}\mathbf{Z}_{\geq 0}$. Then $E\backslash A_n \subset \bigcup_{s \geq t} P_s$, which is a positive set for $\nu - t\mu$, so $\nu(E\backslash A_n) \geq t\mu(E\backslash A_n)$. Since $\nu(E\backslash A_n) < \infty$ and t can be taken arbitrarily large, $\mu(E\backslash A_n) = 0$.

Also fix some $B \subset A_n$. For $t \in 2^{-n}\mathbf{Z}$, write

$$B_t = \{x \in B : f_n(x) = t\}.$$

Observe that $B = \bigsqcup_t B_t$, and $B_t \subset P_t$. In addition for all $x \in B_t$, $x \notin P_{t+2^{-n}}$, hence $B_t \subset N_{t+2^{-n}}$. Therefore,

$$\int_{B_t} f_n(d) d\mu(x) = t\mu(B_t) \leq \nu(B_t) \leq (t + 2^{-n})\mu(B_t).$$

We get

$$\left|\nu(B_t) - \int_{B_t} f_n(x) d\mu(x)\right| \leq 2^{-n}\mu(B_t).$$

Summing up these inequalities for t, we get

$$\left|\nu(B) - \int_B f_n d\mu(x)\right| \leq 2^{-n}\mu(B).$$

If $B \subset A := \bigcap A_n$, then the above inequality is valid for all n and by the monotone convergence theorem, we have

$$\int_B f(x) d\mu(x) = \lim_{n\to\infty} \int_B f_n d\mu(x) = \nu(B).$$

We also have that $E\backslash A = \bigcup_n (E\backslash A_n)$ is a μ null set. □

Proof of Theorem 7. We first prove the existence of the Radon-Nikodym derivative. If $\nu \ll \mu$, then $\nu(E\backslash A) = 0$. Now let $B \in \mathcal{B}$ be arbitrary. We have $\nu(B) = \nu(A \cap B)$. On the other hand

$$\int_B f(x) d\mu(x) = \int_{B\cap A} f(x) d\mu(x),$$

because $\mu(B\backslash A) = 0$. Thus

$$\nu(B) = \int_B f(x) d\mu(x)$$

holds for all $B \in \mathcal{B}$, hence f is a Radon-Nikodym derivative.

Now suppose to the contrary that g is such that

$$\int_B f(x) d\mu(x) = \nu(B) = \int_B g(x) d\mu(x)$$

for all $B \in \mathcal{B}$ but $f = g$ does not hold almost everywhere. Then there is some $\varepsilon > 0$ such that one of the sets

$$B_1 := \{x : f(x) - g(x) > \varepsilon\}, \qquad B_2 := \{x : f(x) - g(x) < \varepsilon\}$$

has positive measure. Suppose it is the first one. Then

$$\nu(B_1) - \nu(B_1) = \int_{B_1} f(x) - g(x) d\mu(x) > \varepsilon\mu(B_1) > 0,$$

a contradiction. The other case is similar. □

Proof of Lebesgue decomposition. We first prove existence. We put $\nu_s = \mu|_{E\backslash A}$ and $\nu_a = \mu|_A$. Note that ν_s is concentrated on $E\backslash A$, while μ is concentrated on A, so ν_s is indeed singular. On the other hand, f is a Randon Nikodym derivative for ν_a, hence it is absolutely continuous.

We turn to uniqueness. Let

$$\nu = \nu_a + \nu_s = \widetilde{\nu}_a + \widetilde{\nu}_s$$

be two Lebesgue decompositions. We show $\nu_a = \widetilde{\nu}_a$. Let D and $\widetilde{D}$ be μ-null sets where ν_s and $\widetilde{\nu}_s$ are concentrated. Note that $F := D \cup \widetilde{D}$ is a μ-null set on which both ν_s and $\widetilde{\nu}_s$ are concentrated. Let $B \in \mathcal{B}$. Since $\nu_a, \widetilde{\nu}_a \ll \mu$,

$$\nu_a(B \cap F) = \widetilde{\nu}_a(B \cap F) = 0.$$

On the other hand,

$$\nu_s(B\backslash F) = \widetilde{\nu}_s(B\backslash F) = 0$$

because both measures are concentrated on F. Therefore

$$\nu_a(B\backslash F) = \widetilde{\nu}_a(B\backslash F) + \widetilde{\nu}_s(B\backslash F) - \nu_s(B\backslash F) = \widetilde{\nu}_a(B\backslash F).$$

Combining this with our previous identity, we get $\nu_a(B) = \widetilde{\nu}_a(B)$, so $\nu_a = \widetilde{\nu}_a$, indeed. From this, $\nu_s = \widetilde{\nu}_s$ follows, as well. □

Example 10. Fix some numbers $\lambda, p \in (0,1)$, and let $X_1, X_2, \ldots$ be a sequence of independent random variables taking the values 0 and 1 with probabilities $1-p$ and p, respectively. The measure $\nu_{\lambda,p}$ that is the distribution of the random variable

$$Y = \sum_{n=0}^{\infty} \lambda^n X_n$$

is called a Bernoulli convolution.

Now let $\lambda = 1/3$ and $p = 1/2$. Let A be the set of values that Y can take. We show that A has Lebesgue measure 0. On the other hand, $\nu_{1/3,1/2}(\mathbf{R}\backslash A) = 0$, so this shows that $\nu_{1/3,1/2} \perp dx$. However, $\nu_{1/3,1/2}(\{x\}) = 0$ for all $x \in \mathbf{R}$, so $\nu_{1/3,1/2}$ is not a discrete measure.

The values of $X_0, \ldots, X_{N-1}$ determine the value of Y up to an error at most $\sum_{n=N}^{\infty}(1/3)^n = 3^{-N+1}/2$. There are 2^N choices for these values, so A can be covered by 2^N intervals of length $3^{-N+1}/2$. Therefore, $|A| \le (2/3)^{N-1} \to 0$ so $|A| = 0$, indeed.

On the first example sheet, you will see that $\nu_{1/2,p} \perp dx$ for all $p \neq 1/2$. This is arguably a more interesting example than the one given above, because for these choices of the parameters the random variable Y takes all values in $[0,1]$.

It is a deep result of Solomyak that $\nu_{\lambda,1/2} \ll dx$ for almost all $\lambda \in [1/2, 1)$. However, in a later example sheet you will see that $\nu_{\theta,1/2} \perp dx$ for $\theta = (\sqrt{5}-1)/2$, the golden ratio. It is a major open problem to decide weather $\nu_{2/3,1/2} \perp dx$ or $\nu_{2/3,1/2} \ll dx$. (We do know that one of the two must hold.)

4. The Lebesgue differentiation theorem

Given an absolutely continuous measure $\mu \ll dx$ how do we find its Radon-Nikodym derivative? One reasonable attempt would be to take the limit $\lim_{r\to 0} \frac{\mu(B(x,r))}{|B(x,r)|}$, where $B(x,r)$ denotes the ball of radius r around x for $x \in \mathbf{R}^d$ and $r > 0$. It turns out that this limit exists for Lebesgue almost every x and equals the Radon-Nikodym derivative $d\mu/dx$. This can be deduced from the Lebesgue differentiation theorem, see the first example sheet.

In the theory of Riemann integration, we learnt the fundamental theorem of calculus, which states that for $F(x) = \int_0^t f(t)dt$ we have $F'(x) = f(x)$ for all continuous $f : \mathbf{R} \to \mathbf{C}$. This is valid under the weaker hypothesis that $f \in L^1(\mathbf{R})$ with the weaker conclusion that $F'(x) = f(x)$. In particular, if a probability distribution is absolutely continuous, the probability density function is the derivative of the distribution function almost everywhere. This is also a consequence of the Lebesgue differentiation theorem, see the first example sheet.

Definition 11. Let $f \in L^1(\mathbf{R}^d)$. A point $x \in \mathbf{R}^d$ is a Lebesgue point of f if

$$\lim_{r\to 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y) - f(x)|dy = 0.$$

Observe that if f is a Lebesgue point of x, then

$$\Big|\frac{1}{|B(x,r)|} \int_{B(x,r)} f(y)dy - f(x)\Big| \le \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y) - f(x)|dy \to 0.$$

Theorem 12 (Lebesgue differentiation theorem)**.** *Let $f \in L^1(\mathbf{R}^d)$. Then Lebesgue almost every $x \in \mathbf{R}^d$ is a Lebesgue point of x.*

Definition 13. Let $f \in L^1(\mathbf{R}^d)$. The Hardy-Littlewood maximal function Mf is defined as

$$Mf(x) = \sup_{r>0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y)|dy.$$

Theorem 14 (Maximal inequality)**.** *Let $f \in L^1(\mathbf{R}^d)$. Then*

$$|\{x : Mf(x) \ge t\}| \le 5^d t^{-1} \|f\|_1.$$

If $g \in L^1(\mathbf{R}^d)$, then Markov's inequality gives

$$|\{x : |g(x)| > t\}| = \int_{x:|g(x)|>t} dx \le \int_{x:|g(x)|>t} \frac{|g(x)|}{t} dx \le \frac{\|g\|_1}{t}.$$

The claim of the maximal inequality would follow if we had $\|Mf\|_1 \le 5^d \|f\|_1$. However, $Mf \in L^1(\mathbf{R}^d)$ does not hold in general, so we need to have a slightly weaker conclusion. A measurable function g is said to be in weak L^1 if $|\{x : |g(x)| > t\}| \le Ct^{-1}$ for some constant C and all $t \in \mathbf{R}_{>0}$.

Maximal inequalities are fundamentally important in harmonic analysis. Among other uses, they can be used to control errors in proof of almost sure convergence. You have already seen a similar result during the proof of the Birkhoff ergodic theorem in Part II Probability and Measure.

The constant 5^d can be substantially improved. The best constant is known to grow at most linearly in the dimension. It is an open problem whether the inequality is true with a constant independent of d.

In the proof of the Lebesgue differentiation theorem we also use the following result that will be proved later.

Lemma 15. *For any $f \in L^1(\mathbf{R}^d)$ and $\varepsilon > 0$, there is $g \in C_c(\mathbf{R}^d)$ such that $\|f - g\|_1 < \varepsilon$.*

Proof of the Lebesgue differentiation theorem. We begin by observing that the theorem holds for continuous functions. In fact, every point is a Lebesgue point in that case. Indeed, if f is continuous at x, then for all ε, there is some $r(\varepsilon)$ such that $|f(x)-f(y)| \le \varepsilon$ for all $y \in B(x, r(\varepsilon))$. In particular,

$$\frac{1}{|B(x,r)|}\int_{B(x,r)} |f(y) - f(x)|dy \le \varepsilon$$

for $r \le r(\varepsilon)$, which proves that every point is a Lebesgue point of a continuous function.

It is enough to prove that for all $\varepsilon_1 > 0$, the set

$$A(f,\varepsilon_1) := \Big\{x : \limsup_{r\to 0} |B(x,r))|^{-1}\int_{B(x,r)} |f(y) - f(x)|dy \ge \varepsilon_1\Big\}$$

is of measure 0. In fact, it is enough to show that $|A(f, \varepsilon_1)| < \varepsilon_2$ for all $\varepsilon_2 > 0$.

Let $g \in C_c(\mathbf{R}^d)$. Since

$$\Big|\int_{B(x,r)} |(f(y)-g(y))-(f(x)-g(x))|dy - \int_{B(x,r)} |f(y)-f(x)|dy\Big| \le \int_{B(x,r)} |g(y)-g(x)|dy$$

and every point is a Lebesgue point of a continuous function, $A(f, \varepsilon_1) = A(f-g, \varepsilon_1)$. Note that if $x \in A(f-g, \varepsilon_1)$, then either $|(f-g)(x)| \ge \varepsilon_1/2$ or $M(f - g)(x) \ge \varepsilon_1/2$. Therefore,

$$|A(f - g, \varepsilon_1)| \le 2\varepsilon_1^{-1}\|f - g\|_1 + 2 \cdot 5^d\varepsilon_1^{-1}\|f - g\|_1 \le 5^{d+1}\varepsilon_1^{-1}\|f - g\|_1$$

by Markov's inequality and the maximal inequality.

We use the lemma to choose g in such a way that $\|f-g\|_1 \le 5^{-d-1}\varepsilon_1\varepsilon_2$ and conclude $|A(f, \varepsilon_1)| \le \varepsilon_2$ as required. □

Begining the proof of the maximal inequality. Fix some $t > 0$. If $x \in \mathbf{R}^d$ is such that $Mf(x) > t$, then there is some $r(x)$ such that

$$t^{-1} \int_{B(x,r(x))} |f(y)| dy \geq |B(x, r(x))|. \tag{2}$$

Write $U(x) = B(x, r(x))$. Notice that $\bigcup_{x:Mf(x)>t} U(x)$ contains all points with $Mf(x) > t$, so we would be done if we could show that

$$|\bigcup U(x)| \leq t^{-1} \int_{\bigcup U(x)} |f(y)| dy \leq t^{-1} \|f\|_1.$$

If the sets $U(x)$ were disjoint, this would follow by σ-additivity and the properties of the integral by summing (2). If there are overlaps, summation of (2) overcounts the overlaps on both side of the inequality, and it is not so clear to see what happens.

The balls $U(x)$ are very unlikely to be disjoint, in fact, we have uncountably many of them! However, the next lemma helps us to resolve this issue. □

Lemma 16 (Vitali covering lemma)**.** *Let $\mathcal{U}$ be a collection of balls in $\mathbf{R}^d$ whose diameter is bounded Then there is (finite or countably infinite) subcollection $\{V_1, V_2, \ldots\} \subset \mathcal{U}$ of disjoint balls such that*

$$\bigcup \mathcal{U} \subset \bigcup_j 5 \cdot V_j$$

or $|\bigcup_j V_j| = \infty$.

Here $5 \cdot V_j$ denotes the dilate of V_j around its centre by a factor of 5.

In this lemma, balls could be replaced by other convex sets with bounded eccentricity, but the proof completely breaks down if we allow arbitrary shapes. Understanding how general convex sets overlap is a very active area of research with some recent exciting developments under the banner of the Kakeya problem.

Proof. We define $V_1, V_2, \ldots$ recursively using a greedy algorithm. Let V_1 be such that $\operatorname{diam}(V_1) \geq \operatorname{diam}(U)/2$ for any $U \in \mathcal{U}$. Once $V_1, \ldots, V_n$ have been selected for some $n \geq 1$, we select V_{n+1} such that it is disjoint from $V_1 \cup \ldots \cup V_n$ and $\operatorname{diam}(V_{n+1}) \geq \operatorname{diam}(U)/2$ for all $U \in \mathcal{U}$ for which U is disjoint from $V_1 \cup \ldots \cup V_n$.

The V_j are clearly disjoint by construction. Suppose $|\bigcup_j V_j| < \infty$ and let $U \in \mathcal{U}$. Then $\{V_1, V_2, \ldots\}$ is either finite or $\operatorname{diam} V_j \to 0$. In both cases, for all i, there is some n such that U is not disjoint from $V_1 \cup \ldots \cup V_n$. Let n be the smallest such value. Then $U \cap V_n \neq \varnothing$ and $\operatorname{diam}(U) \leq 2 \operatorname{diam}(V_n)$, for otherwise we would have selected a larger ball for V_n. Now elementary geometry gives $U \subset 5 \cdot V_n$, and the lemma is proved. □

Completing the proof of the maximal inequality. By (2), $|U(x)|$ and hence $\operatorname{diam}(U(x))$ can be bounded independently of x, so the lemma can be applied to the collection $\mathcal{U} = \{U(x) : Mf(x) > t\}$. Let V_j be as in the conclusion of the lemma. Since the V_j are disjoint,

$$\sum_j |V_j| = \Big|\bigcup_j V_j\Big| \le t^{-1}\int_{\bigcup V_j} |f(y)|dy \le t^{-1}\|f\|_1.$$

In particular $|\bigcup V_j| < \infty$, so the first alternative of the lemma must hold, and

$$\Big|\bigcup \mathcal{U}\Big| \le \Big|\bigcup_j 5\cdot V_j\Big| \le \sum_j |5\cdot V_j| = 5^d \sum_j |V_j| \le 5^d t^{-1}\|f\|_1.$$

□

5. Measures on compact metric spaces

In this section, we discuss the approximation of measurable functions by continuous functions.

In this section and the next, E is a compact topological space endowed with a metric dist. We state results in this setting, which makes the statements and proofs simpler, but everything is true in greater generality. If you want to work with non-compact metric spaces, which nevertheless have an abundant supply of compact sets, e.g. $\mathbf{R}^n$, then you can usually get what you want by restricting everything to large compact subsets and applying the results there, or by embedding your space in a compact space. If your space is not metric, things get a bit more complicated and you need to work with regular or Radon measures (which roughly means that the conclusion of Proposition 17 holds for your measures), or you need to work with the σ-algebra of Baire sets rather than Borel sets. You may find definitions and more general results in [4, Chapter 2], [3, Chapter 14] or [2, Chapter 10] in increasing order of sophistication.

Proposition 17. *Let $\mathcal{B}$ be the Borel σ-algebra of a compact metric space (E, dist), and let μ be a finite measure on $(E, \mathcal{B})$. Then for all $A \in \mathcal{B}$ and $\varepsilon > 0$, there is a compact set $K \subset E$ and an open set $U \subset E$ such that $K \subset A \subset U$ and $\mu(U \backslash K) < \varepsilon$.*

I believe this was covered in Part II Probability and Measure, the proof will not be lectured, but it is here for completeness.

Proof. We show that the conclusion holds if A is an open set, and that the collection of sets for which the conclusion holds is a σ-algebra. Since $\mathcal{B}$ is contained in any σ-algebra containing all open sets, this proves the claim.

Suppose A is open, and fix $\varepsilon > 0$. Consider

$$K_n = \{x \in E : \mathrm{dist}(x, y) \ge 1/n \text{ for all } y \in E\backslash A.\}$$

We show that K_n is closed, hence compact. If $x_0 \notin K_n$, then there is $y \in E\backslash A$ such that $\text{dist}(x_0, y) < 1/n$. Then the same is true for all x in a neighbourhood of x_0. Thus $E\backslash K_n$ is open and K_n is closed.

The sequence of sets K_n clearly increases. We show its union is A. If not, then there is $x \in A$ such that $x \notin K_n$ for any n. Then for all n, there is $y_n \in E\backslash A$ with $\text{dist}(y_n, x) < 1/n$. Since $E\backslash A$ is closed, $\lim y_n = x \in E\backslash A$, a contradiction.

Now we have $\bigcap_n (A\backslash K_n) = \varnothing$, so $\mu(A\backslash K_n) < \varepsilon$ if n is large enough. The claim follows if we take $K = K_n$ and $U = A$.

Now suppose $A \in \mathcal{B}$ is such that the claim holds. Fix $\varepsilon > 0$ and let K_1 be compact and U_1 be open such that $K_1 \subset A \subset U_1$ and $\mu(U_1\backslash K_1) < \varepsilon$. Observe that $K = E\backslash U_1$ is compact and $U = E\backslash K_1$ is open, $K \subset E\backslash A \subset U$ and $\mu(U\backslash K) = \mu(U_1\backslash K_1) < \varepsilon$. Therefore, the claim also holds for $E\backslash A$.

Finally, let $A_1, A_2, \ldots \in \mathcal{B}$ be sets for which the claim holds, and fix $\varepsilon > 0$. For each $j \in \mathbf{Z}_{>0}$, let K_j be compact, U_j be open such that $K_j \subset A_j \subset U_j$ and $\mu(U_j\backslash K_j) < \varepsilon/2^j$. We define $K = K_1 \cup \ldots \cup K_N$ for a suitable $N \in \mathbf{Z}_{>0}$ and $U = \bigcup U_j$. Whatever the value of N, we have $K \subset \bigcup A_j \subset U$, K is compact and U is open. Moreover,

$$\mu(U\backslash \bigcup K_j) \leq \sum_j \mu(U_j\backslash K_j) < \varepsilon.$$

Note that

$$\lim_{N\to\infty} \mu(U\backslash(K_1 \cup \ldots \cup K_N)) = \mu(U\backslash \bigcup K_j),$$

so $\mu(U\backslash K) < \varepsilon$ if we choose N large enough. Therefore $\bigcup A_j$ also satisfies the claim, and this completes the proof. $\square$

Theorem 18 (Lusin)**.** *Let E be a compact metric space, and let μ be a finite measure on the Borel σ-algebra $\mathcal{B}$. Let $f : E \to \mathbf{C}$ be measurable and let $\varepsilon > 0$. Then there is a continuous $g : E \to \mathbf{C}$ with $\|g\|_\infty \leq \|f\|_\infty$ and*

$$\mu(x : f(x) \neq g(x)) \leq \varepsilon.$$

Moreover, if $A \subset E$ is a closed set such that $f|_A \equiv 0$, then g can be chosen so that $g|_A \equiv 0$.

Lemma 19. *Let $A \in \mathcal{B}$ and let $\varepsilon > 0$. Then there is a continuous function $g : E \to [0, 1]$ such that*

$$\mu(x : 1_A(x) \neq g(x)) \leq \varepsilon.$$

Here, and everywhere in these notes, we write

$$1_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Let $K \subset A \subset U$ be such that K is compact, U is open and $\mu(U \backslash K) \le \varepsilon$. Use Uryshon's lemma to find $g : E \to [0,1]$ such that $g|_K = 1$ and $g|_{E\backslash U} = 0$, or simply take

$$g(x) = \frac{\mathrm{dist}(x, E\backslash U)}{\mathrm{dist}(x,K) + \mathrm{dist}(x, E\backslash U)}.$$

□

Proof of Lusin's theorem. We prove the theorem first in the special case that $f(E) \subset [0,1)$. Fix $\varepsilon > 0$. For $n \in \mathbf{Z}_{>0}$, write A_n for the set of points $x \in E$ such that the n'th digit in the binary expansion of $f(x)$ is 1. (If $f(x)$ has two binary expansions, then we may use either as long as the choice is applied consistently.) In other words $x \in A_n$ if and only if $a + 2^{-n} \le f(x) < a + 2^{-n+1}$ for some $a \in 2^{-n+1} \cdot \mathbf{Z}$.

Observe that

$$f(x) = \sum_{n=1}^{\infty} 2^{-n} 1_{A_n}(x)$$

for all $x \in E$. Now we apply the lemma to find $g_n : E \to [0,1]$ such that

$$\mu(x : 1_{A_n}(x) \neq g_n(x)) \le \varepsilon/2^n.$$

Now

$$g(x) = \sum_{n=1}^{\infty} 2^{-n} g_n(x)$$

is the uniform limit of continuous functions and we have $g(x) = f(x)$ for x in the set

$$E' = \{x : g_n(x) = 1_{A_n}(x) \text{ for all } n\}.$$

The complement of this set has measure less than $\sum_{n=1}^{\infty} \varepsilon/2^n = \varepsilon$.

Let $U_0 = E\backslash A$, where A is a closed set with $f|_A \equiv 0$. Then $A_n \subset U_0$ for all sets A_n in the above construction, so we can select the set U in the proof of Lemma 19 so that $U \subset U_0$ each time, and then the resulting function g will vanish on A.

The general case of the theorem can be reduced to the special case in many different ways. One option is to reduce the complex valued case to the case $f : E \to \mathbf{R}_{\ge 0}$ by approximating the positive and negative parts of the real and imaginary parts separately. Then we may consider a strictly monotone continuous bijection $h : \mathbf{R}_{\ge 0} \to [0,1)$, apply the special case to the function $h \circ f$, and then compose the approximating function by h^{-1}. (This may prove the theorem with a worse bound for $\|g\|_\infty$, but we can remedy this if approximate $|f|$ and $\arg f$ instead of the real and imaginary parts. We leave the details as an exercise.) □

Proof of Lemma 15. Let $f \in L^1(\mathbf{R}^d)$ and $\varepsilon > 0$. We show that there is $g \in C_c(\mathbf{R}^d)$ such that $\|f - g\|_1 \le \varepsilon$. For $R \in \mathbf{R}_{>0}$ write $f_R(x) = f(x)$ if $|x| < R$ and $|f(x)| < R$ and $f_R(x) = 0$ otherwise. By the dominated convergence theorem, $\|f - f_R\|_1 \to 0$ as $R \to \infty$. Let R be large enough so that $\|f - f_R\|_1 < \varepsilon/10$. By Lusin's theorem applied for the ball of radius $2R$ around 0, and the set $A = \{x : |x| \geq R\}$, we can find a continuous function g such that $f_R(x) = g(x)$ outside a set of Lebesgue measure less than ε_1, where $\varepsilon_1 > 0$ is for us to choose. Moreover, $g(x) = 0$ for $|x| \geq R$, so we may extend it to $\mathbf{R}^d$ by setting it to 0 outside the ball of radius $2R$.

Now

$$\|f-g\|_1 \leq \|f_R-f\|_1+\|f_R-g\|_1 \leq \varepsilon/10+\varepsilon_1(\|f_R\|_\infty+\|g\|_\infty) \leq \varepsilon/10+2\varepsilon_1\|f_R\|_\infty < \varepsilon$$

provided we choose ε_1 small enough. □

Theorem 20 (Egorov). *Let $(E, \mathcal{B}, \mu)$ be a finite measure space. Let $f, f_1, f_2, \ldots$ be a sequence of measurable functions such that*

$$f(x) = \lim_{n\to\infty} f_n(x)$$

for almost every x. Then for every $\varepsilon > 0$, there is a set $A \in \mathcal{B}$ such that $f_1|_A, f_2|_A, \ldots$ converges uniformly and $\mu(E\backslash A) < \varepsilon$.

Proof. Fix some $k, N \in \mathbf{Z}_{>0}$. Define

$$A_{k,N} = \{x : |f(x) - f_n(x)| < 1/k \text{ for all } n > N\}.$$

This sequence of sets is increasing as N increases, and $\bigcup A_{k,N}$ contains all points where f_n converges to f. Therefore, $\mu(E\backslash \bigcup_N A_{k,N}) = 0$, and we may choose $N(k)$ in such a way that

$$\mu(E\backslash A_{k,N(k)}) \leq \varepsilon/2^k.$$

Now we take $A = \bigcap_k A_{k,N(k)}$. We observe that

$$\mu(E\backslash A) \leq \sum_{k=1}^{\infty} \varepsilon/2^k = \varepsilon,$$

and for $x \in A$ and $n > N(k)$ we have $|f_n(x) - f(x)| \leq 1/k$. The choice of $N(k)$ is independent of x, therefore the convergence is uniform on A. □

6. Riesz representation theorem

In this section, E is a compact topological space endowed with a metric dist, and $\mathcal{B}$ denotes the Borel σ-algebra. we denote by $C(E)$ the space of continuous functions on E. This is a Banach space with the supremum norm

$$\|f\|_\infty = \sup |f| = \max |f|.$$

(The definition of a Banach space will be recalled later, for now we do not need to know what this means.)

A bounded linear functional on $C(E)$ is a map $L : C(E) \to \mathbf{C}$ such that $L(a_1 f_1 + a_2 f_2) = a_1 L(f_1) + a_2 L(f_2)$ for all $a_1, a_2 \in \mathbf{C}$ and $f_1, f_2 \in C(E)$ and $\|Lf\| \leq A\|f\|$ for some $A \in \mathbf{R}_{\geq 0}$ that is independent of $f \in C(E)$. The infimum of all values of A for which the inequality holds is the norm of L, and it is denoted by $\|L\|$. We say that L is positive if $L(f) \geq 0$ for all $f \in C(E)$ with $f(\mathbf{R}) \subset \mathbf{R}_{\geq 0}$.

One way to construct a positive bounded linear functional is by taking the integral of a function with respect to a finite measure. That is,

$$L(f) = \int f d\mu$$

is a bounded positive linear functional on $C(E)$ for any finite Borel measure μ. The next theorem show that this is the only way to construct such functionals.

Theorem 21 (Riesz representation)**.** *Let E be a compact metric space, and let L be a positive bounded linear functional on $C(E)$. Then there is a unique finite Borel measure μ such that*

$$L(f) = \int f d\mu$$

for all $f \in C(E)$. Moreover $\|L\| = \mu(E)$.

The result has a version for not necessarily positive functionals.

Theorem 22 (Riesz representation)**.** *Let E be a compact metric space, and let L be a bounded linear functional on $C(E)$. Then there is a unique complex Borel measure μ such that*

$$L(f) = \int f d\mu$$

for all $f \in C(E)$. Moreover $\|L\| = \|\mu\|$.

This is a great theorem for at least two reasons. It gives us another way of thinking about measures. Sometimes the easiest way to describe a measure is by specifying the integral of continuous functions. Second, it realizes the space of measures as the dual of a Banach space. We will see later, that on dual spaces there is a topology called the weak-$*$ topology that has very nice properties. Restricting it to probability measures is related to convergence in distribution.

In this course, we only prove the version for positive functionals.

Proof of uniqueness. Let μ_1, μ_2 be two finite Borel measures such that

$$\int f d\mu_1 = \int f d\mu_2$$

for all $f \in C(E)$. We prove that $\mu_1 = \mu_2$.

Let $A \in \mathcal{B}$ and fix $\varepsilon > 0$. By Proposition 17, there are K_1, K_2 compact and U_1, U_2 open such that $K_j \subset A \subset U_j$ and $\mu_j(U_j \backslash K_j) < \varepsilon$ for $j = 1, 2$. We take $U = U_1 \cap U_2$ and $K = K_1 \cup K_2$. Then $K \subset A \subset U$

and $\mu_j(U\backslash K) < \varepsilon$ for both $j = 1, 2$. We let $f \in C(E)$ be such that $f(x) = 0$ for $x \in E\backslash U$ and $f(x) = 1$ for $x \in K$. Then

$$\left|\mu_j(A) - \int f d\mu_j\right| < \varepsilon$$

for both $j = 1, 2$ and $\int f d\mu_1 = \int f d\mu_2$. Thus $|\mu_1(A) - \mu_2(A)| < \varepsilon$. Since ε was arbitrary, $\mu_1(A) = \mu_2(A)$. Since A was arbitrary, $\mu_1 = \mu_2$. ☐

Now we turn to the proof of existence. If U is an open set and $f \in C(E)$, we write $f \prec U$ if $f(E) \subset [0, 1]$ and $\operatorname{supp} f \subset U$. This is a mildly stronger condition than $f \leq 1_U$. We fix a positive bounded linear functional L on $C(E)$. We define

$$\mu(U) = \sup(L(f) : f \prec U)$$

for open sets $U \subset X$ and

$$\mu(A) = \inf(\mu(U) : A \subset U,\ U \text{ is open})$$

for arbitrary $A \subset E$. Note that the two definitions are compatible for open sets.

We will show that μ is an outer measure, that is,

(1) $\mu(\varnothing) = 0$,
(2) $\mu(A_1) \leq \mu(A_2)$ whenever $A_1 \subset A_2$ and
(3) $\mu(\bigcup A_n) \leq \sum \mu(A_n)$ for arbitrary sets $A_n \subset E$.

The first two properties are trivial.

Then we will show that open sets are μ-measurable, in the sense of outer measures, that is

$$\mu(A) = \mu(A \cap U) + \mu(A\backslash U)$$

for all $A \subset E$ and open U. Then we will use a result from Part II Probability and Measure that says that the set of μ-measurable sets form a σ-algebra and μ restricted to it is a measure. Therefore, Borel sets are included in this σ-algebra, and $\mu|_{\mathcal{B}}$ is a measure. Finally, we will show that

$$L(f) = \int f d\mu$$

for all $f \in C(X)$. We observe that $\|L\| = L(1) = \int 1 d\mu = \mu(X)$, and this completes the proof.

We begin by a technical lemma, which implies that we could replace $f \prec U$ by $f \leq 1_U$ in the definition of $\mu(U)$.

Lemma 23. *Let $U \subset E$ be open, and let $f : E \to \mathbf{R}$ be a continuous function such that $f(x) \leq 1_U(x)$ for all x, then $L(f) \leq \mu(U)$.*

Proof. Since L is positive and $f(x) \leq \max(f(x), 0)$, we have $L(f) \leq L(\max(f, 0))$, so it is enough to prove the lemma for non-negative functions. We assume f is non-negative.

Fix $\varepsilon > 0$ and consider the function $g(x) = \max(f - \varepsilon, 0)$. Note that $\|f - g\| \le \varepsilon$ hence $|L(f) - L(g)| \le \varepsilon\|L\|$. On the other hand,

$$\operatorname{supp} g \subset \{x : f(x) \ge \varepsilon\} \subset U,$$

so $g \prec U$ and by definition of $\mu(U)$,

$$\mu(U) \ge L(g) \ge L(f) - \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, $\mu(U) \ge L(f)$. □

Proof that μ is σ-subadditive. We first show the property for open sets. Let $U_1, U_2, \ldots$ be open sets. We show that $\mu(\bigcup U_n) \le \sum \mu(U_n)$. To this end, it is enough to show that $L(f) \le \sum \mu(U_n)$ for all continuous $f \prec \bigcup U_n$.

Fix such a function f. Note that $\operatorname{supp} f$ is a compact set contained in $\bigcup U_n$. Therefore, $\operatorname{supp} f \subset U_1 \cup \ldots \cup U_N$ if N is large enough, which we assume.

We will find a decomposition $f = f_1 + \ldots + f_N$ such that $f_j \le 1_{U_j}$ for each j. Then it will follow that

$$L(f) = \sum_{j=1}^{N} L(f_j) \le \sum_{j=1}^{n} \mu(U_j),$$

and this proves the claim for open sets.

We employ a construction called a partition of unity, which can be explicitly constructed as follows in our situation. Let

$$g_j(x) = \frac{\operatorname{dist}(x, E\backslash U_j)}{\operatorname{dist}(x, E\backslash U_1) + \ldots + \operatorname{dist}(x, E\backslash U_N) + \operatorname{dist}(x, \operatorname{supp} f)}.$$

First observe that the denominator is never 0. Indeed, if $\operatorname{dist}(x, \operatorname{supp} f) = 0$, then $x \in \operatorname{supp} f$ and $x \in U_j$ for some j, and then $\operatorname{dist}(x, E\backslash U_j) > 0$. Therefore each g_j is continuous. Moreover $g_j \le 1_{U_j}$. Finally, $g_1(x) + \ldots + g_N(x) = 1$ for all $x \in \operatorname{supp} f$. Now it is easy to see that $f_j = f g_j$ satisfies all our requirements.

It remains to prove the claim in the general case. Let $A_1, A_2 \ldots$ be arbitrary sets, and fix $\varepsilon > 0$. For each j, let $U_j \supset A_j$ be a open such that $\mu(U_j) \le \mu(A_j) + \varepsilon/2^j$. Then

$$\mu(\bigcup A_j) \le \mu(\bigcup U_j) \le \sum \mu(U_j) \le \sum (\mu(A_j) + \varepsilon/2^j) \le \varepsilon + \sum \mu(A_j).$$

Since ε was arbitrary, this proves the claim. □

Proof that open sets are μ-measurable. Let $A \subset X$ be arbitrary and let U be open. We need to show $\mu(A) \ge \mu(A \cap U) + \mu(A\backslash U)$. The reverse inequality follows from sub-additivity, which we already proved. To this end, it is enough to show that $\mu(V) \ge \mu(A \cap U) + \mu(A\backslash U)$ for all open $V \supset A$ by the definition of $\mu(A)$. This will immediately follow if we show

$$\mu(V) \ge \mu(V \cap U) + \mu(V\backslash U) - \varepsilon \tag{3}$$

for all $\varepsilon > 0$.

We need to construct some $f \prec V$ such that $L(f)$ is at least as large as the right hand side of (3). We first find some $f_1 \prec V \cap U$ with $L(f_1) \geq \mu(V \cap U) - \varepsilon/2$, which exists by definition. Then we consider the set $V_2 = V \backslash \operatorname{supp} f_1$. This is an open set and $V_2 \supset V \backslash U$ so there exists $f_2 \prec V_2$ with

$$L(f_2) \geq \mu(V_2) - \varepsilon/2 \geq \mu(V \backslash U) - \varepsilon/2.$$

We take $f = f_1 + f_2$. Since f_1 and f_2 have disjoint supports contained in V, $f \prec V$, and

$$\mu(V) \geq L(f) = L(f_1) + L(f_2) \geq \mu(V \cap U) + \mu(V \backslash U) - \varepsilon.$$

□

Lemma 24. *Let $A \in \mathcal{B}$, and let $f \in C(E)$ with $f(E) \subset [0,1]$. If $f(x) \leq 1_A(x)$ for all x then $L(f) \leq \mu(A)$. If $f(x) \geq 1_A(x)$ for all x then $L(f) \geq \mu(A)$.*

Proof. Let $U \supset A$ open. Then $f \leq 1_A \leq 1_U$, and Lemma 23 gives $\mu(U) \geq L(f)$. Since U is arbitrary, this implies $\mu(A) \geq L(f)$ by definition.

For the other inequality, we use the statement we just proved for $E \backslash A$ in the role of A. If $f \geq 1_A$ then $1 - f \leq 1_{E \backslash A}$.

$$\mu(E \backslash A) \geq L(1_E - f) = L(1_X) - L(f) = \mu(E) - L(f).$$

Here we used that $\mu(X) = L(1_E)$, which follows from the fact that X is open, 1_E is continuous with $\operatorname{supp}(1_E) \subset X$ and $f \leq 1_E$ for all $f \in C(E)$ with $f(E) \subset [0,1]$. Now we get $\mu(A) = \mu(E) - \mu(E \backslash A) \leq L(f)$, as required. □

Proof that μ represents f. Let $f \in C(E)$. We prove $L(f) = \int f d\mu$. Since both integration and L are linear and any continuous function can be written as a linear combination of ones with values in $[0,1]$, we assume as we may that $f(E) \subset [0,1]$. Fix some $n \in \mathbf{Z}_{>0}$, and define

$$A_j := \{x : f(x) \geq j/n\},$$
$$f_j := \min(\max(f - j/n, 0), 1/n)$$

for $j = 0, \ldots, n$. Observe that $f = f_0 + \ldots + f_{n-1}$ and

$$\frac{1}{n} 1_{A_{j+1}} \leq f_j \leq \frac{1}{n} 1_{A_j}$$

for $j = 0, \ldots, n-1$.

Using the lemma and the monotonicity of integration, we have

$$\frac{1}{n}\mu(A_{j+1}) \leq L(f_j) \leq \frac{1}{n}\mu(A_j),$$
$$\frac{1}{n}\mu(A_{j+1}) \leq \int f_j d\mu \leq \frac{1}{n}\mu(A_j)$$

and hence
$$|L(f_j) - \int f_j d\mu| \le \frac{\mu(A_j) - \mu(A_{j+1})}{n}.$$
We sum this up for $j = 0, \ldots, n-1$, and get
$$\Big|L(f) - \int f d\mu\Big| \le \sum_{j=0}^{n-1} \frac{\mu(A_j) - \mu(A_{j+1})}{n} = \frac{\mu(A_n) - \mu(A_0)}{n} \le \mu(E)/n.$$
Taking $n \to \infty$, the claim follows. □

7. L^p SPACES

Recall the following definitions and facts from Part II Probability and Measure. Let $(E, \mathcal{B}, \mu)$ be a measure space. If $f : E \to \mathbf{R}$ is measurable, and $p \in [1, \infty)$ we write
$$\|f\|_p = \Big(\int_E |f(x)|^p d\mu(x)\Big)^{1/p}.$$
We also write
$$\|f\|_\infty = \inf\{\lambda \in \mathbf{R}_{\ge 0} : |f(x)| \le \lambda \ \mu\text{-almost everywhere}\}.$$
(The infimum of the empty set is ∞.) We write $L^p(E) = L^p(E, \mathcal{B}, \mu)$ for the collection of functions with $\|f\|_p < \infty$.

Recall also Minkowski's inequality
$$\|f_1 + f_2\|_p \le \|f_1\|_p + \|f_2\|_p$$
for $p \in [1, \infty]$ and $f_1, f_2 \in L^p(E)$. If we identify functions in $L^p(E)$ that are equal μ-almost everywhere, then $\|\cdot\|_p$ turns $L^p(E)$ into a normed space. Moreover, $L^p(E)$ is complete with respect to this norm, therefore a Banach space.

Another important inequality is Holder's:
$$\Big|\int_E f(x) g(x) d\mu(x)\Big| \le \|f\|_p \|g\|_q$$
whenever $p, q \in [1, \infty]$ and $p^{-1} + q^{-1} = 1$.

The following is a useful generalization of Minkowski's inequality.

Theorem 25 (Minkowski's integral inequality)**.** *Let $(E, \mathcal{A}, \mu)$ and $(F, \mathcal{B}, \nu)$ be σ-finite measure spaces. Let $G : E \times F \to \mathbf{R}$ be a $\mathcal{A} \otimes \mathcal{B}$-measurable function. Let $p \in [1, \infty]$. Suppose that $G(\cdot, y) \in L^p(E)$ for almost all y and*
$$\int \|G(\cdot, y)\|_p d\nu(y) < \infty.$$
Then
$$g(x) = \int G(x, y) d\nu(y)$$
exists for μ-almost all x, and
$$\|g\|_p \le \int \|G(\cdot, y)\|_p d\nu(y).$$

When $F = \{y_1, y_2\}$ and $\nu(y_1) = \nu(y_2) = 1$, the theorem is equivalent to the above stated form of Minkowski's inequality.

Proof. If $p = \infty$, we write $h(y) = \|G(\cdot, y)\|_\infty$, and observe that $|G(x, y)| \le h(y)$ for $\mu \otimes \nu$ almost every (x, y). We conclude $|g(x)| \le \int |h(y)| d\nu(y)$ for almost all x, which proves the claim.

Assume $p < \infty$. It is enough to prove

$$\Big(\int \Big(\int |G(x,y)| d\nu(y)\Big)^p d\mu(x)\Big)^{1/p} \le \int \Big(\int |G(x,y)|^p d\mu(x)\Big)^{1/p} d\nu(y).$$

Since this is finite by assumption, and this implies that $G(x, y)$ is integrable in y for almost all fixed x.

If $p = 1$, the claim follows by Fubini's theorem. If $p > 1$, we prove it using Holder's inequality for exponents p and $p/(p-1)$ in the same way as the special case was done in Probability and Measure. We assume as we may that $G(x, y) \ge 0$, and write

$$\begin{aligned}
\int g(x)^p d\mu(x) &= \int\int G(x,y) d\nu(y) g(x)^{p-1} d\mu(x) \\
&= \int\int G(x,y) g(x)^{p-1} d\mu(x) d\nu(y) \\
&\le \int \Big(\int G(x,y)^p d\mu(x)\Big)^{1/p} \Big(\int (g(x)^{p-1})^{p/(p-1)} d\mu(x)\Big)^{(p-1)/p} d\nu(y) \\
&= \int \|G(\cdot, y)\|_p d\nu(y) \|g\|_p^{p-1}.
\end{aligned}$$

The desired inequality follows upon dividing both sides by $\|g\|_p^{p-1}$. □

In what follows, we study $L^p(\mathbf{R}^d)$, where we consider $\mathbf{R}^d$ endowed with the Lebesgue measure. We introduce a piece of notation which we only need now for some questions on the example sheet, but which we will encounter later also in the lectures. We say that $f : \mathbf{R}^d \to \mathbf{C}$ is locally in L^p, in notation $f \in L^p_{loc}(\mathbf{R}^d)$, if $f|_{B(0,R)} \in L^p(B(0, R))$ for all $R \in \mathbf{R}_{>0}$.

We record a fact that we have already seen for $L^1(\mathbf{R}^d)$. The proof generalizes to $p \in [1, \infty)$ without changes.

Proposition 26. *The set of compactly supported continuous functions $C_c(\mathbf{R}^d)$ is dense in $L^p(\mathbf{R}^d)$ for all $p \in [1, \infty)$.*

The result is not true for $p = \infty$.

Proposition 27. *The space $L^p(\mathbf{R}^d)$ is separable for all $1 \le p < \infty$.*

This is again false for $p = \infty$.

Proof. Let $f \in L^p(\mathbf{R}^d)$ and $\varepsilon > 0$. We will construct a function $g \in L^p(\mathbf{R}^d)$ in such a way that $\|f - g\|_p < \varepsilon$ and we will show that the functions that can arise in the role of g come from a countable collection of family that is independent of f and ε.

We start by using the previous proposition to find a function h : $\mathbf{C}_c(\mathbf{R}^d)$ such that $\|f-h\| < \varepsilon/2$. Since h is compactly supported and continuous, it is uniformly continuous. This means for all $\eta > 0$, there is some $\delta > 0$ such that $|h(x) - h(y)| < \eta$ whenever $|x - y| < \delta$. We choose $n \in \mathbf{Z}_{>0}$ sufficiently large so that the diameter of a cube with side length $1/n$ is less than δ. We subdivide $\mathbf{R}^d$ as a partition of translates of $[0, 1/n)^d$ by the vectors $((1/n)\mathbf{Z})^d$. For each cube K in this partition, we choose a rational number a_K such that $|a_K - h(x)| < \eta$ for all $x \in K$. If h vanishes on K, we take $a_K = 0$. We define $g(x) = a_K$ for $x \in K$ for all K.

We show that $\|g-h\|_p < \varepsilon/2$ if η is small enough. To this end, note that $|g(x) - h(x)| < \eta$ for all $x \in \mathbf{R}^d$. Moreover, the total measure of the cubes on which h does not vanish stays bounded as n grows, because h is compactly supported. In other words, there is a number $C > 0$ such that measure of the points x such that $g(x) \neq h(x)$ is less than C for all n. Therefore, $\|g-h\|_p \le (C\eta)^{1/p} < \varepsilon/2$ if η is sufficiently small.

Now we show that the set of functions that can arise as g is countable. We have countably many choices for n. If n is chosen, there are finitely many cubes on which h does not vanish, and these can be selected in countably many ways. Finally, for each of these finitely many cubes, we need to select a rational number as the value of g on that cube, and this can be done in countably many ways. □

For $a \in \mathbf{R}^d$, we write $\tau_a f(x) = f(x - a)$ for the translation by a operator.

Proposition 28. *Let $p \in [1,\infty)$ and let $f \in L^p(\mathbf{R}^d)$. Then the map $a \mapsto \tau_a f$ is continuous in $L^p(\mathbf{R}^d)$. In other words, $\lim_{b\to a} \|\tau_a f - \tau_b f\|_p = 0$.*

Proof. Fix $\varepsilon > 0$, and let $\widetilde{f} \in C_c(\mathbf{R}^d)$ be with $\|f - \widetilde{f}\|_p < \varepsilon$. Since $\widetilde{f}$ is uniformly continuous, for all $\varepsilon_2 > 0$ there is $\delta > 0$ such that $|\widetilde{f}(x) - \widetilde{f}(y)| < \delta$ if $|x - y| < \varepsilon_2$. Therefore, $|\tau_a\widetilde{f}(x) - \tau_b\widetilde{f}(x)| < \varepsilon_2$ for all x if $|a - b| < \delta$. Moreover, $\tau_a\widetilde{f}(x)f - \tau_b\widetilde{f}(x)$ may be not 0 only if $x - a$ or $x - b$ is in the support of $\widetilde{f}$, and the measure of such points x is bounded by a constant $C = C(\widetilde{f})$ depending only on $\widetilde{f}$. We get

$$\|\tau_a f - \tau_b f\|_p \le \|\tau_a f - \tau_a\widetilde{f}\|_p + \|\tau_a\widetilde{f} - \tau_b\widetilde{f}\|_p + \|\tau_b\widetilde{f} - \tau_b f\|_p \le \varepsilon + \varepsilon_2 C^{1/p} + \varepsilon < 3\varepsilon$$

provided we choose ε_2 sufficiently small depending on ε and $\widetilde{f}$, so ultimately only on ε. This proves the claim. □

Theorem 29. *Let $(E, \mathcal{B}, \mu)$ be a σ-finite measure space. Let $p \in [1,\infty)$, and let q be such that $p^{-1} + q^{-1} = 1$. Then for all $g \in L^q$,*

$$L_g(f) = \int f(x)g(x)d\mu(x)$$

is a bounded linear functional on $L^p(E)$ *with* $\|L_g\| = \|g\|_q$*. Conversely, for every bounded linear functional* L *on* $L^p(E)$*, there is a unique* $g \in L^q(E)$ *such that* $L = L_g$*.*

Proof. The fact that L_g is a linear functional with $\|L_g\| \le \|g\|_q$ follows by the linearity of integration and Holder's inequality. To show that $\|L_g\| \ge \|g\|_q$ if $p > 1$, one may take $f(x) = g(x)^{-1}|g(x)|^q$ (define $f(x) = 0$ if $g(x) = 0$) and observe $\|f\|_p^p = \|g^{p(q-1)}\|_1 = \|g\|_q^q$ using $p(q-1) = q$ and $L_g(f) = \|g\|_q^q = \|f\|_p\|g\|_q$ using $1 + q/p = q$. If $p = 1$ and hence $q = \infty$, we let $A = \{x : |g(x)| > \|L\|\}$ and $f = g^{-1}|_A$. Then $\|f\|_1 < \|L\|^{-1}\mu(A)$ and

$$L(f) = \int fg d\mu = \mu(A),$$

which yields $\mu(A) \le \|L\| \cdot \|L\|^{-1}\mu(A)$ and $\mu(A) = 0$.

Uniqueness of the converse follows by $\|L_{g_1} - L_{g_2}\| = \|L_{g_1-g_2}\| = \|g_1 - g_2\|_q$.

Let L be a bounded linear functional on $L^p(E)$. We show that $L = L_g$ for some $g \in L^q(E)$. We first consider the case $\mu(E) < \infty$. In this case, $L^\infty(E) \subset L^p(E)$, and we may define the set function $\nu(A) = L(1_A)$ for $A \in \mathcal{B}$. Note that ν is finitely additive by the linearity of L. To show σ-additivity, it is enough to show that $\lim \nu(A_n) = 0$ for any sequence of sets $A_n \in \mathcal{B}$ with $\bigcap_n A_n = \varnothing$. This follows from

$$|\nu(A_n)| = |L(1_{A_n})| \le \|L\|\|1_{A_n}\|_p = \|L\|\mu(A_n)^{1/p} \to 0.$$

We proved that ν is a complex measure. If $\mu(A) = 0$, then $1_A = 0$, hence $\nu(A) = L(1_A) = 0$, so ν is absolutely continuous with respect to μ. By the Radon-Nikodym theorem (applied for $(\operatorname{Re}\nu)^\pm, (\operatorname{Im}\nu)^\pm$), there is $g \in L^1(X)$ such that

$$L(1_A) = \nu(A) = \int 1_A g d\mu.$$

Since L and integration are linear, we have $L(f) = \int fg d\mu$ for all simple functions f. We show that this identity also holds for $f \in L^\infty(E)$. Let f_n be a sequence of simple functions that converge to f pointwise and such that $\|f_n\|_\infty \le \|f\|_\infty$. Then $\|f_n - f\|_p \to 0$ by dominated convergence, and by boundedness of L, $L(f_n) \to L(f)$. Similarly, $\int f_n g d\mu \to \int fg d\mu$ again by dominated convergence. Therefore, $L(f) = \int fg d\mu$, indeed.

Now we show $\|g\|_q \le \|L\|$. Suppose $p > 1$ and hence $q < \infty$. Let $R \in \mathbf{R}_{>0}$ and define $g_R(x) = g(x)$ when $|g(x)| \le R$ and $g_R(x) = g(x)R/|g(x)|$ otherwise. Let $f_R(x) = g_R(x)^{-1}g_R^q(x)$. By the calculation we have done before, we have $\|f_R\|_p = \|g_R\|_q^{q-1}$ and

$$L(f_R) = \int f_R g d\mu \ge \int f_R g_R d\mu \ge \|g_R\|_q^q.$$

We conclude $\|g_R\|_q^q \le \|g_R\|_q^{q-1}\|L\|$, hence $\|g_R\|_q \le \|L\|$. Now the claim follows by taking $R \to \infty$. If $p = 1$, $\|g\|_\infty \le \|L\|$ follows by the argument at the beginning of the proof, note that the function f we used there is bounded.

We showed that $L(f) = L_g(f)$ for $f \in L^\infty(E, \mu)$. Since the latter functions are dense in $L^p(E, \mu)$, $L = L_g$ follows.

Now consider the case when μ is only σ-finite. Let $E_1 \subset E_2 \subset \ldots \subset E$ be such that $\bigcup E_n = E$ and $\mu(E_n) < \infty$ for all n. Note that there are natural inclusions $L^p(E_1) \subset L^p(E_2) \subset \ldots \subset L^p(E)$ if we identify the elements of $L^p(E_n)$ with functions in $L^p(E)$ that vanish outside E_n. Applying the case already proved for $L|_{L^p(E_n)}$ we find $g_n \in L^q(E_n)$ such that $L|_{L^p(E_n)} = L_{g_n}$. For $m > n$, we have $L_{g_m}|_{L^p(E_n)} = L_{g_n}$, which by the injectivity of $g \mapsto L_g$ implies $g_m|_{E_n} = g_n$. It follows that there is a function $g : E \to \mathbf{C}$ such that $g|_{E_n} = g_n$ for all n. In addition, we have $\|g_n\|_q = \|L|_{L^p(E_n)}\| \le \|L\|$ so $\|g\|_q \le \|L\|$ follows by the monotone convergence theorem. We have $L(f) = L_g(f)$ for all $f \in \bigcup L^p(E_n)$. Since this space is dense in $L^p(E)$, the property holds for all $f \in L^p(E)$. □

8. Convolution

Let $f, g : \mathbf{R}^n \to \mathbf{C}$ be measurable functions. We define their convolution as

$$f * g(x) = \int f(t)g(x-t)dt$$

provided the integral exists for almost every $x \in \mathbf{R}^d$.

If the integral exists for some x, then

$$f * g(x) = \int f(t)g(x-t)dt = \int f(x-s)g(s)ds = g * f(x)$$

using the substitution $s = x - t$. In addition,

$$\tau_a(f * g) = \tau_a(f) * g = f * \tau_a(g)$$

provided the convolutions exist.

Convolution is a very important operation in different branches of mathematics. It can be used to write down the solution of some PDE's. When f and g are the probability density functions of some independent random variables X and Y then the probability density function of $X + Y$ is $f * g$. In this lecture, the main reason for their interest is that convolution makes functions nicer, as we are going to see.

Theorem 30 (Young's inequality)**.** *Let $p, q, r \in [1, \infty]$ satisfy $1/p + 1/q = 1/r + 1$, and let $f \in L^p(\mathbf{R}^n)$, $g \in L^q(\mathbf{R}^n)$. Then the integral defining $f * g$ exists for almost every x and $f * g \in L^r(\mathbf{R}^n)$. Moreover,*

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

Only the following special cases will be proved in the lectures, which are sufficient for most applications

Proof of the case $p = 1$, $q = r$ via Minkowski's integral inequality. We can write

$$\Big(\int |f * g(x)|^q dx\Big)^{1/q} \le \Big(\int \Big(\int |f(t)g(x-t)|dt\Big)^q dx\Big)^{1/q}$$
$$\le \int \Big(\int |f(t)g(x-t)|^q dx\Big)^{1/q} dt$$
$$= \|g\|_q \int |f(t)|dt$$
$$= \|g\|_q \|f\|_1.$$

This shows, in particular, that $|f(t)g(x-t)|$ is integrable in t for almost all x. □

Proof of the $r = \infty$ case via Holder's inequality. Since $p^{-1} + q^{-1} = 1$, we have

$$|f * g(x)| \le \int |f(t)g(x-t)|dt \le \|f\|_p \|g\|_q$$

by Holder's inequality. □

Note that the proof shows that $\int f(t)g(x-t)dt$ exists and the inequality is valid for all not just almost all x.

We include the proof of the general case for the sake of completeness.

Proof of the general case via Hölder's inequality. Similarly to the previous cases, we may assume that $f(x), g(x) \ge 0$ for all $x \in \mathbf{R}$. Using

$$\frac{1}{r} + \frac{r-p}{pr} + \frac{r-q}{qr} = \frac{1}{r} + \frac{1}{p} - \frac{1}{r} + \frac{1}{q} - \frac{1}{r} = 1,$$

we apply Hölder's inequality as follows

$$\int (f * g(x))^r dx = \int\int f(t)g(x-t)dt(f * g(x))^{r-1}dx$$
$$= \int\int [f(t)^{p/r} g(x-t)^{q/r}][f(t)^{1-p/r}(f * g(x))^{r/p-1}]$$
$$\times [g(x-t)^{1-q/r}(f * g(x))^{r/q-1}]dtdx$$
$$\le \|f(t)^{p/r} g(x-t)^{q/r}\|_r \|f(t)^{1-p/r}(f * g(x))^{r/p-1}\|_{pr/(r-p)}$$
$$\times \|g(x-t)^{1-q/r}(f * g(x))^{r/q-1}\|_{qr/(r-q)}$$

Now we can write

$$\|f(t)^{p/r} g(x-t)^{q/r}\|_r = (\int\int [f(t)^{p/r} g(x-t)^{q/r}]^r dxdt)^{1/r}$$
$$= (\int\int f(t)^p \|g\|_q^q dt)^{1/r} = \|f\|_p^{p/r} \|g\|_q^{q/r}.$$

Using this, and similar identities, we may continue our previous inequality by

$$\int (f * g(x))^r dx \leq \|f\|_p^{p/r}\|g\|_q^{q/r}\|f\|_p^{1-p/r}\|f * g\|_r^{r/p-1}\|g\|_q^{1-q/r}\|f * g\|_r^{r/q-1}$$
$$= \|f\|_p\|g\|_q\|f * g\|_r^{r-1},$$

where we used $r/p + r/q - 2 = r - 1$. We get the claim upon dividing both sides by $\|f * g\|_r^{r-1}$. □

Lemma 31. *Let $f, g \in L^1(\mathbf{R}^d)$ and $h \in L^\infty(\mathbf{R}^d)$. Then we have $(f * g) * h = f * (g * h)$.*

The statement remains true for $f \in L^p(\mathbf{R}^d)$, $g \in L^q(\mathbf{R}^d)$ and $h \in L^r(\mathbf{R}^d)$ if $p^{-1} + q^{-1} + r^{-1} \geq 2$, but we do not prove this.

Proof. The integrals defining the convolutions exist by Young's inequality, and we can write,

$$(f * g) * h(x) = \int f * g(t)h(x - t)dt = \int\int f(s)g(t - s)h(x - t)dsdt.$$

The next step is to apply Fubini's theorem, and to this end, we need to check that

$$\int\int |f(s)g(t - s)h(x - t)|dsdt < \infty,$$

which follows again by Young's inequality, or by direct calculation. Therefore, we can write

$$\begin{aligned}\int\int f(s)g(t - s)h(x - t)dsdt &= \int\int f(s)g(t - s)h(x - t)dtds\\ &= \int\int f(s)g(t)h(x - (t + s))dtds\\ &= \int f(s)(g * h)(x - s)ds\\ &= f * (g * h)(x).\end{aligned}$$

□

Proposition 32. *Let $p, q \in [1, \infty]$ satisfy $p^{-1} + q^{-1} = 1$. Let $f \in L^p(\mathbf{R}^d)$ and $g \in L^q(\mathbf{R}^d)$. Then $f * g \in C(\mathbf{R}^d)$.*

Proof. We assume as we may that $q < \infty$. As we commented above, $f * g(x)$ exists and Young's inequality is valid for all, not just almost all x because $p^{-1} + q^{-1} = 1$. For $x, y \in \mathbf{R}^d$, we can write

$$|f * g(x) - f * g(y)| = |f * g(x) - \tau_{y-x}(f * g)(x)| = |f * (g - \tau_{y-x}g)(x)|$$
$$\leq \|f\|_p\|g - \tau_{y-x}g\|_q \to 0$$

as $y \to x$ by the continuity of $a \mapsto \tau_a g$ in $L^q(\mathbf{R}^d)$. □

Given a so-called multi-index $\alpha \in \mathbf{Z}_{\geq 0}^d$ and a function $f : \mathbf{R}^d \to \mathbf{C}$, we write $|\alpha| = |\alpha_1| + \ldots + |\alpha_d|$ and

$$D^\alpha f(x) = \frac{\partial^{|\alpha|} f(x)}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}$$

whenever this partial derivative exists. The Schwartz space $\mathcal{S} = \mathcal{S}(\mathbf{R}^d)$ consists of functions $f \in C^\infty(\mathbf{R}^d)$ such that

$$\| |x|^n D^\alpha f(x) \|_\infty < \infty$$

for all $n \in \mathbf{Z}_{\geq 0}$ and $\alpha \in \mathbf{Z}_{\geq 0}^d$.

Proposition 33. *Let $f \in L^p(\mathbf{R}^d)$ for some $p \in [1, \infty]$, and let $g \in \mathcal{S}(\mathbf{R}^d)$. Then $f * g \in C^\infty(\mathbf{R}^d)$ and*

$$D^\alpha (f * g) = f * (D^\alpha g)$$

for every multi-index $\alpha \in \mathbf{Z}_{\geq 0}^d$.

The assumption that $g \in \mathcal{S}$ is an overkill.

Proof. It is enough to prove this in the special case $|\alpha| = 1$, and then one can iterate. Suppose $|\alpha| = 1$. Then

$$\begin{aligned} &\frac{f * g(x + t\alpha) - f * g(x)}{t} - f * (D^\alpha g)(x) \\ &= \frac{\tau_{-t\alpha}(f * g)(x) - f * g}{t} - f * (D^\alpha g)(x) \\ &= f * \Big(\frac{\tau_{-t\alpha} g - g}{t} - D^\alpha g \Big)(x). \end{aligned}$$

If we show

$$\Big\| \frac{\tau_{-t\alpha} g - g}{t} - D^\alpha g \Big\|_q \to 0$$

for the exponent q with $p^{-1} + q^{-1} = 1$, the claim will follow by Young's inequality.

Using the mean value theorem twice, we write

$$\frac{\tau_{-t\alpha} g(x) - g(x)}{t} - D^\alpha g(x) = t D^{2\alpha} g(x + s(x)\alpha)$$

for some $0 \leq s(x) \leq t$. Now we observe that

$$|D^{2\alpha} g(x)| \leq C(|x| + 10)^{-d-1}$$

with $C = \|(|x| + 10)^{d+1} D^{2\alpha} g(x)\|_\infty$. This gives

$$\Big\| \frac{\tau_{-t\alpha} g(x) - g(x)}{t} - D^\alpha g(x) \Big\|_q \leq Ct \|(|x| + 9)^{-d-1}\|_q$$

for $0 \leq t \leq 1$. We leave it as an exercise that the L^q norm on the right is finite. □

A sequence of functions $f_n \in L^1(\mathbf{R}^d)$ is called an approximate identity if $f_n(x) \geq 0$ for all $x \in \mathbf{R}^d$ and $\int f_n dx = 1$ for all $n \in \mathbf{Z}_{\geq 0}$ and

$$\int_{|x|>r} f_n(x)dx \to 0 \tag{4}$$

for all $r > 0$.

Lemma 34. *Let $f \in L^1(\mathbf{R}^d)$ be non-negative with $\int f dx = 1$. Then the sequence*

$$f_n(x) = n^d f(nx)$$

is an approximate identity.

Proof. Non-negativity and $\int f_n dx = 1$ is immediate from the definition. We also have

$$\int_{|x|>rn} f(x)dx \to 0$$

by the dominated convergence theorem, and a change of variables gives (4). □

Theorem 35. *Let (f_n) be an approximate identity, and let $p \in [1, \infty)$. Then $\lim \|f_n * g - g\|_p = 0$ for all $g \in L^p(\mathbf{R}^d)$.*

It is useful to think about the special case when $\operatorname{supp} f_n \subset B(0, r)$ with some $r = r(n) \to 0$ as $n \to \infty$. In this case, we can think of

$$f_n * g(x) = \int_{B(x,r)} f_n(x-t)g(t)dt$$

as an average of g in the ball $B(x, r)$ with respect to the weight function f_n. Intuitively, $f_n * g$ "averages out" the oscillations of g at scales below r "without changing the behaviour" of g at coarser scales.

Proof. We reduce first to the case when $f_n(x) = 0$ if $|x| > 1$ for all n. To this end, consider the sequence

$$\widetilde{f}_n(x) = \begin{cases} f_n(x)\Big(\int_{|x|\leq 1} f_n dx\Big)^{-1} & \text{if } |x| \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to check that $\|f_n - \widetilde{f}_n\|_1 \to 0$. Therefore, $\|f_n * g - \widetilde{f}_n * g\|_p \leq \|f_n - \widetilde{f}_n\|_1 \|g\|_p \to 0$ by Young's inequality, and $\|f_n * g - g\|_p \to 0$ if and only if $\|\widetilde{f}_n * g - g\|_p \to 0$.

From this point on, we assume $f_n(x) = 0$ if $|x| > 1$. Fix $\varepsilon > 0$ and let $h \in C_c(\mathbf{R}^d)$ with $\|g - h\|_p \leq \varepsilon$. Since h is uniformly continuous, there is $\delta > 0$ such that $|x - y| < \delta$ implies $|h(x) - h(y)| < \varepsilon$. Now let

n be sufficiently large so that $\int_{|x|>\delta} f_n(x) < \varepsilon$. We can write

$$
\begin{aligned}
|f_n * h(x) - h(x)| &\le \int f_n(t)|h(x-t) - h(x)|dt \\
&= \int_{|t|<\delta} f_n(t)|h(x-t) - h(x)|dt \\
&\qquad + \int_{|t|>\delta} f_n(t)|h(x-t) - h(x)|dt \\
&\le \int f_n(t)\varepsilon dt + \int_{|t|>\delta} f_n(t)\cdot 2\|h\|_\infty dt \\
&\le \varepsilon + \varepsilon \cdot 2\|h\|_\infty.
\end{aligned}
$$

Let $R > 0$ be large enough so that $h(x) = 0$ for $|x| > R$. Note that R is independent of n. Then $h * f_n(x) = 0$ if $|x| > R+1$. We conclude that

$$\|f_n * h - h\|_p \le C_R^{1/p}\varepsilon(1 + 2\|h\|_\infty),$$

where C_R is the volume of the ball of radius $R + 1$.

Finally, we can write

$$\|f_n*g-g\|_p \le \|f_n*(g-h)\|_p+\|f_n*h-h\|_p+\|g-h\|_p \le 2\varepsilon+C_R^{1/p}\varepsilon(1+2\|h\|_\infty).$$

Since $\varepsilon > 0$ can be chosen arbitrarily small, this proves the theorem. □

Proposition 36. *The space of compactly supported infinitely differentiable functions, $C_c^\infty(\mathbf{R}^d)$ is dense in $L^p(\mathbf{R}^d)$ for all $p \in [1,\infty)$.*

Proof. Let $f \in L^p$ and fix $\varepsilon > 0$. Let $g \in L^p$ be of compact support with $\|f - g\|_p < \varepsilon/2$. Let $h \in C^\infty(\mathbf{R}^d)$ be a function with compact support with $h \ge 0$ and $\int h dx = 1$, for example

$$h(x) = \begin{cases} c \cdot e^{\frac{-1}{1-|x|^2}} & \text{if } |x| \le 1 \\ 0 & \text{if } |x| > 1 \end{cases}$$

for a suitable chosen number $c \in \mathbf{R}_{>0}$. Define $h_n(x) = n^d h(nx)$. Then h_n is an approximate identity and we have $\|g - g * h_n\| < \varepsilon/2$ for n large enough, so $\|f - g * h_n\| < \varepsilon$. Now $g * h_n \in C^\infty$ by Proposition 33. It is immediate from the definition of convolution that the convolution of compactly supported functions is also compactly supported. □

9. Hahn-Banach theorem

A Banach space is a vector space X over $\mathbf{C}$ endowed with a map $\|\cdot\| : X \to \mathbf{R}_{\ge 0}$ that satisfies the following properties. First, $\|\cdot\|$ is a norm, that is $\|\lambda x\| = |\lambda|\|x\|$ for $\lambda \in \mathbf{C}$, $x \in X$, $\|x+y\| \le \|x\| + \|y\|$ for $x, y \in X$ and $\|x\| = 0$ is equivalent to $x = 0$. Second, X is a complete metric space with respect to the distance $\|x-y\|$ induced by the norm.

In this course, we are mostly concerned with complex vector spaces, but in this section, we sometimes need to work with spaces over $\mathbf{R}$. If we do not say that a vector space is real, then it is complex.

If X is a Banach space, and there is a sesquilinear form $\langle \cdot, \cdot \rangle : X^2 \to \mathbf{C}$, such that $\|x\|^2 = \langle x, x \rangle$ for all $x \in X$, then X is called a Hilbert space.

Let X be a normed vector space. (Same as above but without assuming completeness.) We write B_X for the unit ball in X, that is the set of elements $x \in X$ with $\|x\| \le 1$. A linear map $f : X \to \mathbf{C}$ is called a linear functional. We define

$$\|f\| = \sup_{x \in B_X} |f(x)|. \tag{5}$$

If $\|f\| < \infty$, we say that f is bounded. An important fact is that f is continuous with respect to the norm topology if and only if f is bounded. The set of bounded linear functionals on X forms a Banach space with respect to the above norm. It is called the dual of X and it is denoted by X'. In what follows we write $f(x) = \langle x, f \rangle$ for $f \in X'$ and $x \in X$.

Theorem 37 (Hahn-Banach)**.** *Let X be a normed vector space, and let $Y \subset X$ be a not (necessarily closed) linear subspace of X. Let $g \in Y'$. Then there is $f \in X'$ that extends g, that is $f(y) = g(y)$ for all $y \in Y$ and $\|f\| = \|g\|$.*

One useful consequence of the Hahn-Banach theorem is that

$$\|x\| = \max_{f \in B_{X'}} |\langle x, f \rangle|. \tag{6}$$

To see this, consider the linear functional g defined on the one dimensional $\mathbf{C} \cdot x$ space by $g(\lambda x) = \lambda$. Observe that $\langle x, g \rangle = 1$, and $\|g\| = 1$. By the Hahn-Banach theorem, we can extend this to a functional $f_0 \in B_{X'}$ such that $\|x\| = \langle x, f_0 \rangle$. On the other hand, $|\langle x, f \rangle| \le \|x\|$ holds for all $f \in B_{X'}$ by the definition of $\|f\|$.

We can think of (6) as a dual of the definition of the norm of a functional. However, the supremum is not realized as a maximum in (5) in general.

We will prove the following more general version of the Hahn-Banach theorem.

Theorem 38 (Hahn-Banach)**.** *Let X be a vector space over $\mathbf{R}$, and let $p(x) : X \to \mathbf{R}_{\ge 0}$ with the following properties*

- *$p(\lambda x) = \lambda p(x)$ for all $\lambda \in \mathbf{R}_{\ge 0}$ and $x \in X$,*
- *$p(x + y) \le p(x) + p(y)$ for all $x, y \in X$.*

Let $Y \subset X$ be a linear subspace of X, and let $g : Y \to \mathbf{R}$ be a linear functional such that $g(x) \le p(x)$ for all $x \in Y$. Then there is a linear functional $f : X \to \mathbf{R}$ that extends g, that is $f(y) = g(y)$ for all $y \in Y$ and $f(x) \le p(x)$ for all $x \in X$.

The assumptions on p are weaker than requiring that p is a norm in that $p(x) = p(-x)$ is not assumed and $p(x) = 0$ may hold for non-zero x. This is not significant, but will come in handy later. Note also the lack of absolute values in the inequalities $g(x) \leq p(x)$ and $f(x) \leq p(x)$.

An explanation is needed to connect this with the previous version, which was stated for complex vector spaces and functionals. Let $Y \subset X$ be complex normed vector spaces, and let $g \in Y'$. Now $\mathrm{Re}(g)$ is a real linear functional on Y if we regard it as a real vector space, which satisfies $\mathrm{Re}(g)(x) \leq p(x)$ with $p(x) := \|x\| \cdot \|g\|$ so we may extend it to a real functional $\widetilde{f} : X \to \mathbf{R}$ satisfying $\widetilde{f}(x) \leq p(x)$. Then $f(x) = \widetilde{f}(x) - i\widetilde{f}(ix) : X \to \mathbf{C}$ is a complex linear functional that extends g. (Exercise: check that $f(ax) = af(x)$ for all $a \in \mathbf{C}$.) In addition, for all $x \in X$, there is $a \in \mathbf{C}$ with $|a| = 1$ such that

$$|f(x)| = af(x) = f(ax) = \widetilde{f}(x) \leq p(x) = \|x\| \cdot \|g\|$$

hence $\|f\| = \|g\|$.

Begining the proof. Let $x \in X \backslash Y$. We show that it is possible to extend g to $Y^+ = Y + \mathbf{R}x$ so that it stays bounded by p. This is a simple calculation, but it is instructive to think about the geometric picture. Write $B = \{x \in X : p(x) \leq 1\}$, and note that this set is convex. The affine halfspace

$$H = \{y \in Y : g(y) \leq 1\}$$

contains $B \cap Y$. To find an extension of g, we need to extend H to a halfspace of Y^+ that contains $B(X) \cap Y^+$.

Now we do the calculation that shows this can be done. Define $g_a(y + \lambda x) = g(y) + a\lambda$, where is $a \in \mathbf{R}$ is a parameter to be chosen. This is a linear functional on Y^+ that extends g. We need to show that a may be chosen in such a way that for all $y + \lambda x \in B \cap Y^+$, we have $g(y) + a\lambda \leq 1$. This is equivalent to

$$\sup_{y,\lambda:\lambda>0,y-\lambda x\in B} \frac{g(y)-1}{\lambda} \leq a \leq \inf_{y,\lambda:\lambda>0,y+\lambda x\in B} \frac{1-g(y)}{\lambda}.$$

If no such a exists, then we may choose $\lambda_1, \lambda_2 > 0$ and $y_1, y_2 \in Y$ such that $y_1 - \lambda_1 x, y_2 + \lambda_2 x \in B$ and

$$\frac{g(y_1)-1}{\lambda_1} > \frac{1-g(y_2)}{\lambda_2},$$

which gives

$$\lambda_2 g(y_1) + \lambda_1 g(y_2) > (\lambda_1 + \lambda_2).$$

But then

$$\frac{\lambda_2}{\lambda_1+\lambda_2}(y_1 - \lambda_1 x) + \frac{\lambda_1}{\lambda_1+\lambda_2}(y_2 - \lambda_2 x) = \frac{\lambda_2 y_1 + \lambda_1 y_2}{\lambda_1+\lambda_2} \in B$$

by the triangle inequality, and

$$g\left(\frac{\lambda_2 y_1 + \lambda_1 y_2}{\lambda_1 + \lambda_2}\right) = \frac{\lambda_2 g(y_1) + \lambda_1 g(y_2)}{\lambda_1 + \lambda_2} > 1,$$

which contradicts the assumption $g(y) \le p(y) \le 1$ for all $y \in B$.

We proved that g can be extended to Y^+ preserving the inequality $g(y) \le p(y)$. Now we want to use an inductive argument to show that g can be extended to the whole space. If the space X is big, we may need uncountably many steps to do this, and we need to use a tool from set theory to complete the argument. This is called the axiom of choice, which has several equivalent forms. The simplest one to use is called Zorn's lemma, which we will now introduce. □

Let A be a set endowed with a relation $\preceq$. We say that $\preceq$ is a partial order if it is

- reflexive, that is $x \preceq x$ for all $x \in A$,
- antisymmetric, that is $x \preceq y$ and $y \preceq x$ implies $x = y$,
- and transitive, that is $x \preceq y$ and $y \preceq z$ implies $x \preceq z$.

A subset $B \subset A$ is totally ordered if at least one of $x \preceq y$ or $y \preceq x$ holds for all $x, y \in B$. An element $m \in B$ is called maximal in B if there is no element $x \neq m \in B$ such that $m \preceq x$. An element $x \in A$ is an upper bound for B if $y \preceq x$ holds for all $y \in B$.

Lemma 39 (Zorn's lemma). *Let $A \neq \varnothing$ be partially ordered by a relation $\preceq$. Assume that all totally ordered subset $B \subset A$ has an upper bound in A. Then A has at least one maximal element.*

Completing the proof of the Hahn–Banach theorem. Let U be the collection of all partial extensions of g to a subspace of X that remain bounded by p. That is, each element h of U is a functional $h : X_h \to \mathbf{R}$, where $Y \subset X_h \subset X$, $h|_Y = g$ and $h(x) \le p(x)$ for all $x \in X_h$. This comes with the natural partial order $h_1 \preceq h_2$ if h_2 extends h_1.

We show that Zorn's lemma can be applied to $(U, \preceq)$. Let $V \subset U$ be a set of partial extensions totally ordered by $\preceq$. We take $X_V : \bigcup_{h \in V} X_h$ and for $x \in X_V$, define $h_V(x) = h(x)$ for any $h \in V$ such that $x \in X_h$. It is easy to check that this is well-defined (it does not matter which h we use in the definition) and it is an extension of g to a linear functional on X_V preserving $h_V \le p$. Moreover, h_V is an upper bound for V with respect to $\preceq$.

We verify linearity as an example. Let $x_1, x_2 \in X_V$. Then there are $h_1, h_2 \in V$ such that $x_j \in X_{h_j}$. Since $\preceq$ is a total order on V, we have $h_1 \preceq h_2$ or $h_2 \preceq h_1$. Suppose it is the former. Then $x_1, x_2 \in X_{h_2}$, and we have

$$h_V(x_1 + x_2) = h_2(x_1 + x_2) = h_2(x_1) + h_2(x_2) = h_V(x_1) + h_V(x_2).$$

Now we can apply Zorn's lemma, and conclude that there is at least one maximal extension $\widetilde{g}$ of g to a functional $\widetilde{Y} \to \mathbf{R}$ for some subspace

$\widetilde{Y} \subset X$ with $\widetilde{g} \leq p$. However, the argument in the first part of the proof applied with $Y = \widetilde{Y}$ shows that if $\widetilde{Y} \subsetneq X$, then there is a larger subspace $Y' \supsetneq \widetilde{Y}$ to which $\widetilde{g}$ can be extended to a functional bounded by p. This contradicts the maximality of $\widetilde{g}$, so we must have $\widetilde{Y} = X$, and the theorem is proved. □

9.1. **Separation theorems.**

Theorem 40. *Let X be a real normed vector space. Let $A, B \subset X$ be convex subsets such that A is open and $A \cap B = \varnothing$. Then there is some $f \in X'$ and a number $\lambda \in \mathbf{R}$ such that $f(A) > \lambda$ and $f(B) \leq \lambda$.*

The geometric meaning of the statement is that the hyperplane $\{x : f(x) = \lambda\}$ separates A and B.

Corollary 41 (Mazur's theorem)**.** *Let X be a real normed vector space. Let $Y \subset X$ be a subsapce and $A \subset X$ an open convex subset such that $Y \cap A = \varnothing$. Then there is some $f \in X'$ such that $f(Y) = 0$ and $f(A) > 0$.*

Geometrically this means that there is a hyperplane containing Y that is disjoint from A.

Proof. By the theorem, there is some $f \in X'$ and some $\lambda \in \mathbf{R}$ such that $f(Y) \leq \lambda$ and $f(A) > \lambda$. Since Y is a subspace and f is linear, $f(Y) = 0$ or $f(Y) = \mathbf{R}$. The second option is ruled out by $f(Y) \leq \lambda$. So $f(Y) = 0$, $\lambda \geq 0$, and the claim follows. □

Corollary 42. *Let X be a real normed vector space and let $Y \subset X$ be a subspace that is not dense, and let $x \in X \backslash \overline{Y}$. Then there is $f \in X'$ such that $f(Y) = 0$ and $f(x) \neq 0$.*

Proof. Let A be an open ball around x that is disjoint from Y. We apply the previous corollary to find f with $f(Y) = 0$ and $f(A) > 0$. □

Proof of the separation theorem. We first consider the special case where $B = \{y\}$ is a single point. Since both the assumptions and the conclusion are translation invariant, there is no harm in assuming that $0 \in A$.

We define

$$p(x) = \inf\{\lambda > 0 : p(\lambda^{-1}x) \in A\}.$$

By a question on the second example sheet, this map satisfies the properties $p(\lambda x) = \lambda p(x)$ for all $\lambda \in \mathbf{R}_{\geq 0}$ and $x \in X$ and $p(x + y) \leq p(x) + p(y)$ for all $x, y \in X$, which are required to apply the Hahn-Banach theorem.

Now consider $g : \mathbf{R}y \to \mathbf{R}$ defined by $g(\lambda y) = \lambda$. Since $y \notin A$, we have $p(y) \geq 1 = g(y)$. By homogeneity, it follows that $p(\lambda y) \geq g(\lambda y)$ for all $\lambda \in \mathbf{R}$. We can apply the Hahn-Banach theorem to find some $f \in X'$ with $f(x) \leq p(x)$ for all $x \in X$ that extends g. For all $x \in A$, we have $f(x) \le p(x) < 1$. On the other hand, $f(y) = 1$, and this proves the theorem in the special case $B = \{y\}$.

In the general case, we consider the set $\widetilde{A} = A - B$. This is the union of translates of A by elements of $-B$, hence it is open. It is also easy to check convexity, which we leave as an exercise. Note that $A \cap B = \varnothing$ implies $0 \notin \widetilde{A}$. We can apply the special case for $\widetilde{A}$ and $\widetilde{B} = \{0\}$ and find f and λ with $f(\widetilde{A}) < \lambda$ and $0 = f(0) \ge \lambda$. Therefore $f(\widetilde{A}) < 0$ and $f(x - y) < 0$ for all $x \in A$ and $y \in B$. We take $\lambda = \sup\{f(x) : x \in A\}$ and note $f(y) \ge \lambda$ for all $y \in B$. Since A is open, $f(A)$ is open by a question on the example sheet, so $f(A) \le \lambda$ implies that $f(A) < \lambda$. (If X was a Banach space, we could use the open mapping theorem.) The proof is complete. □

Remark 43. In the context of Theorem 40 if we assume that A and B are closed convex sets and one of them is compact instead of assuming that A is open, then the conclusion can be upgraded as follows. There is $f \in X'$ and $\lambda_1 < \lambda_2$ such that $f(A) \le \lambda_1$ and $f(B) \ge \lambda_2$, see [1, Theorem 1.7].

10. Weak and weak-∗ topologies

Banach spaces come equipped with a nice topology, the norm topology, but this is not suitable for all purposes. To motivate this, consider the following problem. Given numbers a, b, A, what is the function $f \in C^1([0,1])$ such that $f(0) = a$, $f(1) = b$, the area under its graph is A and the graph of f is of the shortest possible length subject to these constraints?

The answer is a circular arc, at least when the prescribed area is in a suitable range depending on a and b. To prove this, one may attempt to show that if the graph of f is not a circular arc, then one can find suitable perturbation $f + g$ that satisfies the constraints while reducing the length. However, such an argument has a crucial gap. The argument shows that only the circular arc can be the minimizer, but what if the minimum is not attained?

There is one powerful tool in analysis that guaranties the existence of a minimizer. A continuous (or even a lower semi-continuous) function attains its minimum on a compact set. In finite dimensional normed spaces, closed bounded sets are compact thanks to the Bolzano–Weierstrass theorem. However, this fails in infinite dimensional normed spaces. If the closed unit ball is compact in a normed space, the space is finite dimensional.

For this reason, it is useful to introduce other topologies on infinite dimensional vector spaces. The following definition gives a general procedure for this.

Definition 44. Let X be a vector space over $\mathbf{C}$. A seminorm, is a map $p : X \to \mathbf{R}_{\ge 0}$ such that $p(ax) = |a|p(x)$ for all $a \in \mathbf{C}$ and $x \in X$

and $p(x+y) \le p(x) + p(y)$ for all $x, y \in X$. (This is weaker than a norm in that we do not require $p(x) > 0$ for $x \neq 0$.)

Given a family of seminorms $\mathcal{P}$, we can define a topology on X as follows. For any $n \in \mathbf{Z}_{\ge 1}$, $p_1, \ldots, p_n \in \mathcal{P}$ and $\varepsilon > 0$, we define a neighbourhood of a point x as

$$N(x; p_1, \ldots, p_n, \varepsilon) := \{y \in X : p_j(x-y) < \varepsilon \text{ for } j = 1, \ldots, n\}.$$

This collection of sets for all choices of n, $p_1, \ldots, p_n$ and ε form a neighbourhood base of x in a topology, which we call the topology induced by $\mathcal{P}$. A set $U \subset X$ is open in this topology if and only if for all $x \in U$ there are $n, p_1, \ldots, p_n$ and ε such that $N(x; p_1, \ldots, p_n, \varepsilon) \subset U$.

We leave it as an exercise to check that the topology described above is indeed a topology, and that the sets $N(x; p_1, \ldots, p_n, \varepsilon)$ are open sets in it.

We say that a family of seminorms $\mathcal{P}$ is separating if for all $x \in X$, there is $p \in \mathcal{P}$ such that $p(x) > 0$.

Lemma 45. *If $\mathcal{P}$ is a separating family of seminorms on a vector space X, then the induced topology is Hausdorff.*

Proof. Let $x \neq y \in X$ and let $p \in \mathcal{P}$ such that $p(x-y) = 2\varepsilon > 0$. Then $N(x; p, \varepsilon)$ and $N(y; p, \varepsilon)$ are disjoint open sets containing x and y respectively. ☐

When $\mathcal{P} = \{p_1, p_2, \ldots\}$ is countable then the topology induced is metrizable with the metric

$$\operatorname{dist}(x, y) = \sum_{n=1}^{\infty} \min(p_n(x-y), 2^{-n}).$$

We leave it as an exercise to show that this metric induces the same topology. However, in general, there is no norm on X that induces the same topology. If $\mathcal{P}$ is uncountable, then the induced topology may be non-metrizable.

Now we consider a few examples, we will encounter more later in the course. When $\mathcal{P} = \{\|\cdot\|\}$ for a single norm $\|\cdot\|$, then the induced topology is simply the norm topology.

Given a normed space X, the family

$$\mathcal{P} = \{x \mapsto |\langle x, f\rangle| : f \in X'\}$$

is a separating family of seminorms thanks to the Hahn-Banach theorem. The induced topology is called the weak topology on X. For convergence of sequences in this topology we will use the notation $\operatorname{w-lim} x_n = x$ or $x_n \rightharpoonup x$.

On the dual X' of a normed vector space, the family

$$\mathcal{P} = \{f \mapsto |\langle x, f\rangle| : x \in X\}$$

is a separating family of seminorms, because a non-zero functional evaluates to a non-zero number at some element x. The induced topology is called the weak-$*$ topology on X'. For convergence of sequences in this topology we will use the notation w*-lim $f_n = f$ or $f_n \overset{*}{\rightharpoonup} f$.

In what follows, we write $N(x; f_1, \ldots, f_n, \varepsilon)$ and $N(f; x_1, \ldots, x_n, \varepsilon)$ for the neighbourhoods in the weak and weak-$*$ topologies that correspond to the seminorms $x \mapsto |\langle x, f_j \rangle|$ and $f \mapsto |\langle x_j, f \rangle|$, respectively.

On the dual space X' we can also consider the weak topology if we apply the definition with X' in the place of X. This topology is induced by the seminorms $f \mapsto |\varphi(f)|$ for $\varphi \in X''$.

For each element $x \in X$, there corresponds of functional $J(x) \in X''$ given by $J(x)f = \langle x, f \rangle$. It is easy to check that for each x, the map $J(x) : X' \to \mathbf{C}$ is indeed linear, and we have

$$\|J(x)\|_{X''} = \sup_{f \in B_{X'}} |J(x)f| = \sup_{f \in B_{X'}} |\langle x, f \rangle| = \|x\|_X.$$

Therefore, J embeds X isometrically in X''.

The weak-$*$ topology on X' is induced by the seminorms $f \mapsto |\varphi(f)|$, where $\varphi \in J(X) \subset X''$, which is a subset of the seminorms used to define the weak topology. Therefore, the weak-$*$ topology is coarser then the weak topology, that is, it contains fewer open sets, it is easier for a sequence to converge and it is more difficult for a function defined on X' to be continuous. If $J(X) = X''$, we say that X is a reflexive Banach space. In this case, the weak and weak-$*$ topologies coincide on X'. (The converse is also true.) Moreover, in this case, we can identify X with the dual of X' and the weak topology on X will coincide with the weak-$*$ topology when we identify X with the dual of X' via J.

The next statement collects some basic properties of weak and weak-$*$ topologies and their relationship to the norm topology.

Proposition 46. *Let X be a Banach space.*

1. *The weak and weak-$*$ topologies are Hausdorff.*
2. *The map $x \mapsto \langle x, f \rangle$ is continuous in the weak topology for all $f \in X'$ and the map $f \mapsto \langle x, f \rangle$ is continuous in the weak-$*$ topology for all $x \in X$.*
3. *Let $x, x_1, x_2, \ldots \in X$. Then*
$$\text{w-lim}\, x_n = x \quad \textit{if and only if} \quad \lim \langle x_n, f \rangle = \langle x, f \rangle$$
for all $f \in X'$.
4. *If* $\lim x_n = x$ *in the norm topology, then* w-lim $x_n = x$.
5. *Let $f, f_1, f_2, \ldots \in X'$. Then*
$$\text{w*-lim}\, f_n = f \quad \textit{if and only if} \quad \lim \langle x, f_n \rangle = \langle x, f \rangle$$
for all $f \in X'$.
6. *If* $\lim f_n = f$ *in the norm topology of X' then* w*-lim $f_n = f$.

(7) *Let* $x, x_1, x_2, \ldots \in X$. *If* $\text{w-lim}\, x_n = x$, $\|x_n\|$ *is bounded and* $\liminf \|x_n\| \geq \|x\|$.
(8) *Let* $f, f_1, f_2, \ldots \in X$. *If* $\text{w*-lim}\, f_n = f$, $\|f_n\|$ *is bounded and* $\liminf \|f_n\| \geq \|f\|$.
(9) *Let* $x, x_1, x_2, \ldots \in X$, *and* $f, f_1, f_2, \ldots \in X'$. *Suppose that* $\text{w-lim}\, x_n = x$ *and* $\lim f_n = f$ *or* $\lim x_n = x$ *and* $\text{w*-lim}\, f_n = f$. *Then*

$$\lim \langle x_n, f_n \rangle = \langle x, f \rangle.$$

Before we give the proof some remarks are in order. The weak and respectively the weak-$*$ topologies are the coarsest (that is the one that contains the fewest open sets) that make the maps $x \mapsto \langle x, f \rangle$ and respectively $f \mapsto \langle x, f \rangle$ continuous.

These topologies are not metrizable if X is infinite dimensional, that is, there is no metric that give rise to these topologies. This means that these topologies cannot be characterized by the convergence of sequences. For example, a set that contains all weak limits of sequences of its elements may not be closed in the weak topology.

Item (5) says that a sequence of functionals converge in the weak-$*$ topology if and only if they converge pointwise on X. By linearity, this is equivalent to pointwise convergence on the unit ball B_X. In contrast, norm convergence is equivalent to uniform convergence on B_X.

In the proof of Items (7) and (8), we use the uniform boundedness principle (Banach–Steinhaus theorem) from Part II Linear Analysis, which we recall.

Theorem 47. *Let* $T_n : X \to Y$ *be a sequence of bounded linear maps from a Banach space* X *to a normed space* Y. *Suppose that* $\lim_{n\to\infty} T_n(x) = T(x)$ *for all* $x \in X$, *where* $T : X \to Y$ *is a map. Then* T *is linear and bounded,* $\sup \|T_n\| < \infty$ *and* $\|T\| \leq \liminf \|T_n\|$.

The last property was only given as an exercise in Part II Linear Analysis. Here is how to do it. For every $x \in B_X$, we can write

$$\|Tx\| = \lim \|T_n x\| \leq \liminf \|T_n\|.$$

Now taking the supremum over x gives the required inequality.

Proof. (1) We have already proved this.

(2) Let $U \subset \mathbf{R}$ be open. We show that $f^{-1}(U) \subset X$ is open in the weak topology. Let $x \in f^{-1}(U)$. Then $f(x) \in U$ so there is some $\varepsilon > 0$ such that $(f(x) - \varepsilon, f(x) + \varepsilon) \subset U$. Now we see that $N(x; f, \varepsilon) \subset f^{-1}(U)$. Since x was arbitrary, $f^{-1}(U)$ is open.

The proof for the weak-$*$ topology is the same.

(3) Suppose $\text{w-lim}\, x_n = x$ and let $f \in X'$. Then $\langle x_n, f \rangle \to \langle x, f \rangle$ by Item (2).

Now suppose $\lim \langle x_n, f \rangle = \langle x, f \rangle$ for all $f \in X'$. Fix some $f_1, \ldots, f_k$ and $\varepsilon > 0$. If n is large enough depending on $f_1, \ldots, f_k, \varepsilon$, then $|\langle x_n -$

$x, f_j\rangle| < \varepsilon$ for $j = 1, \dots, k$, hence $x_n \in N(x, f_1, \dots, f_k, \varepsilon)$. Therefore w-lim $x_n = x$.

(4) Suppose $x_n \to x$, and let $f \in X'$. Then

$$|\langle x_n - x, f\rangle| \le \|x_n - x\|\|f\| \to 0,$$

hence $\lim\langle x_n, f\rangle = \langle x, f\rangle$. Since f was arbitrary, item (3) implies $x_n \rightharpoonup x$.

(5) Same as (3).

(6) Same as (4).

(7) Suppose w-lim $x_n = x$. By the discussion above, item (3) implies that $J(x_n)f = J(x)f$ for all $f \in X'$. That is, the sequence of functionals $J(x_n)$ converge pointwise to $J(x)$. By the uniform boundedness principle, $\|x_n\| = \|J(x_n)\|$ is bounded and

$$\|x\| = \|J(x)\| \le \liminf \|J(x_n)\| = \liminf \|x_n\|.$$

(8) Same as (7), just easier. We can apply the uniform boundedness principle directly for f_n.

(9) Suppose $x_n \rightharpoonup x$ and $f_n \to f$. Then

$$\begin{aligned}|\langle x_n, f_n\rangle - \langle x, f\rangle| &\le |\langle x_n - x, f\rangle| + |\langle x_n, f_n - f\rangle| \\ &\le |\langle x_n - x, f\rangle| + (\sup \|x_n\|)\|f_n - f\| \to 0.\end{aligned}$$

Here we used that $(\sup \|x_n\|)$ is bounded by item (7).

The proof of the other claim is the same. □

Example 48. Let E be a compact metric space. Let $\mu, \mu_1, \mu_2, \dots$ be complex Borel measures on E. Then $\mu_n \overset{*}{\rightharpoonup} \mu$ if and only if

$$\int f d\mu_n \to \int f d\mu$$

for all $f \in C(E)$.

Let $p, q \in [0, \infty]$ with $p^{-1}+q^{-1} = 1$ and $p > 1$. Let $f, f_1, f_2, \dots L^p(\mathbf{R}^d)$. Then $f_n \overset{*}{\rightharpoonup} f$ if and only if

$$\int f_n g dx \to \int f g dx$$

for all $g \in L^q(\mathbf{R}^d)$.

Let H be a Hilbert space and $x, x_1, x_2, \dots \in H$. Then $x_n \rightharpoonup x$ (or equivalently $x_n \overset{*}{\rightharpoonup} x$) if and only if

$$\langle x_n, y\rangle \to \langle x, y\rangle$$

for all $y \in H$.

Here are a few examples of weakly convergent sequences in $L^2(\mathbf{R})$ that are not norm convergent:

(1) $f_n(x) = \exp(inx) \cdot 1_{[0,1]}$,
(2) $f_n(x) = 2^{n/2} \cdot 1_{[0,2^{-n}]}$,
(3) $f_n(x) = 1_{[n,n+1]}$.

We leave it as an exercise to show that each of these sequences converge to 0 in the weak topology, but they are not Cauchy sequences in the norm topology. (For the first example, the Riemann-Lebesgue lemma to be discussed later will be helpful.) In fact, $\|f_n\|_2 = 1$ in all three cases for all n, but the weak limit has norm 0. It turns out that in Hilbert spaces the only way a weakly convergent sequence can fail to be norm convergent is if the norm of the weak limit drops compared to the norm of the elements of the sequence.

Theorem 49. *Let X be a Hilbert space, and let $x, x_1, x_2, \dots$ be such that $\text{w-lim}\, x_n = x$. Then x_n converges in the norm topology if and only if $\lim \|x_n\| = \|x\|$. If this is the case the limit in the norm topology is also x.*

The proof of this result is a question on an example sheet. It remains true for sufficiently nice, so called, uniformly convex Banach spaces. For the definition of what this means and proofs, see [1, Chapter 3.7].

A Banach space X is separable if it has a countable dense subset (with respect to the norm topology). As mentioned above, the weak and weak-$*$ topologies are not metrizable on infinite dimensional Banach spaces. However, the restriction of the weak-$*$ topology to the unit ball $B_{X'}$ of the dual of a separable Banach space is metrizable, as the next result shows.

Proposition 50. *Let X be a separable Banach space, and let $x_1, x_2, \dots \in B_X$ be a countable dense subset of B_X. Then*

$$\operatorname{dist}(f, g) = \sum_{n=1}^{\infty} \frac{|\langle x_n, f - g\rangle|}{2^n}$$

is a metric on $B_{X'}$ and it generates the restriction of the weak-$*$ topology on $B_{X'}$.

The proof of this result is a question on the example sheet.

Theorem 51 (Banach-Alaoglu)**.** *Let X be a Banach space. Then $B_{X'}$ is compact in the weak-$*$ topology.*

The usual proof of this result embeds $B_{X'}$ into $\{z : |z| \le 1\}^{B_X}$, that is, we take a copy of $\{z : |z| \le 1\}$ for each element of B_X and then take their product. The embedding is via the map $f \mapsto (\langle x, f\rangle)_x$. Then it can be checked that the image of $B_{X'}$ is closed in the product topology, and the embedding is a a homeomorphism between $B_{X'}$ with the weak-$*$ topology and its image with the restriction of the product topology. Then the theorem follows from Tychonoff's theorem that an arbitrary product of compact topological spaces is compact. A proof of this for finite products was given in Part II Analysis and topology.

Instead of this, we give a proof only in the case where X is separable, which is more instructive. In this case, the weak-$*$ topology on $B_{X'}$ is metrizable, hence compactness is equivalent to sequential compactness. Therefore, it is enough to prove the following result.

Theorem 52. *Let X be a separable Banach space. Then an arbitrary sequence $f_1, f_2, \ldots \in B_{X'}$ has a weak-$*$ convergent subsequence.*

Proof. Let $x_1, x_2 \ldots$ be a countable dense subset of B_X. We define a subsequence $f_j^{(m)}$ of $f_1, f_2, \ldots$ for each $m \in \mathbf{Z}_{\geq 1}$. Let $f_j^{(1)}$ be a subsequnce of $f_1, f_2, \ldots$ such that $\langle x_1, f_j^{(1)} \rangle$ converges to some number $\alpha_1 \in [-1, 1]$. Such a subsequence exists by the Bolzano–Weierstrass theorem. If $m > 1$ and $f_j^{(m-1)}$ is already defined, we define $f_j^{(m)}$ to be a subsequence of $(f_j^{(m-1)})_j$ such that $\langle x_m, f_j^{(m)} \rangle$ converges to some number $\alpha_m \in [-1, 1]$.

Now we take $g_n = f_n^{(n)}$, and observe that this is a subsequence of $(f_j^{(m)})$ for all m, hence $\lim_n \langle x_m, g_n \rangle = \alpha_m$ for all m.

Now let $x \in B_X$ be arbitrary and fix some $\varepsilon > 0$. Let x_m be such that $\|x - x_m\| < \varepsilon/10$. Then there is some N such that $|\langle x_m, g_{n_1} - g_{n_2} \rangle| < \varepsilon/10$ for all $n_1, n_2 > N$, hence

$$|\langle x, g_{n_1} - g_{n_2} \rangle| \leq 2\|x - x_m\| + |\langle x_m, g_{n_1} - g_{n_2} \rangle| < \varepsilon.$$

This shows that $\langle x, g_n \rangle$ is Cauchy and hence convergent. By homogeneity, $\langle x, g_n \rangle$ is convergent for all $x \in X$ not just on B_X.

It remains to show that the limit $h(x) := \lim \langle x, g_n \rangle$ is a linear functional in $B_{X'}$ and this follows from the uniform boundedness principle, Theorem 47. ☐

11. Fourier transform

For $f \in L^1(\mathbf{R}^d)$, its Fourier transform is defined by

$$\widehat{f}(\xi) = \int f(x) e^{-2\pi i \langle x, \xi \rangle} dx.$$

Intuitively, one may think about $\widehat{f}(\xi)$ as the "coefficient" of $e^{2\pi i \langle x, \xi \rangle}$ in a "decomposition" of f. The reason why this is interesting is that $e^{2\pi i \langle x, \xi \rangle}$ is an eigenfunction of the translation operator τ_a for all a. Indeed,

$$\tau_a e^{2\pi i \langle x, \xi \rangle} = e^{-2\pi i \langle a, \xi \rangle} \cdot e^{2\pi i \langle x, \xi \rangle}.$$

It is a fact of linear algebra that the eigenspaces of a linear transformation are invariant under any linear transformation that commutes with the first one. Motivated by this, we expect that the functions $e^{2\pi i \langle x, \xi \rangle}$ will be eigenfunctions of linear maps that commute with translations. So the Fourier transform will "diagonalize" such linear maps. There are many important linear maps in analysis that commute with translations, including differentiation and convolution. It is difficult to make this informal discussion rigorous, because the functions $e^{2\pi i \langle x, \xi \rangle}$ are not in $L^2(\mathbf{R}^d)$ which would be the most natural Hilbert space to work with. Nevertheless we will see that Fourier transform turns differentiation and convolution into pointwise multiplication, as expected.

Proposition 53. *Let $f, g \in L^1(\mathbf{R}^d)$. Then*

1. $\widehat{f} \in C(\mathbf{R}^d) \cap L^\infty(\mathbf{R}^d)$ *and* $\|f\|_\infty \leq \|f\|_1$,
2. $\widehat{(\lambda f)} = \lambda \widehat{f}$ *for all* $\lambda \in \mathbf{C}$,
3. $\widehat{f + g} = \widehat{f} + \widehat{g}$,
4. $\widehat{\tau_a f} = e^{-2\pi i \langle a, \xi \rangle} \widehat{f}$ *for all* $a \in \mathbf{R}^d$,
5. $\widehat{f \circ U} = |\det(U)|^{-1} \widehat{f} \circ (U^{-1})^T$ *for all invertible* $U \in \mathbf{R}^{d \times d}$,
6. $\widehat{\overline{f}} = \overline{\check{\widehat{f}}}$,
7. $\widehat{f * g} = \widehat{f} \cdot \widehat{g}$.

In item (6), we use the notation $\check{f}(x) = f(-x)$.

Proof. (1) We have

$$|\widehat{f}(\xi)| \leq \int |f(x) e^{2\pi i \langle x, \xi \rangle}| dx = \|f\|_1.$$

Continuity follows from

$$|f(x) e^{2\pi i \langle x, \xi \rangle}| \leq |f(x)| \in L^1(\mathbf{R}^d),$$

the continuity of the integrand in ξ and dominated convergence.

(2) Follows from the homogeneity of integration.

(3) Follows from additivity of integration.

(4) Unwinding the definitions and then using translation invariance of Lebesgue measure, we write

$$\begin{aligned}\widehat{\tau_a f}(\xi) = \int f(x-a) e^{-2\pi i \langle x, \xi \rangle} dx &= e^{-2\pi i \langle a, \xi \rangle} \int f(x-a) e^{-2\pi i \langle (x-a), \xi \rangle} dx \\ &= e^{-2\pi i \langle a, \xi \rangle} \int f(x) e^{-2\pi i \langle x, \xi \rangle} dx \\ &= e^{-2\pi i \langle a, \xi \rangle} \widehat{f}(\xi).\end{aligned}$$

(5) Using the change of variables $y = Ux$, we write

$$\begin{aligned}\widehat{f \circ U}(\xi) = \int f(Ux) e^{-2\pi i \langle x, \xi \rangle} dx &= \int f(Ux) e^{-2\pi i \langle Ux, (U^{-1})^T \xi \rangle} dx \\ &= \int f(y) e^{-2\pi i \langle y, (U^{-1})^T \xi \rangle} |\det U|^{-1} dy \\ &= |\det U|^{-1} \widehat{f}((U^{-1})^T \xi).\end{aligned}$$

(6) Exchanging complex conjugation and integration, we write

$$\widehat{\overline{f}}(\xi) = \int \overline{f}(x) e^{-2\pi i \langle x, \xi \rangle} dx = \int \overline{f(x) e^{-2\pi i \langle x, -\xi \rangle}} dx = \overline{\widehat{f}}(-\xi).$$

(7) Using Fubini's theorem and then translation invariance in the x variable we write

$$\begin{aligned}\widehat{f*g}(\xi) &= \int\int f(t)g(x-t)dte^{-2\pi i\langle x,\xi\rangle}dx\\ &= \int\int f(t)g(x-t)e^{-2\pi i\langle x,\xi\rangle}dxdt\\ &= \int\int f(t)e^{-2\pi i\langle t,\xi\rangle}g(x-t)e^{-2\pi i\langle x-t,\xi\rangle}dxdt\\ &= \int\int f(t)e^{-2\pi i\langle t,\xi\rangle}g(x)e^{-2\pi i\langle x,\xi\rangle}dxdt\\ &= \widehat{f}(\xi)\widehat{g}(\xi).\end{aligned}$$

To justify the use of Fubini, we note that $\int\int |f(t)g(x-t)|dtdx = \int |f|*|g|(x)dx < \infty$ by Young's inequality. $\square$

Recall the definition of the space $\mathcal{S}(\mathbf{R}^d)$ of Schwartz functions and the notation D^α from Section 8. Given a multiindex $\alpha \in \mathbf{Z}^d_{\geq 0}$ and a vector $\xi = (\xi_1, \ldots, \xi_d)$, we write $\xi^\alpha = \xi_1^{\alpha_1} \cdots \xi_d^{\alpha_d}$.

Proposition 54. *Let $f \in \mathcal{S}(\mathbf{R}^d)$ and let $\alpha \in \mathbf{Z}^d_{\geq 0}$. Then*

$$\widehat{D^\alpha f}(\xi) = (2\pi i)^{|\alpha|}\xi^\alpha \widehat{f}(\xi).$$

Proof. It is enough to prove this in the case $|\alpha| = 1$, and then one may iterate this case to get the general case. For simplicity, we assume $D^\alpha = \partial/\partial x_1$.

We use integration by parts (from Riemann integration theory, say) to write

$$\begin{aligned}\int_a^b \frac{\partial f}{\partial x_1}(x)e^{-2\pi i\langle x,\xi\rangle}dx_1 =& f(b,x_2,\ldots,x_d)e_{b,x} - f(a,x_2,\ldots,x_d)e_{a,x}\\ &- (-2\pi i\xi_1)\int_a^b f(x)e^{-2\pi i\langle x,\xi\rangle}dx_1\end{aligned}$$

for all $a, b \in \mathbf{R}$ and $x \in \mathbf{R}^d$, where $e_{a,x}, e_{b,x}$ are some complex numbers of modulus 1 depending on a, b, x.

Taking $a, b \to \infty$, we get

$$\int_{\mathbf{R}} \frac{\partial f}{\partial x_1}(x)e^{-2\pi i\langle x,\xi\rangle}dx_1 = 2\pi i\xi_1 \int_{\mathbf{R}} f(x)e^{-2\pi i\langle x,\xi\rangle}dx_1$$

in the limit. Here we used that $|f(x)|, |\partial f(x)/\partial x_1| \leq C(1+|x|)^{-(d+1)}$ for some C, so $f(a,x_2,\ldots,x_d), f(b,x_2,\ldots,x_d) \to 0$ and the integrals converge by dominated convergence. We integrate out the rest of the variables, and using Fubini, we get the claim. $\square$

Proposition 55. *Let $f \in \mathcal{S}(\mathbf{R}^d)$. Then $\widehat{f} \in C^\infty(\mathbf{R}^d)$ and*

$$D^\alpha \widehat{f} = ((-2\pi i)^{|\alpha|}x^\alpha f)^\wedge$$

Proof. It is enough to prove that partial derivatives of order 1 exists and they are given by the claimed formula. Then item (1) in Proposition 53 implies continuity of the partial derivatives, which implies that $f \in C^1(\mathbf{R}^d)$ and one may iterate this to conclude the claim for all α.

For simplicity suppose that $D^\alpha = \partial/\partial\xi_1$. Then we can write

$$\frac{\partial \widehat{f}}{\partial \xi_1} = \frac{\partial}{\partial \xi_1}\int f(x)e^{-2\pi i\langle x,\xi\rangle}dx = \int f(x)\frac{\partial e^{-2\pi i\langle x,\xi\rangle}}{\partial \xi_1}dx$$
$$= \int f(x)(-2\pi i x_1)e^{-2\pi i\langle x,\xi\rangle}dx.$$

Differentiation under the integral can be justified by estimating the difference quotient using the mean value theorem, the decay of Schwartz functions and the dominated convergence theorem. □

Proposition 56. *If $f \in \mathcal{S}(\mathbf{R}^d)$, then $\widehat{f} \in \mathcal{S}(\mathbf{R}^d)$.*

Proof. We have already seen that $\widehat{f} \in C^\infty(\mathbf{R}^d)$. It remains to show that $|x|^n D^\alpha \widehat{f}(\xi)$ is bounded for all $\alpha \in \mathbf{Z}^d_{\geq 0}$ and $n \in \mathbf{Z}_{\geq 0}$. Since any partial derivative of the Fourier transform of a Schwartz function is the Fourier transform of a Schwartz function by Proposition 55, it is enough to prove the claim for $\alpha = 0$.

Proposition 54 gives

$$\widehat{\Delta f}(\xi) = ((2\pi i\xi_1)^2 + \ldots + (2\pi i\xi_d)^2)\widehat{f}(\xi) = -4\pi^2|\xi|^2\widehat{f}(\xi).$$

Here $\Delta = \sum_j \partial^2/\partial x_j^2$ is the Laplace operator. Iterating this, we get $\widehat{\Delta^k f}(\xi) = (-4\pi^2)^k|\xi|^{2k}\widehat{f}(\xi)$. Item (1) in Proposition 53 implies that $|\xi|^{2k}\widehat{f}(\xi)$ is bounded for all $k \in \mathbf{Z}_{\geq 0}$, and the claim follows. □

Observe that to prove differentiability of $\widehat{f}$, we used that f decays fast as $|x| \to \infty$. To show that $\widehat{f}$ decays we used differentiability. There is a duality between smoothness and decay properties of a function and it Fourier transform. The smoother f is, the faster $\widehat{f}$ decays and vice versa. The assumption $f \in \mathcal{S}(\mathbf{R}^d)$ in the above two propositions is an overkill. We will come back to this point later.

Theorem 57 (Riemann-Lebesgue lemma)**.** *Let $f \in L^1(\mathbf{R}^d)$. Then $\widehat{f}(\xi) \to 0$ as $|\xi| \to \infty$.*

Proof. Fix $\varepsilon > 0$, and let $g \in C^\infty$ be a compactly supported function with $\|f - g\|_1 < \varepsilon$. Then $\widehat{g}$ is a Schwartz function and hence $\widehat{g}(\xi) \to 0$ as $\xi \to \infty$. By (1) in Proposition 53, $\|\widehat{f} - \widehat{g}\|_\infty < \varepsilon$, hence $\limsup_{|\xi|\to\infty} |\widehat{f}(\xi)| < \varepsilon$. The claim follows by taking $\varepsilon \to 0$. □

Theorem 58 (Fourier inversion)**.** *Let $f \in L^1(\mathbf{R}^d)$ be such that $\widehat{f} \in L^1(\mathbf{R}^d)$. Then*

$$f(y) = \int \widehat{f}(\xi)e^{2\pi i\langle y,\xi\rangle}d\xi.$$

We may write

$$\int \widehat{f}(\xi)e^{2\pi i\langle y,\xi\rangle}d\xi = \int\int f(x)e^{-2\pi i\langle x,\xi\rangle}dx e^{2\pi i\langle y,\xi\rangle}d\xi$$
$$= \int\int f(x)e^{2\pi i\langle y-x,\xi\rangle}dxd\xi.$$

Then we would like to apply Fubini's theorem to switch the order of integration, and argue that the integral of $e^{-2\pi i\langle y-x,\xi\rangle}$ in ξ is 0 when $y-x\neq 0$ and it is ∞ when $y-x=0$. Then we hope that integrating in x would give $f(y)\cdot\infty\cdot 0 = f(y)$. This does not make any sense, of course, because $f(x)e^{-2\pi i\langle y-x,\xi\rangle}$ is not in L^1, and $e^{-2\pi i\langle y-x,\xi\rangle}$ is not integrable on $\mathbf{R}$, not to mention the final equation. However, this vague heuristic can be turned into a proof if we multiply the integrand with an L^1 function of ξ.

Proof. Let $g\in L^1(\mathbf{R}^d)$, and write

$$\int \widehat{f}(\xi)e^{2\pi i\langle y,\xi\rangle}g(\xi)d\xi = \int\int f(x)e^{2\pi i\langle y-x,\xi\rangle}g(\xi)dxd\xi$$
$$= \int\int f(x)e^{2\pi i\langle y-x,\xi\rangle}g(\xi)d\xi dx$$
$$= \int f(x)\widehat{g}(x-y)dx$$
$$= f*\check{\widehat{g}}(y).$$

Now we want to pick a sequence g_n with the properties that $\|g_n\|_\infty\leq 1$ for all n and $g_n(\xi)\to 1$ for (almost) all ξ. Then the dominated convergence theorem gives

$$\int \widehat{f}(\xi)e^{2\pi i\langle y,\xi\rangle}d\xi = \lim_{n\to\infty}\int \widehat{f}(\xi)e^{2\pi i\langle y,\xi\rangle}g_n(\xi)d\xi.$$

In addition to this, we want to choose g_n in such a way that $\check{\widehat{g}}_n$ is an approximate identity, then $f*\check{\widehat{g}}_n$ converges in L^1 to f. By passing to a subsequence if necessary, we also have convergence for almost every y and the claim follows.

It remains to construct the sequence g_n. One way to do this is to pick an explicit sequence, for example the Gaussians:

$$g_n(x) = e^{-\pi|x/n|^2}, \qquad \widehat{g}_n(\xi) = n^d e^{-\pi|n\xi|^2}.$$

It is immediate from the definition that $\|g_n\|_\infty\leq 1$ and $g_n(x)\to 1$ for all x. Lemma 34 implies that $\widehat{g}_n$ is an approximate identity if we show that $\int\widehat{g}_1(\xi)d\xi = 1$. Since $\widehat{g}_1 = g_1$, this actually follows by

$$\int \widehat{g}_1 dx = \int g_1 dx = \widehat{g}_1(0) = 1.$$

We do not prove the formula stated above for $\widehat{g}_n$ in this course. □

Proposition 59 (Plancherel's formula for Schwartz functions)**.** *Let* $f, g \in \mathcal{S}(\mathbf{R}^d)$*. Then*

$$\int f(x)\overline{g}(x)dx = \int \widehat{f}(\xi)\widehat{\overline{g}}(\xi)d\xi.$$

Proof. Using (5) and (6) in Proposition 53, we can write $\widehat{\check{\overline{g}}} = \widehat{\overline{g}}$. Using (7) from the same proposition, we can write $\widehat{f * \check{\overline{g}}} = \widehat{f} \cdot \widehat{\overline{g}}$. Since the Fourier transform of Schwartz functions are Schwartz functions, and their products are also Schwartz functions, we see that $\widehat{f} \cdot \widehat{\overline{g}}$ is a Schwartz function, and, therefore, it is in L^1. We use Fourier inversion for $\widehat{f * \check{\overline{g}}}$, to write

$$\int f(x)\overline{g}(x)dx = f * \check{\overline{g}}(0) = \int \widehat{f} \cdot \widehat{\overline{g}}(\xi)e^{2\pi i\langle 0,\xi\rangle}d\xi.$$

This proves the claim. □

The proposition shows that the map $f \mapsto \widehat{f}$ is a linear transformation on $\mathcal{S}(\mathbf{R}^d)$ that preserves the inner product defining the Hilbert space $L^2(\mathbf{R}^d)$. Since $\mathcal{S}(\mathbf{R}^d)$ is dense in $L^2(\mathbf{R}^d)$, this allows us to extend the Fourier transform to $L^2(\mathbf{R}^d)$.

This is done as follows. Proposition 59 applied with $f = g$ shows that $\|f\|_2 = \|\widehat{f}\|_2$ for Schwartz functions. In particular the Fourier transform of the elements of a Cauchy sequence is a Cauchy sequence. If $f \in L^2(\mathbf{R}^d)$, and $(f_n) \subset \mathcal{S}(\mathbf{R}^d)$ is a sequence with $\lim f_n = f$, then $(\widehat{f_n})$ is a Cauchy sequence and has a limit. We define $\widehat{f} := \lim \widehat{f}_n$.

It is easy to see that the definition does not depend on the choice of f_n. Indeed, if $g_n \to f$ is another sequence, then $\|g_n - f_n\|_2 \to 0$ implies $\|\widehat{g_n - f_n}\|_2 = \|\widehat{g}_n - \widehat{f}_n\|_2 \to 0$, so $\lim \widehat{g}_n = \lim \widehat{f}_n = \widehat{f}$.

Moreover, when $f \in L^1(\mathbf{R}^d) \cap L^2(\mathbf{R}^d)$, the new definition of $\widehat{f}$ is compatible with the old one. Indeed, in this case, we may pick a sequence of Schwartz functions that converge to f in both $L^1(\mathbf{R}^d)$ and $L^2(\mathbf{R}^d)$. The proof in Proposition 36 gives a sequence independent of p. Now $\widehat{f}_n(\xi)$ converges for every ξ to the L^1 definition of $\widehat{f}(\xi)$ by $\|f - f_n\|_1 \to 0$ and it converges in L^2 to the L^2 definition of $\widehat{f}$. We may take a subsequence to turn L^2 convergence into pointwise convergence almost everywhere, and this shows that the two definitions of $\widehat{f}$ coincide.

With this definition of the Fourier transform on $L^2(\mathbf{R}^d)$, it inherits items (2)–(6) from Proposition 53, and it also has the following properties.

Proposition 60. *Let* $f, g \in L^2(\mathbf{R}^d)$*.*

1. $\widehat{f} \in L^2(\mathbf{R}^d)$ *and* $\|\widehat{f}\|_2 = \|f\|_2$,
2. $\langle f, g\rangle = \langle \widehat{f}, \widehat{g}\rangle$,

(3) $\hat{\hat{f}} = \check{f}$,
(4) $\widehat{f \cdot g} = \hat{f} * \hat{g}$.

Proof. (1) and (2) follows by approximating f and g with Schwartz functions and Plancherel's formula for such functions.

(3) For Schwartz functions, this follows by Fourier inversion, and the formula extends to general L^2 functions by taking the limit.

(4) For Schwartz functions, this follows by item (7) in Proposition 53 applied for $\hat{f}$ and $\hat{g}$ in the role of f and g and Fourier inversion:

$$\widehat{\hat{f} * \hat{g}} = \hat{\hat{f}} \cdot \hat{\hat{g}} = \check{f} \cdot \check{g},$$

which after applying Fourier inversion to both sides gives

$$\hat{f} * \hat{g} = \widehat{f \cdot g}.$$

Now we take sequences of Schwartz functions $f_n \to f$ and $g_n \to g$ in $L^2(\mathbf{R}^d)$. We observe that $\hat{f}_n * \hat{g}_n \to \hat{f} * \hat{g}$ in L^∞ by Young's inequality and $f_n \cdot g_n \to f \cdot g$ in L^1 by the Cauchy-Schwarz inequality:

$$\begin{aligned}\|f_n g_n - fg\|_1 &\le \|(f_n - f)g\|_1 + \|f_n(g_n - g)\|_1 \\ &\le \|f_n - f\|_2 \|g\|_2 + \|f_n\|_2 \|g_n - g\|_2 \to 0.\end{aligned}$$

□

11.1. **Fourier series of periodic functions, Poisson summation formula.** We say that a function $\varphi : \mathbf{R}^d \to \mathbf{C}$ is periodic if $\varphi(x + a) = \varphi(x)$ for all $a \in \mathbf{Z}^d$ and $x \in \mathbf{R}^d$. We write $L^p(\mathbf{T}^d)$ for spaces of measurable periodic functions endowed with the norm

$$\|\varphi\|_{L^p(\mathbf{T}^d)} = \Big(\int_{[0,1]^d} |\varphi(x)|^p dx \Big)^{1/p}$$

for $p \in [1, \infty)$, while for $p = \infty$ the norm is the essential supremum.

To explain the notation, $\mathbf{T}^d$ stands for $\mathbf{R}^d/\mathbf{Z}^d$, which can be endowed with the structure of a manifold and a group, and there is a unique probability measure on it that is invariant under translations with respect to the group operation. Periodic functions can be identified with functions on $\mathbf{R}^d/\mathbf{Z}^d$ and we can talk about L^p spaces with respect to the above measure. This turns out to be equal to the above norm. We do not need this viewpoint in this course.

For $\varphi \in L^1(\mathbf{T}^d)$, we define its Fourier coefficients by

$$\hat{f}(\xi) = \int_{[0,1]^d} f(x) e^{-2\pi i \langle x, \xi \rangle} dx.$$

Unlike in the case of functions defined on $\mathbf{R}^d$, the functions $e^{2\pi i \langle x, \xi \rangle}$ form an orthonormal basis in $L^2(\mathbf{T}^d)$ and the Fourier coefficients $\hat{\varphi}(\xi)$ are the coefficients of the basis elements when we decompose φ. Note that $L^2(\mathbf{T}^d) \subset L^1(\mathbf{T}^d)$ so we do not need to introduce a separate definition for the Fourier coefficients for L^2 functions. We do not prove the above facts in this course.

We first explain the relationship between the Fourier coefficients and the Fourier transform discussed above. To this end, it is helpful to introduce a map $P : L^1(\mathbf{R}^d) \to L^1(\mathbf{T}^d)$ that is defined by

$$Pf(x) = \sum_{a \in \mathbf{Z}^d} f(x-a) = \sum_{a \in \mathbf{Z}^d} \tau_a f(x).$$

The function Pf is clearly periodic, and the identity

$$\begin{aligned}\int_{\mathbf{T}^d} Pf dx &:= \int_{[0,1]^d} Pf dx \\ &= \int_{[0,1]^d} \sum_{a \in \mathbf{Z}^d} \tau_a f dx \\ &= \sum_{a \in \mathbf{Z}^d} \int_{[0,1]^d} \tau_a f dx \\ &= \sum_{a \in \mathbf{Z}^d} \int_{[0,1]^d + a} f dx \\ &= \int_{\mathbf{R}^d} f dx\end{aligned}$$

together with $|Pf(x)| \leq P|f|(x)$ implies $\|Pf\|_{L^1(\mathbf{T}^d)} \leq \|f\|_{L^1(\mathbf{R}^d)}$. To exchange the integral and the sum, we used Fubini's theorem. To justify its application, we show the finiteness of $\int_{[0,1]^d} \sum_{a \in \mathbf{Z}^d} |\tau_a f| dx$ using the same calculation.

This map has a partial inverse. Let $u \in L^1(\mathbf{R}^d)$ be any function with $Pu \equiv 1$. We may take $u = 1_{[0,1)^d}$ for now, but later we will want a smooth function with this property. It is immediate from the definitions that $\varphi = Pu\varphi$ for all $\varphi \in L^1(\mathbf{T}^d)$. On the other hand, uPf may differ from f in general. This shows that $\varphi \mapsto u\varphi$ is a partial inverse of P. In particular, P is surjective but not injective.

Using these maps, we have the following relationship between the Fourier transform on $L^1(\mathbf{R}^d)$ and the Fourier coefficients on $L^1(\mathbf{R}^d/\mathbf{Z}^d)$:

$$\begin{aligned}\widehat{Pf}(\xi) = \hat{f}(\xi) &\qquad \text{for } f \in L^1(\mathbf{R}^d) \text{ and } \xi \in \mathbf{Z}^d, \\ \hat{\varphi}(\xi) = \widehat{u\varphi}(\xi) &\qquad \text{for } \varphi \in L^1(\mathbf{T}^d) \text{ and } \xi \in \mathbf{Z}^d.\end{aligned}$$

In the above formula $\hat{\cdot}$ stands for either the Fourier transform or for a Fourier coefficient. The meaning can be inferred from the type of the function to which it is applied.

Using this relationship many of the properties of Fourier transform is inherited by the Fourier coefficients. We prove an appropriate version of Fourier inversion.

Theorem 61. *Let $\varphi \in L^1(\mathbf{T}^d)$ and suppose that $\sum_{\xi\in\mathbf{Z}^d} |\widehat{\varphi}(\xi)| < \infty$. Then*

$$\varphi(x) = \sum_{\xi\in\mathbf{Z}^d} \widehat{\varphi}(\xi)e^{2\pi i\langle x,\xi\rangle}$$

for almost every x.

Under the assumptions of the theorem, the sum converges uniformly for all x to a continuous function. This means that f equals almost everywhere to a continuous function. If we assume that $f \in C(\mathbf{T}^d)$, then the convergence in the theorem holds for all x.

Proof. The argument is similar to Fourier inversion on $\mathbf{R}^d$. Let $a_n(\xi) \in [0, 1]$ for $n \in \mathbf{Z}_{>0}$ and $\xi \in \mathbf{Z}^d$ be such that $\lim_{n\to\infty} a_n(\xi) = 1$. Then

$$\sum_{\xi\in\mathbf{Z}^d} \widehat{\varphi}(\xi)e^{2\pi i\langle x,\xi\rangle} = \lim_{n\to\infty} \sum_{\xi\in\mathbf{Z}^d} \widehat{\varphi}(\xi)e^{2\pi i\langle x,\xi\rangle} a_n(\xi).$$

by dominated convergence. Let $u = 1_{[-1/2,1/2)^d}$. We write

$$\sum_{\xi\in\mathbf{Z}^d} \widehat{\varphi}(\xi)e^{2\pi i\langle x,\xi\rangle} a_n(\xi) = \sum_{\xi\in\mathbf{Z}^d} \int_{\mathbf{R}^d} u\varphi(t)e^{2\pi i\langle x-t,\xi\rangle} a_n(\xi)dt = (u\varphi) * g_n(x),$$

where

$$g_n(x) = \sum_{\xi\in\mathbf{Z}^d} e^{2\pi i\langle x,\xi\rangle} a_n(\xi).$$

By a question on the example sheet, a_n may be chosen in such a way that ug_n forms an approximate identity as $n \to \infty$. By Theorem 35, and after passing to a subsequence, $(u\varphi) * (ug_n) \to u\varphi$ almost everywhere. Note that $g_n = \sum_{a\in\mathbf{Z}^d} \tau_a(ug_n)$, and

$$(u\varphi) * (\tau_a(ug_n)) \to \tau_a(u\varphi)$$

almost everywhere. For all $x \in \mathbf{R}^d$, $(u\varphi) * (\tau_a(ug_n)) = 0$ for all but finitely many a, so we have

$$(u\varphi) * g_n \to \varphi$$

almost everywhere, and this completes the proof. □

Theorem 62 (Poisson summation formula)**.** *Let $f \in \mathcal{S}(\mathbf{R}^d)$. Then*

$$\sum_{x\in\mathbf{Z}^d} f(x) = \sum_{\xi\in\mathbf{Z}^d} \widehat{f}(\xi).$$

Proof. Since $f(x) \le C(1 + |x|)^{-d-1}$ for a suitable C, the sum defining $Pf(x)$ converges uniformly for $|x| \le \sqrt{d}$, so Pf is continuous and $Pf(0) = \sum_{x\in\mathbf{Z}^d} f(x)$. In particular, Fourier inversion is valid for all x, and we conclude

$$Pf(0) = \sum_{\xi\in\mathbf{Z}^d} \widehat{Pf}(\xi)e^{2\pi i\langle 0,\xi\rangle} = \sum_{\xi\in\mathbf{Z}^d} \widehat{f}(\xi).$$

□

12. Distributions

In this section, we are going to define a space of objects called distributions or generalised functions that contain all function spaces we have encountered so far. We will extend important operations, such as multiplication by smooth functions, differentiation and convolution to this space. In particular, all distributions will be infinitely differentiable in this space. This has similar benefits to extending the field of real numbers to the field of complex numbers, where all algebraic equations have solutions.

The idea of this construction, is that functions can be identified with functionals on an appropriate space. For example L^p for $p > 1$ is the dual of L^q for the appropriate q, and even L^1 embeds into the dual of L^∞. Now all these functionals can be restricted to any subspace of L^q. We will call this subspce the space of test functions. If we choose a stronger topology on the test functions (that is, we add more open sets) then the functionals corresponding to functions remain continuous, and there will be potentially many more continuous functionals.

On L^p, differentiation is a linear map that is defined on a proper subspace. Its adjoint on L^q, which is also differentiation as we will see, is also defined on a proper subspace. If we choose a space that is closed under differentiation for our test functions, then we will be able to define differentiation on the dual as the adjoint of differentiation on test functions and obtain a linear map that is defined on the whole space. We are not going to make this informal discussion rigorous, but it will motivate are definitions.

In this section $\Omega \subset \mathbf{R}^d$ is a fixed open domain.

Theorem 63. *There is a topology on the space $C_c^\infty(\Omega)$ of infinitely differentiable functions supported on a compact subset of Ω such that the following holds.*

(1) *A sequence $(f_n) \subset C_c^\infty(\Omega)$ converges to some function $f \in C_c^\infty(\Omega)$ if and only if, there is a compact set $K \subset \Omega$ such that $\operatorname{supp} f_n \subset K$ for all n, and $D^\alpha f_n \to D^\alpha f$ uniformly for all $\alpha \in \mathbf{Z}_{\geq 0}^d$.*
(2) *A linear map $A : C_c^\infty(\Omega) \to V$ to a topological vector space is continuous if and only if $\lim_{n\to\infty} A(f_n) = A(f)$ whenever $\lim_{n\to\infty} f_n = f$ for some functions $f, (f_n) \subset C_c^\infty(\Omega)$.*

This theorem is not proved in this course. There are more than one topologies on C_c^∞ that satisfy these properties. A particularly nice one is described in Appendix A of the notes of Claude Warnick [5]. This topology could be defined using a family of seminorms, but this is not the usual approach in the literature.

We denote by $\mathcal{D} = \mathcal{D}(\Omega)$ the space $\mathbf{C}_c^\infty(\Omega)$ endowed with a topology described in the theorem. We stress that this topology is not metrizable, and the second item of the theorem does not hold for general non-linear maps. We write $\mathcal{D}'$ for the space of continuous linear functionals on $\mathcal{D}$.

There is another useful characterization of continuity of functionals, which we state without proof.

Lemma 64. *Let $u : \mathcal{D}(\Omega) \to \mathbf{C}$ be a linear functional. Then u is continuous if and only if for all compact sets $K \subset \Omega$, there are $C = C(K)$ and $k = k(K)$ such that*

$$|u(f)| \le C\|f\|_{C^k(K)}$$

for all f with $\operatorname{supp} f \subset K$.

If the property in the lemma holds for some k that is independent of K and some C that may depend on K, then we say that f is of finite order and the smallest such k is the order of u.

Lemma 65. *For each $f \in L^1_{loc}(\Omega)$, the formula*

$$T_f(g) = \int_\Omega fg dx$$

defines an element of $\mathcal{D}'$.

Moreover, $f \mapsto T_f : L^1_{loc}(\Omega) \to \mathcal{D}'$ is an injective linear transformation.

This lemma shows that we can regard functions in L^1_{loc} as distributions. In what follows we often will not distinguish between f and T_f. In particular, we are going to write $u \in X$, where X is a function space when we mean that there is some $f \in X$ such that $u = T_f$.

Proof. If $g, (g_n) \subset \mathcal{D}$ such that $\lim g_n = g$, then there is a compact $K \subset \Omega$ such that $\operatorname{supp} g_n \subset K$ for all n, and $g_n \to g$ uniformly. Therefore,

$$\lim_n \int_\Omega f g_n dx = \lim_n \int_K f g_n dx = \int_K f g dx = \int_\Omega f g dx$$

by dominated convergence because the restriction of f to K is integrable.

The linearity of $f \to T_f$ follows by the linearity of integration. Injectivity is a consequence of the next lemma. □

Lemma 66. *Let $\varphi \in \mathbf{C}^\infty_c(\mathbf{R}^d)$ such that $\varphi(x) \ge 0$ for all x and $\int \varphi dx = 1$, and let $\varphi_n = n^d \varphi(nx)$ for $n \in \mathbf{Z}_{>0}$. Let $f \in L^1_{loc}(\Omega)$ and let $K \subset \Omega$ compact. Then*

$$1_K \cdot T_f(\tau_x \varphi_n) = 1_K \cdot f * \breve{\varphi}_n(x) \to 1_K f(x)$$

in $L^1(\mathbf{R}^d)$.

Since K is compact, there is some $\delta > 0$ such that $\operatorname{dist}(x, \mathbf{R}^d \backslash \Omega) \ge \delta$ for all $x \in K$. If n is large enough that $\operatorname{supp} \varphi_n \subset B(0, \delta)$, so the function in the lemma is well defined for $x \in K$. For $x \notin K$, we just understand the function as 0.

Proof. We can compute

$$T_f(\tau_x\varphi_n) = \int f(t)\varphi_n(t-x)dt = f * \check{\varphi}_n(x)$$

whenever $\operatorname{supp}\tau_x\varphi_n \subset \Omega$.

Now let $L = \{x : \operatorname{dist}(K,x) \leq \delta/2\}$. Then $(1_L \cdot f) * \check{\varphi}_n$ agrees with $f * \check{\varphi}_n$ on K for n large enough, and it converges to $1_L \cdot f$ in $L^1(\mathbf{R}^d)$ by Theorem 35. ☐

Example 67. If μ is a Borel measure on Ω with the property that $\mu(K) < \infty$ for all compact $K \subset \Omega$, then

$$T_\mu(f) = \int f d\mu$$

defines a distribution. Of particular interest to us will be the Dirac delta $\delta_x(f) := f(x)$ for any $x \in \Omega$.

Consider

$$P.V.(1/x)(f) := \lim_{r\to 0}\int_{\mathbf{R}\setminus[-r,r]} \frac{f(x)}{x}dx.$$

Note that

$$\begin{aligned}\int_{[-1,-r]\cup[r,1]} \frac{f(x)}{x}dx &= \int_r^1 \frac{f(x)-f(-x)}{x}dx = \int_r^1\int_{-x}^x x^{-1}f'(t)dtdx \\ &= \int_{-1}^1\int_{\max(|t|,r)}^1 x^{-1}f'(t)dxdt = \int_{-1}^1 -\log(\max(|t|,r))f'(t)dt \\ &\to \int_{-1}^1 -\log|t|f'(t)dt.\end{aligned}$$

we see that

$$P.V.(1/x)(f) = \int_{\mathbf{R}\setminus[-1,1]} \frac{f(x)}{x}dx + \int_{-1}^1 -\log|x|f'(x)dx$$

is a well defined distribution of order at most 1.

Definition 68. Let $u \in \mathcal{D}'(\Omega)$, and let $f \in C^\infty(\Omega)$. We define $fu \in \mathcal{D}'(\Omega)$ by

$$fu(\varphi) = u(f\varphi)$$

for $\varphi \in \mathcal{D}$.

Let $\alpha \in \mathbf{Z}^d_{\geq 0}$. We define $D^\alpha u \in \mathcal{D}'(\Omega)$ by

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

for $\varphi \in \mathcal{D}$.

It is easy to check that the above formulas give rise to well-defined distribution.

Lemma 69. *Let $f \in C^\infty(\Omega)$ and $g \in L^1_{loc}(\Omega)$. Then*

$$fT_g = T_{fg}.$$

Proof. Unwinding the definitions, we can write
$$(fT_g)(h) = T_g(fh) = \int fhgdx = T_{fg}(h)$$
for all $h \in \mathcal{D}(\Omega)$. $\square$

Lemma 70. *Let $k \in \mathbf{Z}_{\geq 0}$. Let $f \in C^k(\Omega)$ and let $\alpha \in \mathbf{Z}^d_{\geq 0}$ with $|\alpha| \leq k$. Then*
$$D^\alpha T_f = T_{D^\alpha f}.$$

Proof. It is enough to prove this for $|\alpha| = 1$ Unwinding the definitions, we can write
$$(D^\alpha T_g)(h) = T_g(D^\alpha h) = \int gD^\alpha hdx = -\int (D^\alpha g)hdx = T_{D^\alpha g}(h)$$
for all $h \in \mathcal{D}(\Omega)$, where we used integration by parts. $\square$

Example 71. Consider the Heaviside function
$$H(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then
$$\frac{d}{dx} T_H(f) = -\int_0^\infty f'(x)dx = f(0).$$
We see that $(d/dx)T_H = \delta_0$.

Definition 72. Let $u \in \mathcal{D}'(\Omega)$. The support of u is defined as
$$\operatorname{supp} u = \bigcap \{K : K \subset \Omega \text{ is closed and } u(f) = 0 \text{ for all } f \in \mathbf{C}_c^\infty(\Omega \backslash K)\}.$$

Lemma 73. *Let $u \in \mathcal{D}'(\Omega)$ and $f \in C_c^\infty(\Omega \backslash \operatorname{supp} u)$, then $u(f) = 0$.*

This proof is not lectured. It is given in the notes for the sake of completeness.

Proof. Let $\mathcal{U}$ be the collection of open sets U that has the property that $u(h) = 0$ for all $h \in C_c^\infty(U)$. By definition $\operatorname{supp} u = \Omega \backslash \bigcup \mathcal{U}$. Let $h_0 \in C_c(\Omega \backslash \operatorname{supp} u)$. Then $\operatorname{supp} h_0$ is covered by the open sets in $\mathcal{U}$, and by compactness, there is a finite cover $\operatorname{supp} h_0 \subset U_1 \cup \ldots \cup U_n$ for some $U_1, \ldots, U_n \in \mathcal{U}$.

We will construct a so-called partition of unity, that is functions $\varphi_1, \ldots, \varphi_n \in \mathcal{D}(\Omega)$ such that $\operatorname{supp} \varphi_j \subset U_j$ for every j and $\sum \varphi_j(x) = 1$ for $x \in K := \operatorname{supp} h_0$. Then $h_0 = \varphi_1 h_0 + \ldots + \varphi_n h_0$ and $\operatorname{supp} \varphi_j h_0 \subset U_j$, hence $u(\varphi_j h_0) = 0$ for all j. We can conclude $u(h_0) = 0$.

It remains to construct $\varphi_1, \ldots, \varphi_n$. Fix $\varepsilon > 0$ such that $\max_j \operatorname{dist}(x, \mathbf{R}^d \backslash U_j) > \varepsilon$ and $|x| < \varepsilon^{-1}$ for all $x \in K$. This is possible, because the function in question is continuous and does not vanish on K. For each $j = 1, \ldots, n$, let K_j be the set of points x such that $|x| \leq \varepsilon^{-1}$ and $\operatorname{dist}(x, \mathbf{R}^d \backslash U_j) \geq \varepsilon$. Then $K_j \subset U_j$ is a compact set for all j and $\bigcup K_j \supset K$. We take a function $\psi \in C_c^\infty(B(0,\varepsilon))$ such that $\psi(x) \geq 1$ for $|x| \leq |\varepsilon|/2$. Then it is easy to check that

$$\varphi_j(x) = \frac{1_{K_j} * \psi}{1_{K_1} * \psi + \ldots + 1_{K_n} * \psi}$$

is a suitable partition of unity. □

Lemma 74. *For $f \in L^1_{loc}(\Omega)$, $\operatorname{supp} T_f =$ ess supp(f).*

Proof. Let $K \subset \Omega$ be a closed set. If $|\{x \in \Omega \backslash K : f(x) \neq 0\}| = 0$, then $T_f(h) = \int f h dx = 0$ for all $h \in \mathcal{D}(\Omega)$ with $\operatorname{supp} h \cap K = \varnothing$. This proves $\operatorname{supp} T_f \subset$ ess supp(f).

Conversely, let $\widetilde{\Omega} = \Omega \backslash \operatorname{supp} T_f$, and consider $T_{f|_{\widetilde{\Omega}}} \in \mathcal{D}'(\widetilde{\Omega})$. Note that $T_{f|_{\widetilde{\Omega}}}(h) = T_f(h) = 0$ for $h \in \mathcal{D}(\widetilde{\Omega})$. Now the injectivity of $f \mapsto T_f$ applied for $\mathcal{D}'(\widetilde{\Omega})$ implies the claim. □

Lemma 75. *Let $v \in \mathcal{D}'(\mathbf{R}^d)$ be of compact support and let $h \in \mathcal{D}(\mathbf{R}^d)$. Then $t \mapsto v(\tau_{-t}h) \in \mathcal{D}(\mathbf{R}^d)$. Moreover,*

$$D_t^\alpha v(\tau_{-t}h) = v(\tau_{-t}D^\alpha h).$$

Proof. If $(t_n) \subset \mathbf{R}^d$ is a sequence with $t_n \to t$ for some t, then $\tau_{-t_n}h \to h$ in the topology of $\mathcal{D}(\mathbf{R}^d)$. (See the example sheet.) Therefore, $t \mapsto v(\tau_{-t}h)$ is continuous.

Now it is enough to prove that the partial derivatives exist and the formula given for them is correct for $|\alpha| = 1$. Write

$$\frac{v(\tau_{-(t_0+s\alpha)}h) - v(\tau_{-t_0}h)}{s} = v\Big(\frac{\tau_{-(t_0+s\alpha)}h - \tau_{-t_0}h}{s}\Big)$$

for $s \in [-1, 1]$. Then

$$\frac{\tau_{-(t_0+s_n\alpha)}h - \tau_{-t_0}h}{s_n} \to \tau_{-t_0}D^\alpha h$$

for any sequence $s_n \to 0$ in the $\mathcal{D}(\mathbf{R}^d)$ topology. (See the example sheet.) The continuity of v implies $D_t^\alpha v(\tau_{-t}h) = v(\tau_{-t}D^\alpha h)$. □

Definition 76. Let $u, v \in \mathcal{D}(\mathbf{R}^d)$, and suppose that $\operatorname{supp} v$ is compact. Then

$$u * v(h) = u(t \mapsto v(x \mapsto \tau_{-t}h(x))).$$

Lemma 77. *Let $f \in L^1_{loc}(\mathbf{R}^d)$ and $g \in L^1(\mathbf{R}^d)$ such that $\operatorname{supp} g$ is compact. Then*

$$T_{f*g} = T_f * T_g.$$

Proof. For $h \in \mathcal{D}(\mathbf{R}^d)$, we write

$$T_f * T_g(h) = \int f(t) T_g(\tau_{-t}(h)) dt = \int\int f(t) g(x) h(x+t) dx dt$$
$$= \int\int f(t) g(x-t) dt h(x) dx = T_{f*g}(h).$$

To justify the application of Fubini, let R be large enough so that $\operatorname{supp} g, \operatorname{supp} h \subset B(0,R)$, and write

$$\begin{aligned}\int\int |f(t)g(x)h(x+t)|dxdt &= \int_{|t|\le 2R}\int_{|x|\le R} |f(t)g(x)h(x+t)|dxdt \\ &\le \|g\|_1\|h\|_\infty \int_{|t|\le 2R} |f(t)|dt < \infty.\end{aligned}$$

□

Lemma 78. *Let $u, v \in \mathcal{D}(\mathbf{R}^d)$, and suppose that $\operatorname{supp} v$ is compact. Let $\alpha \in \mathbf{Z}^d_{\ge 0}$. Then*

$$D^\alpha(u*v) = (D^\alpha u)*v = u*(D^\alpha v).$$

Proof. For $h \in \mathcal{D}(\mathbf{R}^d)$, we write

$$\begin{aligned}(D^\alpha(u*v))(h) &= (-1)^{|\alpha|}(u*v)(D^\alpha h) = (-1)^{|\alpha|}u(t\mapsto v(\tau_{-t}D^\alpha h)) \\ &= u(t\mapsto (-1)^{|\alpha|}v(D^\alpha\tau_{-t}h)) \\ &= u(t\mapsto D^\alpha v(\tau_{-t}h)) = u*(D^\alpha v)(h).\end{aligned}$$

In addition using $v(\tau_{-t}D^\alpha h) = D^\alpha_t v(\tau_{-t}h)$ from Lemma 75, we can write

$$\begin{aligned}(-1)^{|\alpha|}u(t\mapsto v(\tau_{-t}D^\alpha h)) &= (-1)^{|\alpha|}u(D^\alpha_t(t\mapsto v(\tau_{-t}h))) \\ &= (D^\alpha u)(t\mapsto v(\tau_{-t}h)) = (D^\alpha u)*v(h).\end{aligned}$$

□

Definition 79. Let $k \in \mathbf{Z}_{\ge 1}$, and let $a_\alpha \in \mathbf{C}$ for $\alpha \in \mathbf{Z}^d_{\ge 0}$ with $|\alpha| \le k$. A distribution $G \in \mathcal{D}'(\mathbf{R}^d)$ is called a fundamental solution of the partial differential operator

$$L = \sum_{|\alpha|\le k} \alpha_\alpha D^\alpha$$

if it satisfies the equation $LG = \delta_0$.

Theorem 80. *Suppose G is a fundamental solution of L as in the above definition, and let $u_0 \in \mathcal{D}'(\mathbf{R}^d)$ be of compact support. Then $u = G*u_0$ solves the equation $Lu = u_0$.*

Proof. We can write

$$L(G*u_0) = \sum_{|\alpha|\le k}\alpha_\alpha D^\alpha(G*u_0) = \sum_{|\alpha|\le k}\alpha_\alpha(D^\alpha G)*u_0 = \Big(\sum_{|\alpha|\le k}\alpha_\alpha D^\alpha G\Big)*u_0 = \delta_0*u_0.$$

Let $h \in \mathcal{D}(\mathbf{R}^d)$. Then

$$\delta_0 * u_0(h) = \delta_0(t\mapsto u_0(\tau_{-t}h)) = u_0(h),$$

hence $\delta_0 * u_0 = u_0$. □

Example 81 (Poisson equation)**.** Let $G(x) = \frac{-1}{4\pi|x|}$ for $x \in \mathbf{R}^3$. A calculation using the divergence theorem (see example sheet) gives $\Delta T_G = \delta_0$, therefore $u = T_G * T_f$ is the *distributional* solution of Poisson's equation

$$\Delta u = T_f$$

for any compactly supported $f \in L^1(\mathbf{R}^d)$. If $w \in G * f \in C^2(\mathbf{R}^d)$, then it is a *classical solution* of

$$\Delta w = f.$$

It follows from the results discussed above that this holds when $f \in C^2_c(\mathbf{R}^3)$. It turns out that much less regularity is enough, and w is a classical solution if f is Holder continuous for any positive exponent. We do not prove this fact in this course.

13. Tempered distribution

Definition 82. We endow $\mathcal{S}(\mathbf{R}^d)$, the space of Schwartz functions with the family of norms

$$p_N(f) = \max_{\alpha:|\alpha|\le N} \|(1+|x|)^N D^\alpha f(x)\|_\infty$$

These induce a topology on $\mathcal{S}(\mathbf{R}^d)$, which is also a metric space with

$$\mathrm{dist}(f,g) = \sum_{N=1}^{\infty} \min(p_N(f-g), 2^{-N}).$$

The space $\mathcal{S}'(\mathbf{R}^d)$ of continuous linear functionals on $\mathcal{S}(\mathbf{R}^d)$ is called the space of tempered distributions. It is endowed with the weak-$*$ topology.

A sequence of Schwartz functions f_n converges to another Schwartz function f if and only if $p_N(f - f_n) \to 0$ for all N. A sequence of tempered distributions u_n converges to a tempered distribution u if and only if $u_n(f) \to u(f)$ for all $f \in \mathcal{S}(\mathbf{R}^d)$.

Tempered distributions can be restricted to $\mathcal{D}(\mathbf{R}^d)$, and it is easy to see that this yields a continuous functional, that is a distribution in $\mathcal{D}'(\mathbf{R}^d)$. This allows us to extend all of the operations we defined for $\mathcal{D}'$.

Let $t \in \mathbf{R}^d$. We define

$$\tau_t u(h) = u(\tau_{-t} h).$$

Lemma 83. *Let $f : \mathbf{R}^d \to \mathbf{C}$ be a measurable function such that $|f(x)| \le C(1+|x|)^N$ almost everywhere for some $C, N > 0$. Then $T_f : \mathcal{S} \to \mathbf{C}$ defined as*

$$T_f(g) = \int f(x) g(x) dx$$

for $g \in \mathcal{S}(\mathbf{R}^d)$ is a tempered distribution.

For all $t \in \mathbf{R}^d$, we have

$$\tau_t T_f = T_{\tau_t f}.$$

Proof. Let $g_n \to g$ be Schwartz functions. Then $p_{N+d+1}(g_n - g) \to 0$, so

$$\int |f(x)(g_n(x) - g(x))| dx \le \int C(1+|x|)^{N+d+1}(g_n(x) - g(x))(1+|x|)^{-d-1} dx$$
$$\le C p_N(g_n - g)\|(1+|x|)^{-d-1}\|_1 \to 0.$$

We can also write

$$\tau_t T_f(h) = \int f(x)h(x+t)dx = \int f(x-t)h(x)dx = T_{\tau_{-t} f}(h).$$

□

Definition 84. Let $u \in \mathcal{S}'(\mathbf{R}^d)$ be a tempered distribution. The Fourier transform of u is defined as the tempered distribution

$$\widehat{u}(f) = u(\widehat{f})$$

for $f \in \mathcal{S}$.

Proposition 85. *The map $f \mapsto \widehat{f}$ is continuous on $\mathcal{S}(\mathbf{R}^d)$.*
The map $u \mapsto \widehat{u}$ is continuous on $\mathcal{S}'(\mathbf{R}^d)$.

Proof. We have

$$|\xi|^{2n} D^\alpha \widehat{f} = (-2\pi i)^{|\alpha|}(-4\pi^2)^{-n}(\Delta^n(x^\alpha f))^\wedge.$$

If $f_m \to f$ is a convergent sequence in $\mathcal{S}(\mathbf{R}^d)$, then

$$(1+|x|)^{d+1}\Delta^n(x^\alpha f_m) \to (1+|x|)^{d+1}\Delta^n(x^\alpha f)$$

uniformly. Therefore,

$$\Delta^n(x^\alpha f_m) \to \Delta^n(x^\alpha f)$$

in $L^1(\mathbf{R}^d)$. By item (1) in Proposition 53,

$$|\xi|^{2n} D^\alpha \widehat{f}_m \to |\xi|^{2n} D^\alpha \widehat{f}$$

uniformly, hence $\widehat{f}_m \to \widehat{f}$ in $\mathcal{S}(\mathbf{R}^d)$.

Now let $u, u_1, u_2, \ldots \in \mathcal{S}'(\mathbf{R}^d)$ such that $u_m \xrightarrow{*} u$. Let $f \in \mathcal{S}(\mathbf{R}^d)$. Then

$$\widehat{u}_m(f) = u_m(\widehat{f}) \to u(\widehat{f}) = \widehat{u}(f).$$

□

Lemma 86. *For $f \in L^1(\mathbf{R}^d) \cup L^2(\mathbf{R}^d)$, we have*

$$\widehat{T_f} = T_{\widehat{f}}.$$

Proof. First suppose $f \in L^1(\mathbf{R}^d)$ and let $h \in \mathcal{S}(\mathbf{R}^d)$, then

$$\widehat{T_f}(h) = \int f\widehat{h}dx = \int\int f(x)h(y)e^{-2\pi i\langle y,x\rangle}dydx = \int \widehat{f}(y)h(y)dy.$$

This requires an application of Fubini, which is justified by $f, h \in L^1(\mathbf{R}^d)$.

If $f \in L^2(\mathbf{R}^d)$, then the identity $\int f\widehat{h}dx = \int \widehat{f}hdx$ can be proved using a sequence of functions $f_n \in \mathcal{S}(\mathbf{R}^d)$ that converge to f in $L^2(\mathbf{R}^d)$. Alternatively, the identity can be proved using Parseval's formula. □

Lemma 87. *Let $u \in \mathcal{S}'(\mathbf{R}^d)$, $t \in \mathbf{R}^d$ and $\alpha \in \mathbf{Z}^d_{\geq 0}$. Then*

$$(e^{2\pi i\langle x,t\rangle}u)^\wedge = \tau_t\widehat{u}, \quad \widehat{\tau_t u} = e^{-2\pi i\langle \xi,t\rangle}\widehat{u}$$
$$\widehat{D^\alpha u} = (2\pi i)^{|\alpha|}\xi^\alpha\widehat{u}, \quad D^\alpha\widehat{u} = (-2\pi i)^{|\alpha|}\widehat{x^\alpha u},$$
$$\widehat{\widehat{u}} = \check{u}.$$

In the last line, we use the notation $\check{u}(h) = u(\check{h})$. It is easy to check that $\check{T_f} = T_{\check{f}}$ for appropriate f for which T_f is defined.

Proof. Let $h \in \mathcal{S}(\mathbf{R}^d)$. Then

$$(e^{2\pi i\langle x,t\rangle}u)^\wedge(h) = u(e^{2\pi i\langle x,t\rangle}\widehat{h}) = u(\widehat{\tau_{-t}h}) = \tau_t\widehat{u}(h),$$

where we used item (4) in Proposition 53.

We can write

$$\widehat{\tau_t u}(h) = u(\tau_{-t}\widehat{h}) = u((e^{-2\pi i\langle x,t\rangle}h)^\wedge) = e^{-2\pi i\langle x,t\rangle}\widehat{u}(h),$$

where we used

$$(e^{2\pi i\langle x,t\rangle}h)\widehat{}(\xi) = \int h(x)e^{-2\pi i\langle x,\xi-t\rangle}dx = \tau_t\widehat{h}(\xi).$$

We can write

$$\widehat{D^\alpha u}(h) = (-1)^{|\alpha|}u(D^\alpha\widehat{h}) = (-1)^{|\alpha|}u((-2\pi i)^{|\alpha|}\widehat{x^\alpha h}) = (2\pi i)^{|\alpha|}x^\alpha\widehat{u},$$

where we used Proposition 55.

We can write

$$D^\alpha\widehat{u}(h) = (-1)^{|\alpha|}u(\widehat{D^\alpha h}) = (-1)^{|\alpha|}u((2\pi i)^{|\alpha|}x^\alpha\widehat{h}) = (-2\pi i)^{|\alpha|}\widehat{x^\alpha u}(h),$$

where we used Proposition 54.

Finally, we write

$$\widehat{\widehat{u}}(h) = u(\widehat{\widehat{h}}) = u(\check{h}) = \check{u}(h),$$

where we used Fourier inversion. □

14. Periodic distributions

Definition 88. We write $\mathcal{D}(\mathbf{T}^d)$ for the space of functions in $C^\infty(\mathbf{R}^d)$ that are periodic under the action of $\mathbf{Z}^d$, that is, $f \in C^\infty(\mathbf{R}^d)$ is in $\mathcal{D}(\mathbf{T}^d)$ if and only if $\tau_a f = f$ holds for all $a \in \mathbf{Z}^d$.

We consider this space endowed with the metric

$$\mathrm{dist}_{\mathcal{D}}(f,g) = \sum_\alpha \min(\|D^\alpha(f-g)\|_\infty, 2^{-|\alpha|}).$$

We write $\mathcal{D}'(\mathbf{T}^d)$ for the space of continuous functionals on $\mathcal{D}(\mathbf{T}^d)$.

Definition 89. For $f \in L^1(\mathbf{T}^d)$, we write

$$T_f(g) = \int_{[0,1]^d} f(x)g(x)dx$$

for $g \in (\mathbf{T}^d)$ which defines an element of $\mathcal{D}'(\mathbf{T}^d)$.

Definition 90. Let $u \in \mathcal{D}'(T^d)$. Its Fourier coefficients are defined as $\widehat{u}(a) = u(e^{-2\pi i\langle x,a\rangle})$ for $a \in \mathbf{Z}^d$.

Proposition 91. *Let $u \in \mathcal{D}'(T^d)$. Then there is some $N \in \mathbf{Z}_{\geq 0}$ such that*

$$|\widehat{u}(\xi)| \leq (2+|\xi|)^N$$

for all $\xi \in \mathbf{Z}_{\geq 0}$ and

$$u = \sum_{\xi\in\mathbf{Z}^d} \widehat{u}(\xi)e^{2\pi i\langle x,\xi\rangle}$$

in the sense that the partial sums

$$\sum_{|\xi|\leq k} \widehat{u}(\xi)T_{e^{2\pi i\langle x,\xi\rangle}}$$

converge to u in the weak-$*$ topology.

Proof. Suppose to the contrary that there is a sequence ξ_N such that $|\widehat{u}(\xi_N)| > (2+|\xi_N|)^N$ for all N. It is an easy calculation to show that

$$\|D^\alpha e^{-2\pi i\langle x,\xi\rangle}\|_\infty \leq (2\pi|\xi|)^{|\alpha|}$$

It follows that

$$f_N := (2+|\xi_N|)^{-N}e^{-2\pi i\langle x,\xi_N\rangle} \to 0$$

in $\mathcal{D}(\mathbf{T}^d)$. On the other hand $|u(f_N)| > 1$ for all N, which contradicts to the continuity of u.

Let $f \in \mathcal{D}(\mathbf{T}^d)$, and write

$$\sum_{|\xi|\leq k} \widehat{u}(\xi)T_{e^{2\pi i\langle x,\xi\rangle}}(h) = \sum_{|\xi|\leq k} \widehat{u}(\xi)\int h(x)e^{2\pi i\langle x,\xi\rangle}dx = u\Big(\sum_{|\xi|\leq k} \widehat{h}(-\xi)e^{2\pi i\langle x,-\xi\rangle}\Big) \to u(h)$$

using the continuity of u, Fourier inversion, and the the decay of Fourier coefficients for smooth functions. $\square$

Recall the linear map $P : L^1(\mathbf{R}^d) \to L^1(\mathbf{T}^d)$ defined by

$$Pf = \sum_{a \in \mathbf{Z}^d} \tau_a f,$$

Restricting it to $\mathcal{S}(\mathbf{R}^d)$, the image is in $\mathcal{D}(\mathbf{T}^d)$, and this map is continuous with respect to the topologies we defined on these spaces.

We want to show that $P : \mathcal{S}(\mathbf{R}^d) \to \mathcal{D}(\mathbf{T}^d)$ is onto. To this end, we construct a function $\chi \in C_c^\infty(\mathbf{R}^d)$ such that $P\chi = 1$. Then $P\chi f = f$ and $\chi f \in \mathcal{S}(\mathbf{R}^d)$. Let $f \in C_c^\infty(\mathbf{R}^d)$ such that $f(x) \geq 0$ for all x and that $f(x) > 0$ for all $x \in [0,1]^d$. Then $Pf(x) > 0$ for all $x \in \mathbf{R}^d$ and $\chi(x) := f(x)/Pf(x)$ satisfies the required properties.

Since $P : \mathcal{S}(\mathbf{R}^d) \to \mathcal{D}(\mathbf{T}^d)$ is continuous, it induces a map $P^*\mathcal{D}'(\mathbf{T}^d) \to \mathcal{S}'(\mathbf{R}^d)$ via

$$P^*u(f) = u(Pf).$$

In what follows our aim is to identify the image of $\mathcal{D}'(\mathbf{T}^d)$ in $\mathcal{S}'(\mathbf{R}^d)$ and understand the relationship between the Fourier coefficients of $u \in \mathcal{D}'(\mathbf{T}^d)$ and the Fourier transform of Pu.

Lemma 92. *The map P^* is linear, injective and continuous with respect to the weak-$*$ topologies.*

Proof. P^* is linear because it is the composition of linear maps. Let $u \neq 0 \in \mathcal{D}'(\mathbf{T}^d)$. Then there is some $f \in \mathcal{D}(T^d)$ such that $u(f) \neq 0$. Therefore, $P^*u(\chi f) = u(P\chi f) = u(f) \neq 0$, so P^* is injective.

Now let $u, u_1, u_2, \ldots \mathcal{D}'(T^d)$ be such that $u_n \overset{*}{\rightharpoonup} u$. Let $f \in \mathcal{S}(\mathbf{R}^d)$, then

$$P^*u_n(f) = u_n(Pf) \to u(Pf) = P^*u(f),$$

hence $P^*u_n \overset{*}{\rightharpoonup} P^*u$. □

Corollary 93. *Suppose $u \in P^*(\mathcal{D}'(\mathbf{T}^d)) \subset \mathcal{S}'(\mathbf{R}^d)$. Then*

$$u = \sum_{\xi \in \mathbf{Z}^d} u(\chi \cdot e^{-2\pi i \langle x, \xi \rangle}) e^{2\pi i \langle x, \xi \rangle}$$

in the same sense as in Proposition 91.

Proof. We need to show two things. The first is that $P^* T_{e^{2\pi i \langle x,\xi \rangle}} = T_{e^{2\pi i \langle x,\xi \rangle}}$, where the T notation designates a distribution in $\mathcal{D}'(\mathbf{T}^d)$ on the left and a distribution in $\mathcal{S}'(\mathbf{T}^d)$ on the right. This is equivalent to the simple identity

$$\int_{\mathbf{T}^d} Pf(x) e^{2\pi i \langle x,\xi \rangle} dx = \int_{\mathbf{R}^d} f(x) e^{2\pi i \langle x,\xi \rangle} dx$$

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

The other fact we need is that if $v \in \mathcal{D}'(\mathbf{T}^d)$ and $u = P^*v$, then

$$u(\chi e^{-2\pi i \langle x,\xi \rangle}) = v(P\chi e^{-2\pi i \langle x,\xi \rangle}) = v(e^{-2\pi i \langle x,\xi \rangle}) = \widehat{v}(\xi).$$

□

Definition 94. Let $N \in \mathbf{Z}_{\geq 0}$. Given a sequence of points $(x_j) \subset \mathbf{R}^d$ and a sequence $(b_j) \subset \mathbf{C}$ such that

$$\sum_{j:|x_j|<R} |b_j| \leq R^N$$

for all $R \in \mathbf{R}_{\geq 1}$. we define the tempered distribution

$$\Big(\sum_j b_j\delta_j\Big)(f) = \sum_j b_j f(x_j).$$

Note that $\sum_j b_j\delta_j$ may not be a measure.

Corollary 95. *Suppose $u \in P^*(\mathcal{D}'(\mathbf{T}^d)) \subset \mathcal{S}'(\mathbf{R}^d)$. Then*

$$\widehat{u} = \sum_{\xi\in\mathbf{Z}^d} u(\chi \cdot e^{-2\pi i\langle x,\xi\rangle})\delta_\xi.$$

Proof. It is immediate from the definitions that

$$\sum_{|\xi|\leq N} u(\chi \cdot e^{-2\pi i\langle x,\xi\rangle})\delta_\xi \xrightarrow{*} \sum_{\xi\in\mathbf{Z}^d} u(\chi \cdot e^{-2\pi i\langle x,\xi\rangle})\delta_\xi$$

as $N \to \infty$. By the continuity of the Fourier transform on $\mathcal{S}'(\mathbf{R}^d)$, we can conclude by

$$\widehat{T}_{e^{2\pi i\langle x,\xi\rangle}}(h) = \int e^{2\pi i\langle x,\xi\rangle}\widehat{h}(x)dx = h(\xi) = \delta_\xi(h).$$

$\square$

Definition 96. A tempered distribution $u \in \mathcal{S}'(\mathbf{R}^d)$ is periodic if $\tau_a u = u$ for all $a \in \mathbf{Z}^d$.

Proposition 97. *Let $u \in \mathcal{S}'(\mathbf{R}^d)$. Then $u \in P^*(\mathcal{D}'(\mathbf{T}^d))$ if and only if u is periodic.*

Proof. If $v \in \mathcal{D}'(\mathbf{T}^d)$, then

$$\tau_a P^*v(h) = v(\tau_{-a}Ph) = v(Ph) = P^*v(h).$$

This proves that distributions in $P^*(\mathcal{D}'(\mathbf{T}^d))$ are periodic.

Now let u be a periodic distribution. We show that $u = P^*v$ for the distribution $v \in \mathcal{D}'(\mathbf{T}^d)$ defined by $v(h) = u(\chi h)$. We write

$$P^*v(h) = v(Ph) = u(\chi Ph).$$

We need to show that $u(h) = u(\chi Ph)$ for all $h \in \mathcal{S}'(\mathbf{R}^d)$. Since $P\chi = 1$, we can write

$$h = \sum_{a\in\mathbf{Z}^d} (\tau_{-a}\chi)h. \tag{7}$$

We note that

$$u((\tau_{-a}\chi)h) = u(\tau_a((\tau_{-a}\chi)h)) = u(\chi\tau_a h)$$

using that u is periodic. We also have

$$\chi P h = \sum_{a\in\mathbf{Z}^d} \chi\tau_a h \tag{8}$$

The statement will follow if we prove that both of the sums (7) and (8) converge in $\mathcal{S}(\mathbf{R}^d)$.

We prove this only for (7), as the other sum is similar and easier. The derivative $D^\alpha(\tau_{-a}\chi)h$ is a linear combination of products of a derivative of $\tau_{-a}\chi$ and a derivative of h. The order of the derivatives are at most $|\alpha|$. The derivatives of $\tau_{-a}\chi$ are bounded by a constant depending only on $|\alpha|$ and χ. Also $\tau_{-a}\chi h(x) = 0$ if $|x| > |a| - C_1$ for a constant C_1 depending on χ. This implies that $|x| + 1 \geq |a|/C_1$ on the support of $\tau_{-a}\chi h$. It follows that

$$p_N(\tau_{-a}(\chi)h) \leq C_2(N)p_{N+d+1}(h)\frac{C_1^{d+1}}{|a|^{d+1}},$$

where $C_2(N)$ comes from the contribution of the derivatives of $\tau_{-a}\chi$. This is enough to show that the sum (7) is convergent with respect to the p_N norm for all N, therefore in $\mathcal{S}(\mathbf{R}^d)$. □

Example 98. Let $u = \sum_{a\in\mathbf{Z}^d} \delta_a$. This is easily seen to be a periodic distribution. Note that

$$u(\chi \cdot e^{-2\pi i\langle x,\xi\rangle}) = \sum_{a\in\mathbf{Z}^d} \chi(a)e^{-2\pi i\langle a,\xi\rangle} = P\chi(0) = 1$$

for all $\xi \in \mathbf{Z}^d$. It follows that $\widehat{u} = u$.

Observe that this is equivalent to the Poisson summation formula for Schwartz functions.

15. Sobolev spaces

In this section, we introduce some function spaces that are particularly useful in the study of PDEs.

Fix an open set $\Omega \subset \mathbf{R}^d$.

Definition 99. Let $f \in L^1_{loc}(\Omega)$ and let $\alpha \in \mathbf{Z}^d_{\geq 0}$. We say that a function $g \in L^1_{loc}(\Omega)$ is a weak derivative of f of order α, if $D^\alpha T_f = T_g$.

Unwinding the definitions, this is equivalent to

$$\int_\Omega g h dx = (-1)^{|\alpha|}\int_\Omega f D^\alpha h dx$$

for all $h \in C_c^\infty(\Omega)$.

In this case we write $D^\alpha f$ for g.

Definition 100. Let $p \in [1,\infty]$ and let $k \in \mathbf{Z}_{>0}$. We say that $f \in L^p(\Omega)$ belongs to the Sobolev space $W^{k,p}(\Omega)$ if it has weak derivatives $D^\alpha f \in L^p(\Omega)$ for all $\alpha \in \mathbf{Z}^d_{\geq 0}$ with $|\alpha| \leq k$.

We endow this space with the norm

$$\|f\|_{W^{k,p}(\Omega)} := \sum_{|\alpha|\le k} \|D^\alpha f\|_{L^p(\Omega)}.$$

It can be proved that Sobolev spaces are Banach spaces with respect to the above norms.

It is often easier to find "weak" solutions of PDEs in Sobolev spaces rather than in $C^k(\Omega)$. It is then of great interest to find sufficient conditions for when these weak solutions are "classical" solutions, that is when they belong to $C^k(\Omega)$. A very useful tool for this is the Sobolev embedding theorem, which we are going to discuss now.

Definition 101. Let $k \in \mathbf{Z}_{\ge 0}$ and $\kappa \in [0,1]$. The Hölder space $C^{k,\kappa}(\mathbf{R}^d)$ is the space of functions in $C^k(\mathbf{R}^d)$ for which

$$\|f\|_{C^{k,\kappa}(\mathbf{R}^d)} := \sum_{|\alpha|\le k} \|D^\alpha f(x)\|_\infty + \sum_{|\alpha|=k} \sup_{x,y\in\mathbf{R}^d} \frac{|D^\alpha f(x) - D^\alpha f(y)|}{|x-y|^\kappa} < \infty.$$

Definition 102. We say that a Banach space $(X, \|\cdot\|_X)$ continuously embeds into another Banach space $(Y, \|\cdot\|_Y)$ if $X \subset Y$ and $\|f\|_X \le C\|f\|_Y$ for some constant $C = C(X,Y)$.

Theorem 103 (Sobolev embedding)**.** *Let $1 \le p \le q < \infty$ and $k \ge l \in \mathbf{Z}_{\ge 0}$. Suppose*

$$\frac{l}{n} \ge \frac{1}{p} - \frac{1}{q}.$$

Then $W^{p,k}(\mathbf{R}^d)$ embeds continuously into $W^{q,k-l}(\mathbf{R}^d)$.

Let $1 \le p < \infty$, $k \ge l \in \mathbf{Z}_{\ge 0}$ and $\gamma \in (0,1)$ such that

$$\frac{l-\gamma}{n} \ge \frac{1}{p}.$$

Then $W^{p,k}(\mathbf{R}^d)$ embeds continuously into $C^{k-l,\gamma}(\mathbf{R}^d)$.

Note that elements of $W^{p,k}(\mathbf{R}^d)$ are equivalence classes of functions that coincide almost everywhere. What we mean in the second statement is that the equivalence class contains an element that belongs to $C^{l-k,\gamma}(\mathbf{R}^d)$.

The theorem remains true if we replace $\mathbf{R}^d$ with a sufficiently nice domain Ω, for example a bounded open set with C^1 boundary. We do not discuss what this means precisely.

In this course, we only prove a special case of the above theorem. When $p = 2$, the space $W^{2,k}(\Omega)$ is also denoted by $H^k(\Omega)$. When $\Omega = \mathbf{R}^d$, this space can also be described using the Fourier transform.

Proposition 104. *Let $f \in L^2(\mathbf{R}^d)$ and let $k \in \mathbf{Z}_{\ge 0}$. We have $f \in W^{2,k}(\mathbf{R}^d)$ if and only if*

$$\|f\|_{H^k} := \Big(\int |\widehat{f}(\xi)|^2 (1+|\xi|^2)^k d\xi\Big)^{1/2} < \infty.$$

Moreover, the norms $\|f\|_{W^{2,k}}$ *and* $\|f\|_{H^k}$ *are equivalent, that is, there is a constant* $C = C(d,k)$ *such that*

$$C^{-1}\|f\|_{H^k} \leq \|f\|_{W^{2,k}} \leq C\|f\|_{H^k}$$

Proof. This proof relies on the fact that if f has a weak derivative $D^\alpha f \in L^2(\mathbf{R}^d)$, then

$$\widehat{D^\alpha f}(\xi) = (2\pi i)^{|\alpha|}\xi^\alpha \widehat{f}(\xi).$$

Conversely if $\xi^\alpha \widehat{f}(\xi) \in L^2(\mathbf{R}^d)$, then f has a weak derivative $D^\alpha f \in L^2(\mathbf{R}^d)$. This can be proved using tempered distributions. (Caveat: this requires proving that T_f is a tempered distribution when $f \in L^2(\mathbf{R}^d)$, and we did this under a slightly different condition, but the same proof works.)

Now suppose $f \in W^{2,k}(\mathbf{R}^d)$. Then for any $\alpha \in \mathbf{Z}^d_{\geq 0}$ with $|\alpha| \leq k$

$$\int (2\pi)^{2|\alpha|}|\widehat{f}(\xi)|^2\xi^{2\alpha}d\xi = \|(2\pi i)^{|\alpha|}\xi^\alpha\widehat{f}(\xi)\|_2^2 = \|D^\alpha f\|_2^2 \leq \|f\|^2_{W^{2,k}}.$$

Observe that the H^k norm is a certain linear combination of expressions that appear on the left hand side of this inequality so $\|f\|_{H^k} \leq C\|f\|_{W^{2,k}}$ follows.

Conversely, suppose $\|f\|_{H^k} < \infty$. Let $\alpha \in \mathbf{Z}^d_{\geq 0}$ with $|\alpha| \leq k$. Then

$$\|D^\alpha f\|_2 = (2\pi)^{|\alpha|}\|\xi^\alpha \widehat{f}\|_2 \leq (2\pi)^{|\alpha|}\|f\|_{H^k}$$

so $\|f\|_{W^{k,p}} \leq C\|f\|_{H^k}$ follows. □

The definition of the H^k norm makes sense if k is a real number even when it is negative! Motivated by this, we make the following definition.

Definition 105. Let $s \in \mathbf{R}$. The space $H^s(\mathbf{R}^d)$ consists of tempered distributions $u \in \mathcal{S}'(\mathbf{R}^d)$ such that $\widehat{u} \in L^2_{loc}(\mathbf{R}^d)$ and

$$\|u\|_{H^s(\mathbf{R}^d)} = \int |\widehat{u}|^2(1+|\xi|^2)^s d\xi < \infty.$$

It is immediate from the definition that $H^s(\mathbf{R}^d)$ is a Hilbert space with the inner product

$$\langle f, g\rangle_{H^s} := \int (1+|\xi|^2)^s \widehat{f}\overline{\widehat{g}}d\xi.$$

For $s = 1$, we have the useful formula

$$\langle f, g\rangle_{H^1} := \int f\overline{g} + \frac{1}{4\pi^2}\nabla f \cdot \overline{\nabla g}dx.$$

Theorem 106 (Sobolev embedding)**.** *Let* $d \in \mathbf{Z}_{>0}$ *and* $s > d/2$*. Then* $H^s(\mathbf{R}^d)$ *is continuously embedded in* $L^\infty(\mathbf{R}^d)$*, that is,*

$$\|f\|_\infty \leq C\|f\|_{H^s}$$

for all $f \in H^s(\mathbf{R}^d)$, where C is a constant that depends only on d and s.

Proof. We show that $\|\widehat{f}\|_1 \le C\|f\|_{H^s}$, and then the claim follows by the Fourier inversion formula. To this end, we apply the Cauchy-Schwarz inequality as

$$\begin{aligned}\int |\widehat{f}(\xi)|d\xi &= \int |\widehat{f}(\xi)|(1+|\xi|^2)^{s/2}(1+|\xi|^2)^{-s/2}d\xi\\ &\le \||\widehat{f}(\xi)|(1+|\xi|^2)^{s/2}\|_2\Big(\int(1+|\xi|^2)^{-s}d\xi\Big)^{1/2}\\ &= \|f\|_{H^s}\Big(C\int_0^\infty(1+r^2)^{-s}r^{d-1}d\xi\Big)^{1/2}.\end{aligned}$$

Since $2s > d$, the last integral is finite. □

Corollary 107 (Sobolev embedding)**.** *Let $d \in \mathbf{Z}_{>0}$, $k \in \mathbf{Z}_{\ge 0}$, and let $s > d/2 + k$. Then $H^s(\mathbf{R}^d)$ continuously embeds into $C^k(\mathbf{R}^d)$.*

Proof. If $\alpha \in \mathbf{Z}_{\ge 0}^d$ with $|\alpha| \le k$, then $D^\alpha f \in H^{s-k}(\mathbf{R}^d)$ and

$$\|D^\alpha f\|_{H^{s-k}} \le (2\pi)^{|\alpha|}\|f\|_{H^s}.$$

This follows by the formula

$$\widehat{D^\alpha f} = (2\pi i)^{|\alpha|}\xi^\alpha \widehat{f}$$

and the definition of Sobolev norms. Therefore

$$\|f\|_{C^k} = \sum_{\alpha:|\alpha|\le k}\|D^\alpha f\|_\infty \le C\|f\|_{H^s}.$$

It remains to show that $D^\alpha f$ is continuous for all α. To this end, let $\varphi \in C_c^\infty(\mathbf{R}^d)$ be with $\int \varphi dx = 1$ and $\varphi(x) \ge 0$ for all x. Let $\varphi_r(x) = r^d\varphi(rx)$. We claim that $\varphi_r * f \to f$ in $H^s(\mathbf{R}^d)$ as $r \to \infty$. Note that

$$\widehat{\varphi_r * f}(\xi) = \widehat{\varphi}(r^{-1}\xi)\cdot\widehat{f}(\xi),$$

which is bounded by $|\widehat{f}(\xi)|$ and converges to $\widehat{f}(\xi)$ pointwise. So the claim follows by dominated convergence.

Using the inequality

$$\|\varphi_r * f - f\|_\infty \le C\|\varphi_r * f - f\|_{H^s},$$

we see that $\varphi_r * f$ converges to f uniformly. The functions $\varphi_r * f$ are continuous being the convolutions of L^2 functions. We conclude that f is continuous. Applying the argument for $D^\alpha f \in H^{s-k}(\mathbf{R}^d)$ and using $s - k > d/2$, we conclude that $D^\alpha f$ is also continuous. □

15.1. **Traces.** In PDEs one often wants to specify conditions for the behaviour of the solution at the boundary of a domain. When we look for solutions in a Sobolev space, this causes a problem, because such functions are defined only almost everywhere and the boundary is of measure 0. There is a remedy when Sobolev embedding implies the existence of a continuous representative. However, it is often desirable to do this under weaker regularity assumptions, which can be achieved using the trace operator.

We restrict our discussion to the half space $\Omega := \mathbf{R}_{>0} \times \mathbf{R}^{d-1}$, but the results hold true for more general sufficiently regular domains. For $s \in \mathbf{R}_{\geq 0}$, we define the Sobolev space $H^s(\Omega)$ as the space of functions that are restrictions of functions in $H^s(\mathbf{R}^d)$, and we endow this space with the norm

$$\|f\|_{H^s} = \inf(\|g\|_{H^s} : g \in H^s(\mathbf{R}^d), g|_\Omega = f).$$

It is a non-trivial fact that this norm is equivalent to the $W^{s,2}(\Omega)$ norm defined above when $s \in \mathbf{Z}_{\geq 0}$.

Theorem 108 (Trace theorem)**.** *Let $\Omega = \mathbf{R}_{>0} \times \mathbf{R}^{d-1}$ and let $s > 1/2$. Then there is a bounded linear operator $T : H^s(\Omega) \to H^{s-1/2}(\partial\Omega)$ such that*

$$Tf = f|_{\partial\Omega}$$

when $f \in H^s(\Omega) \cap C(\overline{\Omega})$.

The operator that appears in this theorem is called the trace operator. The proof of this result will be on the example sheet.

15.2. **Rellich-Kondrachov theorem.** Let $\Omega \subset \mathbf{R}^d$ be an open set. We may regard elements of $C_c^\infty(\Omega)$ as functions on $\mathbf{R}^d$ if we extend them to take the value 0 outside Ω. We define $H_0^1(\Omega)$ as the closure of $C_c^\infty(\Omega)$ in $H^1(\mathbf{R}^d)$. Formally, the elements of $H_0^1(\mathbf{R}^d)$ are functions on $\mathbf{R}^d$, but they vanish outside Ω (almost surely), so there is no harm in identifying the elements of $H_0^1(\Omega)$ with their restriction to Ω. If Ω is a sufficiently nice domain, for example $\mathbf{R}_{>0} \times \mathbf{R}^{d-1}$ or a bounded domain with C^1 boundary, then $H_0^1(\Omega)$ can be identified with the kernel of the trace operator $H^1(\Omega) \to H^{1/2}(\partial\Omega)$.

Warning: the above discussion cannot be generalized to H^s in place of H^1 for general s. In particular, the analogous definition for $H_0^{1/2}$ would be the wrong one.

The space $H_0^1(\Omega)$ being a closed subspace of $H^1(\mathbf{R}^d)$ is a Hilbert space, so we know that its unit ball $B_{H_0^1}(\Omega)$ is compact in the weak topology, so any sequence has a weakly convergent subsequence. In some applications, it would be desirable to have a norm convergent subsequence. The next theorem shows that this is possible in the L^2 norm.

Theorem 109 (Rellich-Kondrachov)**.** *Let $\Omega \subset \mathbf{R}^d$ be a bounded domain. Let $(f_n) \subset H_0^1(\Omega)$ be a bounded sequence. Then there is a subsequence (f_{n_j}) and $f \in H_0^1(\Omega)$ such that*

(1) *$f_{n_j} \rightharpoonup f$ in the weak topology of $H_0^1(\Omega)$, and*
(2) *$f_{n_j} \to f$ in $L^2(\Omega)$.*

Corollary 110. *Let $\Omega \subset \mathbf{R}^d$ be bounded. Then $B_{H_0^1(\Omega)}$ is compact in the norm topology of $L^2(\Omega)$.*

This can be thought of as a Sobolev space version of the Arzelà Ascoli theorem. Indeed, a consequence of that theorem is that the unit ball of $C^1(\overline{\Omega})$ is compact in $C(\overline{\Omega})$.

Proof. Using the Banach–Alaoglu theorem for the Hilbert space $H_0^1(\Omega)$, we can find a subsequence (f_{n_j}) that is weakly convergent. Let $f = \text{w-lim}\, f_{n_j}$, and $g_j = f_{n_j} - f$.

It remains to show that $g_n \to 0$ in the norm topology of $L^2(\Omega)$. We first show that $g_n \rightharpoonup 0$ in $L^2(\Omega)$. We know that

$$\langle g_n, h_1 \rangle_{H^1} = \int (|\xi|^2 + 1)\widehat{g}_n \overline{\widehat{h}_1} d\xi \to 0$$

for any fixed $h_1 \in H^1(\mathbf{R}^d)$. If $h_2 \in L^2(\mathbf{R}^d)$, then $\widehat{h}_2 \in L^2(\mathbf{R}^d)$ and

$$\int (|\xi|^2 + 1)|(|\xi|^2 + 1)^{-1}\widehat{h}_2|^2 d\xi < \infty.$$

It follows that $(|\xi|^2 + 1)^{-1}\widehat{h}_2 = \widehat{h}_1$ for some $h_1 \in H^1(\mathbf{R}^d)$. We conclude that

$$\langle g_n, h_2 \rangle_{L^2} = \int \widehat{g}_n \overline{\widehat{h}_2} d\xi = \int (1 + |\xi|^2)\widehat{g}_n \overline{\widehat{h}_1} d\xi \to 0,$$

so $g_n \rightharpoonup 0$ in $L^2(\Omega)$.

Now let $h_2 = e^{2\pi i \langle x, \xi \rangle}|_\Omega \in L^2(\Omega)$. We conclude that

$$\widehat{g}_n(\xi) = \langle g_n, h_2 \rangle_{L^2} \to 0$$

for all ξ. We also have

$$|\widehat{g}_n(\xi)| \le \int |g_n| d\xi \le \|g_n\|_{L^2} \|1_\Omega\|_2 \le \|g_n\|_{H^1} \|1_\Omega\|.$$

It follows by the dominated convergence theorem that

$$\int_{|\xi| < R} |\widehat{g}_n(\xi)|^2 d\xi \to 0$$

for any fixed $R > 0$. On the other hand,

$$\int_{|\xi| > R} |\widehat{g}_n(\xi)|^2 d\xi \le R^{-2} \int_{|\xi| > R} (1 + |\xi|^2)|\widehat{g}_n(\xi)|^2 d\xi \le R^{-2} \|g_n\|_{H^1}.$$

It follows that

$$\begin{aligned}\limsup \|g_n\|_{L^2} &= \limsup \int_{|\xi|<R} |\widehat{g}_n(\xi)|^2 d\xi + \limsup \int_{|\xi|>R} |\widehat{g}_n(\xi)|^2 d\xi \\ &\leq 0 + R^{-2} \limsup \|g_n\|_{H^1}.\end{aligned}$$

Taking $R \to \infty$, we conclude $\limsup \|g_n\|_{L^2} = 0$, as required.

The corollary follows from the fact that compactness and sequential compactness is equivalent in metric spaces. □

16. Elliptic PDEs

We are interested in solving the following PDE. Let $\Omega \subset \mathbf{R}^d$ be an open set. Given a function $f : \Omega \to \mathbf{R}$, we look for a function $u : \Omega \to \mathbf{R}$ that satisfies

$$\begin{aligned}-\frac{1}{4\pi^2}\Delta u + u = f \qquad &\text{in } \Omega, \\ u = 0 \qquad &\text{on } \partial\Omega.\end{aligned}$$

A classical solution is a function $u \in C^2(\Omega) \cap C(\overline{\Omega})$ that satisfies the equations. At first, we look for a weak solution in a Sobolev space. To encode the boundary condition, we seek the solution in $H^1_0(\Omega)$. It is possible to prove that if Ω and f are sufficiently regular then a classical solution of the PDE satisfies the boundary condition if and only if it is in $H^1_0(\Omega)$, but this goes beyond this course.

It is not immediately obvious what the PDE itself means for $u \in H^1_0(\Omega)$, because such functions have weak derivatives of first order only. However, we can make sense of it if we work in $\mathcal{D}'(\Omega)$. In that sense, $u \in H^1_0(\Omega)$ satisfies the equation if and only if

$$-\frac{1}{4\pi^2}\Delta T_u(v) + T_u(v) = T_f(v)$$

for all $v \in \mathcal{D}(\Omega)$. We can compute

$$\begin{aligned}-\Delta T_u(v) &= -\sum_j (D^{e_j} D^{e_j} T_u)(v) = \sum_j (D^{e_j} T_u)(D^{e_j} v) \\ &= \sum_j T_{D^{e}_j u}(D^{e_j} v) = \int \nabla u \cdot \nabla v dx.\end{aligned}$$

We see that the equation is equivalent to that

$$\langle u, v\rangle_{H^1} = \langle f, v\rangle_{L^2}.$$

holds for all $v \in \mathcal{D}(\Omega)$. (This involves writing $\overline{v}$ in the place of v.)

Note that $v \mapsto \overline{\langle f, v\rangle_{L^2}}$ is a continuous linear functional on $H^1_0(\Omega)$ for any $f \in L^2(\Omega)$. One way to see this is that the H^1 topology is stronger than L^2 so it is easier for a functional to be continuous than on L^2. We may conclude two things from this. One is that

$$\langle u, v\rangle_{H^1} = \langle f, v\rangle_{L^2}.$$

is valid for all $v \in H_0^1(\Omega)$ by continuity, the other is that using the Riesz representation theorem for the Hilbert space $H_0^1(\Omega)$, there is a unique solution of the equation in $H_0^1(\Omega)$.

Theorem 111. *For all $f \in L^2(\Omega)$, there is a unique $u \in H_0^1(\Omega)$ such that*

$$\langle u, v\rangle_{H^1} = \langle f, v\rangle_{L^2} \tag{9}$$

for all $v \in \mathcal{D}(\Omega)$ or equivalently for all $v \in H_0^1(\Omega)$.

Next we will investigate when the weak solution in the above theorem is a classical solution. As a warm up, we consider the case $\Omega = \mathbf{R}^d$. In this case, $H_0^1(\mathbf{R}^d) = H^1(\mathbf{R}^d)$, because $\mathcal{D}(\mathbf{R}^d)$ is dense in $H^1(\mathbf{R}^d)$. (See the example sheet.)

Taking Fourier transform, we see that (9) is equivalent to

$$\int (1 + |\xi|^2)\widehat{u}\overline{\widehat{v}}d\xi = \int \widehat{f}\overline{\widehat{v}}d\xi.$$

We see that $u = ((1 + |\xi|^2)^{-1}\widehat{f})^{\wedge\vee}$ is a solution of the equation. (In the general case, where Ω is not $\mathbf{R}^d$, this function would still solve (9), however, it likely does not belong to $H_0^1(\Omega)$.) From this, we immediately deduce the following.

Theorem 112. *Suppose $\Omega = \mathbf{R}^d$. If $f \in H^s(\mathbf{R}^d)$ for some $s \geq 0$, then the unique solution u of (9) satisfies $u \in H^{s+2}(\mathbf{R}^d)$. Moreover, $\|u\|_{H^{s+2}} = \|f\|_{H^s}$. If $s > d/2$, the solution is a classical solution.*

We want to prove a similar result for general domains. The behaviour of the solution near the boundary is beyond the scope of this course, but we will study the regularity of the solution in the interior of the domain. To this end, we make a definition.

Definition 113. Let $\Omega \subset \mathbf{R}^d$ be an open set and let $s \in \mathbf{R}_{\geq 0}$. A function $f : \Omega \to \mathbf{C}$ is in the space $H^s_{loc}(\Omega)$ if $fh \in H_0^s(\Omega)$ for all $h \in \mathcal{D}(\Omega)$.

Theorem 114. *Let $\Omega \subset \mathbf{R}^d$ be an open set and let $s \in \mathbf{R}_{\geq 0}$. Suppose $f \in H^s_{loc}(\Omega)$, then the unique solution $u \in H_0^1(\Omega)$ of (9) satisfies $u \in H^{s+2}_{loc}(\Omega)$.*

Lemma 115. *Let $u \in H^1(\mathbf{R}^d)$, $\alpha \in \mathbf{Z}^d_{\geq 0}$ with $|\alpha| = 1$, $f \in \mathcal{D}(\mathbf{R}^d)$. Then we have*

$$D^\alpha(fu) = (D^\alpha f)u + f(D^\alpha u).$$

Proof. Let $h \in \mathcal{D}(\mathbf{R}^d)$. Then by the definition of weak derivative, we have

$$\int D^\alpha(fu)hdx = -\int fuD^\alpha hdx = -\int uD^\alpha(fh) - u(D^\alpha f)hdx$$
$$= \int f(D^\alpha u)h + (D^\alpha f)uhdx.$$

□

If $u \in H^s(\mathbf{R}^d)$ for some $s \in \mathbf{Z}_{\geq 0}$, then it is easy to see from the lemma that $uh \in H^s(\mathbf{R}^d)$ for all $h \in \mathcal{D}(\mathbf{R}^d)$. This remains true for non-integral s, but we do not prove it.

Proof. The strategy of the proof is the following. Suppose we already know that $u \in H^t_{loc}$ for some $t \in [1, s+2]$. We will show that for all $h \in \mathcal{D}(\Omega)$, uh satisfies (9) for all $v \in \mathcal{D}(\mathbf{R}^d)$ for an appropriate $g \in H^{\min(t-1,s)}(\mathbf{R}^d)$. This will imply that $uh \in H^{\min(t+1,s+2)}(\mathbf{R}^d)$ by Theorem 112. Therefore, $u \in H^{\min(t+1,s+2)}_{loc}(\Omega)$. The function g will depend on u, f and h, but this is OK. We already know that $u \in H^1_{loc}(\Omega)$, and we can use induction to show that $u \in H^t_{loc}(\Omega)$ for $t = 2, 3, \ldots$ up to $t = \lceil s \rceil + 1$ and finally $t = s + 2$.

Let $u \in H^1_0(\Omega)$ be the solution of (9), $h \in \mathcal{D}(\Omega)$ and $v \in \mathcal{D}(\mathbf{R}^d)$. We know that

$$\int \frac{1}{4\pi^2} \nabla u \cdot \overline{\nabla(\overline{h}v)} + u\overline{\overline{h}v} dx = \int f\overline{\overline{h}v} dx$$

by (9) applied for the test function $\overline{h}v$. We want to show that there is a function g such that

$$\int \frac{1}{4\pi^2} \nabla(uh) \cdot \overline{\nabla v} + uh\overline{v} dx = \int g\overline{v} dx,$$

where g may depend on u and h but not on v.

To this end, we make the following computation using Lemma 115 and at the end integration by parts, which in this case is just the definition of weak derivatives.

$$\begin{aligned}
&\int \nabla(uh) \cdot \overline{\nabla v} - \nabla u \cdot \overline{\nabla \overline{h} v} dx \\
&\quad = \int h\nabla u \cdot \nabla \overline{v} + u \nabla h \cdot \nabla \overline{v} - (\nabla u \cdot \nabla h)\overline{v} - h \nabla u \cdot \nabla \overline{v} dx \\
&\quad = \int u \nabla h \cdot \nabla \overline{v} - (\nabla u \cdot \nabla h)\overline{v} dx \\
&\quad = \int -(\nabla \cdot u \nabla h)\overline{v} - (\nabla u \cdot \nabla h)\overline{v} dx \\
&\quad = \int -(u\Delta h + 2\nabla u \cdot \nabla h)\overline{v} dx.
\end{aligned}$$

Therefore, we need a function g that satisfies

$$\int g\overline{v} dx - \int f\overline{\overline{h}v} dx = -\frac{1}{4\pi^2} \int (u\Delta h + 2\nabla u \cdot \nabla h)\overline{v} dx.$$

We see that we can take

$$g = fh - \frac{1}{4\pi^2}(u\Delta h + 2\nabla u \cdot \nabla h).$$

We have $fh \in H_0^s(\Omega) \subset H^s(\mathbf{R}^d)$ by the definition of $H^s_{loc}(\Omega)$. We also have $u\Delta h \in H^t(\mathbf{R}^d)$ for the same reason. In addition, $\nabla u \cdot \nabla h = \nabla \cdot (u\nabla h) - u\Delta h$, and $u\nabla h \in H^t(\mathbf{R}^d)$, so $\nabla u \cdot \nabla h \in H^{t-1}(\mathbf{R}^d)$. We conclude $g \in H^{\min(t-1,s)}(\mathbf{R}^d)$, as required. □

16.1. **The spectrum of the Laplacian.** The purpose of this section is the following theorem.

Theorem 116. *Let $\Omega \subset \mathbf{R}^d$ be a bounded domain. Then there is an orthonormal basis $w_1, w_2, \ldots$ of $L^2(\Omega)$ and real numbers $0 \le \lambda_1 \le \lambda_2 \le \ldots$ such that $\lambda_j \to \infty$, $w_j \in H_0^1(\Omega) \cap C^\infty(\Omega)$ and*

$$-\Delta w_j = \lambda_j w_j$$

for all j.

Proof. For $f \in L^2(\Omega)$, we write $Af = u$ for the unique $u \in H_0^1(\Omega)$ that satisfies

$$\langle u, v\rangle_{H^1} = \langle f, v\rangle_{L^2} \tag{10}$$

for all $v \in H_0^1(\Omega)$. This is obviously a linear operator, and we have

$$\|u\|_{L^2}^2 \le \|u\|_{H^1}^2 = \langle f, u\rangle_{L^2} \le \|f\|_{L^2}\|u\|_{L^2},$$

which implies that $\|Af\|_{L^2} = \|u\|_2 \le \|f\|_2$ so $\|A\| \le 1$. By the Rellich-Kodrachov theorem, $A(B_{L^2(\Omega)}) \subset B_{H_0^1(\Omega)}$ is compact in $L^2(\Omega)$, so A is a compact $L^2(\Omega) \to L^2(\Omega)$ operator.

We show that A is Hermitian. To this end, let $f, g \in L^2(\Omega)$. Then

$$\langle Af, g\rangle_{L^2} = \overline{\langle g, Af\rangle_{L^2}} = \overline{\langle Ag, Af\rangle_{H^1}} = \langle Af, Ag\rangle_{H^1} = \langle f, Ag\rangle_{L^2}.$$

Note that $Af, Ag \in H_0^1(\Omega)$, so (10) can be applied.

Now we can use the spectral theorem for compact operators from Part II Linear Analysis and conclude that there is an orthonormal basis $w_1, w_2, \ldots$ of $L^2(\Omega)$ such that $Aw_j = \mu_j w_j$ for some $\mu_j \in \mathbf{R}$ and $\mu_j \to 0$ as $j \to \infty$. We note that $Aw_j \neq 0$ for any j, for otherwise

$$\langle w_j, v\rangle_{L^2} = \langle Aw_j, v\rangle_{H^1} = 0$$

for all $v \in H_0^1(\Omega)$, which is not possible because $H_0^1(\Omega)$ is dense in $L^2(\Omega)$. (Hint: use the argument we had for the density of $C_c^\infty(\mathbf{R}^d)$ in $L^2(\mathbf{R}^d)$ to show that $C_c^\infty(\Omega)$ is dense in $L^2(\Omega)$.) In addition,

$$\mu_j = \langle w_j, Aw_j\rangle_{L^2} = \langle Aw_j, Aw_j\rangle_{H^1} = \|Aw_j\|_{H^1}^2 > 0.$$

This implies that $\mu_j > 0$ for all j, so we may assume that $1 \le \mu_1 \le \mu_2 \le \ldots$.

Note that

$$\langle w_j, v\rangle_{H^1} = \langle \mu_j^{-1} w_j, v\rangle_{L^2}$$

for all $v \in H_0^1(\Omega)$. We know that $w_j \in H_0^1(\Omega)$, and repeating the induction argument in the proof of Theorem 114, we can prove that $w_j \in H^3_{loc}(\Omega), H^5_{loc}(\Omega), \ldots$. By Sobolev embedding, we conclude that $w_j h \in C^\infty(\Omega)$ for all $h \in \mathcal{D}(\Omega)$. For all closed balls $B \subset \Omega$ we may find $h \in \mathcal{D}(\Omega)$ such that $h|_B \equiv 1$, so w_j is smooth on B. Since B is arbitrary, this implies $w_j \in C^\infty(\Omega)$.

Therefore, $\mu_j w_j$ is a classical solution of the equation

$$-\frac{1}{4\pi^2}\Delta\mu_j w_j + \mu_j w_j = w_j.$$

We conclude

$$-\Delta w_j = \frac{4\pi^2}{\mu_j}(1-\mu_j)w_j =: \lambda_j w_j.$$

□

References

[1] H. Brezis, *Functional analysis, Sobolev spaces and partial differential equations*, Universitext, Springer, New York, 2011. MR2759829 ↑34, 39

[2] M. M. Rao, *Measure theory and integration*, Second, Monographs and Textbooks in Pure and Applied Mathematics, vol. 265, Marcel Dekker, Inc., New York, 2004. MR2031535 ↑12

[3] H. L. Royden, *Real analysis*, Third, Macmillan Publishing Company, New York, 1988. MR1013117 ↑12

[4] W. Rudin, *Real and complex analysis*, Third, McGraw-Hill Book Co., New York, 1987. MR924157 ↑6, 12

[5] C. Warnick, *Analysis of functions*. https://www.dpmms.cam.ac.uk/~cmw50/resources/Part-II-AoF/AoF.pdf. ↑49