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 5 February 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

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

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

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

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 $f = g$ hold 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 [3, 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 $E\backslash A_n \subset \bigcup_{s \geq t} P_t$, 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 $\|Mf\|_1 \le 5^d$. 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 $\|f - g\| < \varepsilon$.*

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. $\square$

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 [3, Chapter 2], [2, Chapter 14] or [1, 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 $\operatorname{dist}(x_0, y) < 1/n$. 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 $\operatorname{dist}(y_n, x) < 1/n$. 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. □

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*

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

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 $\mu(U \backslash K) \leq \varepsilon$. 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 $\sum_{n=1}^{\infty} \varepsilon/2^n = \varepsilon$.

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 real valued case by approximating the real and imaginary parts separately. Then for the real valued case, we may consider a strictly monotone continuous bijection $h : \mathbf{R} \to (0,1)$, 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 $\|g\|$, but we can remedy this if we multiply f with a suitable complex number of unit modulus so that $\|\operatorname{Re}(f)\|_\infty = \|\operatorname{Im}(f)\|_\infty$.) □

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, 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. Maybe this does not follow from the statement of Lusin's theorem immediately, but certainly from its proof. Now

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

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

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

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

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

and for $x \in A$ and $n > N(k)$ we have $|f_n(x) - f(x)| \le 1/k$. 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 $L(a_1f_1 + a_2f_2) = a_1L(f_1) + a_2L(f_2)$ for all $a_1, a_2 \in \mathbf{C}$ and $f_1, f_2 \in C(E)$ and $\|Lf\| \le A\|f\|$ for some $A \in \mathbf{R}_{\ge 0}$ 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 $L(f) \ge 0$ for all $f \in C(E)$ with $f(\mathbf{R}) \subset \mathbf{R}_{\ge 0}$.

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 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.

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 $\mu(\bigcup U_n) \leq \sum \mu(U_n)$. To this end, it is enough to show that $L(f) \leq \sum \mu(U_n)$ 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 $f_j \prec U_j$ for each j. Then it will follow that

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

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

$$g_j(x) = \frac{\mathrm{dist}(x, 1_{E\backslash U_j})}{\mathrm{dist}(x, 1_{E\backslash U_1}) + \ldots + \mathrm{dist}(x, 1_{E\backslash U_N}) + \mathrm{dist}(x, \mathrm{supp}\, f)}.$$

First observe that the denominator is never 0. Indeed, if $\mathrm{dist}(x, \mathrm{supp}\, f) = 0$, then $x \in \mathrm{supp}\, f$ and $x \in U_j$ for some j, and then $\mathrm{dist}(x, E\backslash U_j) > 0$. Therefore each g_j is continuous. Moreover $g_j(X) \subset [0, 1]$ and $\mathrm{supp}\, g_j \subset U_j$. Finally, $g_1(x) + \ldots + g_N(x) = 1$ for all $x \in \mathrm{supp}\, f$. Now it is easy to see that $f_j = fg_j$ 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 $L(f_1) \ge \mu(V \cap U) - \varepsilon/2$, which exists by definition. Then we consider the set $V_2 = V \backslash \mathrm{supp}\, f_1$. This is an open set and $V_2 \supset V\backslash U$ so there exists $f_2 \prec V_2$ with

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

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

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

□

Lemma 23. *Let $A \in \mathcal{B}$, and let $f \in C(E)$ with $f(E) \subset [0, 1]$. If $f(x) \le 1_A(x)$ for all x then $L(f) \le \mu(A)$. If $f(x) \ge 1_A(x)$ for all x then $L(f) \ge \mu(A)$.*

Proof. Let $\varepsilon > 0$ be arbitrary and $U \subset A$ open. Note that $\|f - \max(f - \varepsilon, 0)\| \le \varepsilon$ hence

$$|L(f) - L(\max(f - \varepsilon, 0))| \le \varepsilon \|L\|.$$

On the other hand,

$$\operatorname{supp}\max(f-\varepsilon,0)\subset\{x:f(x)\geq\varepsilon\}\subset A\subset U,$$

so by definition of $\mu(U)$,

$$\mu(U)\geq L(\max(f-\varepsilon,0))\geq L(f)-\varepsilon.$$

Since $U\supset A$ and $\varepsilon>0$ are arbitrary, $\mu(A)\geq L(f)$ by the definition of $\mu(A)$.

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 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

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

We sum this up for $j=0,\ldots,n-1$, and get

$$\Big|L(f)-\int f d\mu\Big|\leq\sum_{j=0}^{n-1}\frac{\mu(A_j)-\mu(A_{j+1})}{n}=\frac{\mu(A_n)-\mu(A_0)}{n}\leq\mu(E)/n.$$

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

$$\|f\|_\infty = \inf\{\lambda \in \mathbf{R}_{\geq 0} : f(x) \leq \lambda \ \mu\text{-almost everywhere}\}.$$

(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

$$\|f_1 + f_2\|_p \leq \|f_1\|_p + \|f_2\|_p$$

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:

$$\Big| \int_E f(x)g(x)d\mu(x) \Big| \leq \|f\|_p \|g\|_q$$

whenever $p, q \in [1, \infty]$ and $p^{-1} + q^{-1} = 1$.

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

Theorem 24 (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

$$\|g\|_p \leq \int \|G(\cdot, y)\|_p d\nu(y).$$

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 $|G(x, y)| \leq h(y)$ for $\mu \otimes \nu$ almost every (x, y). We conclude $|g(x)| \leq \int |h(y)| d\nu(y)$ for almost all x, which proves the claim.

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

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

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 $G(x, y) \geq 0$, and write

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

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 25. *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 26. *The space $L^p(\mathbf{R}^d)$ is separable for all $1 \leq p < \infty$.*

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 27. *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 28. *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

$$L(f) = \int fgd\mu = \mu(A),$$

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

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

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 $L(f) = \int fgd\mu$ 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, $\int f_n g d\mu \to \int fgd\mu$ again by dominated convergence. Therefore, $L(f) = \int fgd\mu$, 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 $f, g : \mathbf{R}^n \to \mathbf{R}$ 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 29 (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

$$\begin{aligned}\Big(\int |f*g(x)|^p dx\Big)^{1/p} &\le \Big(\int \Big(\int |f(t)g(x-t)|dt\Big)^p dx\Big)^{1/p}\\ &\le \int \Big(\int |f(t)g(x-t)|^p dx\Big)^{1/p} dt\\ &= \|g\|_p \int |f(t)|dt\\ &= \|g\|_p\|f\|_1.\end{aligned}$$

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

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

Now we can write

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

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

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

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 30. *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 $p^{-1} + q^{-1} + r^{-1} \ge 2$, but we do not prove this.

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

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

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

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

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

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

□

Proposition 31. *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

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

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 $\alpha \in \mathbf{Z}^d_{\geq 0}$ 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 $\alpha \in \mathbf{Z}^d_{\geq 0}$.

Proposition 32. *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 $\alpha \in \mathbf{Z}^d_{\geq 0}$.

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 33. *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 34. *Let f_n be an approximate identity. 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

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

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

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

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

Finally, we can write

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

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

Proposition 35. *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 $h \geq 0$ and $\int h dx = 1$, for example

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

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 32. It is immediate from the definition of convolution that the convolution of compactly supported functions is also compactly supported. □

References

[1] 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

[2] H. L. Royden, *Real analysis*, Third, Macmillan Publishing Company, New York, 1988. MR1013117

[3] W. Rudin, *Real and complex analysis*, Third, McGraw-Hill Book Co., New York, 1987. MR924157